Big data in digital product innovation

Identifying the antecedents and consequences of using big data in digital product innovation – a multiple case study

Qiu, Y.

DOI
10.4233/uuid:c8b86311-fe3d-440e-9983-4e6e4f1dfb67

Publication date
2024

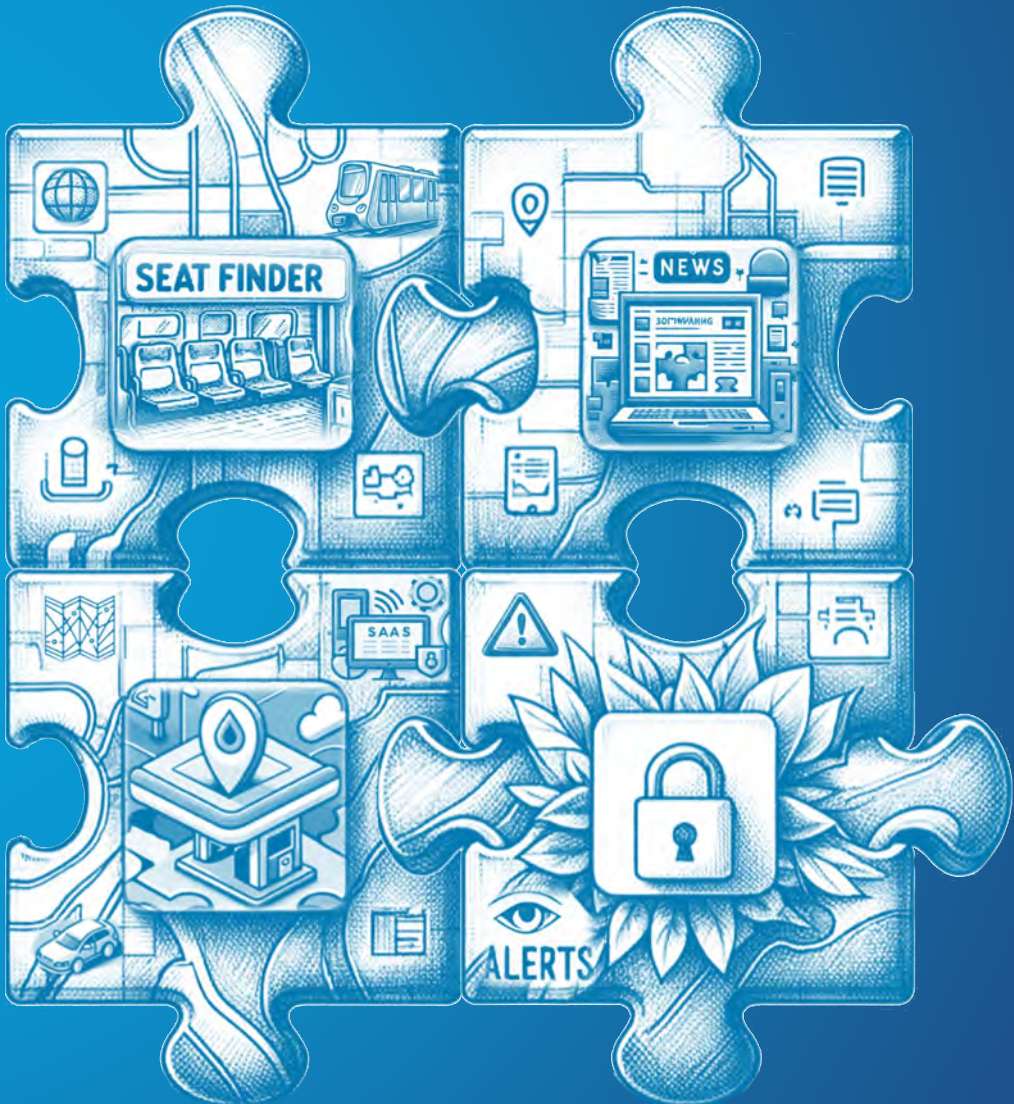
Document Version
Final published version

Citation (APA)
Qiu, Y. (2024). *Big data in digital product innovation: Identifying the antecedents and consequences of using big data in digital product innovation – a multiple case study*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:c8b86311-fe3d-440e-9983-4e6e4f1dfb67

BIG DATA IN DIGITAL PRODUCT INNOVATION

Identifying the antecedents and consequences of using big data in digital product innovation –

a multiple case study



Yunran (Grace) Qiu

BIG DATA IN DIGITAL PRODUCT INNOVATION

Identifying the antecedents and consequences of using big data in digital product innovation – a multiple case study

Yunran (Grace) Qiu

BIG DATA IN DIGITAL PRODUCT INNOVATION

Identifying the antecedents and consequences of using big data
in digital product innovation –
a multiple case study

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J van der Hagen

Chair of the Board for Doctorates
to be defended publicly on
Monday 22 April 2024 at 15:00 o'clock
by

Yunran QIU

Master of Arts in Human Resource Management,
Newcastle University, The United Kingdom
born in Nanchang, China

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. H.J.Hultink	Delft University of Technology, promotor
Dr. P.Cankurtaran	Delft University of Technology, copromotor

Independent members:

Prof. dr. H.M.J.J.Snelders	Delft University of Technology
Prof. mr. dr. ir. S.C.Santema	Delft University of Technology
Prof. dr. ir. J.C.Diehl	Delft University of Technology
Dr. T.J. Marion	Northeastern University, USA
Prof.dr.P.A.Lloyd	Delft University of Technology, reserve member

Keywords: Big Data, Digital Product Innovation, Case Study
Printed by: Gildeprint, www.gildeprint.nl
Cover designed by: Yunran Qiu, OpenAI DALL-E, Ilse Modder
Lay-out by: Ilse Modder, www.ilsemodder.nl

ISBN 978-94-6384-562-5
An electronic version of this dissertation is available at
https://repository.tudelft.nl/

Table of Contents

Summary

Innovation is critical in driving economic growth. It fosters new products and services to move the economy forward. On the other hand, innovation requires new technologies to thrive. One such technology that has revolutionised various industries is big data. By harnessing big data, businesses can unlock valuable insights to enhance decision-making processes and ultimately encourage innovation across sectors. While big data holds the potential to drive innovation and economic growth, managing big data-enabled digital product innovation projects is challenging. Organisations need the resources, knowledge, and skills to manage these projects. In addition, companies should also be aware of the privacy and security risks associated with these big data innovation projects. Considering all the many factors influencing big data innovation projects, it is necessary to understand how they impact these projects in order to manage them well. This doctoral project answers the main research question: How do companies develop big data analytics capabilities in digital product innovation? This doctoral research identifies the different factors that influence big data innovation projects and reveals the mechanisms behind how these factors influence each other and affect digital product innovation.

This research contains a multiple-case study with four cases to explore the answers to the research questions. The case studies analysed four new digital service development projects from four big companies operating in different industries (i.e. transportation, internet, navigation, and cybersecurity) in the past five years. In these cases, big data from various sources are used as input for machine learning and statistical models to enable digital product innovation. For collecting research data for this study, the semi-structured interview protocol was used, focusing on the process of using big data in innovation projects, the challenges and factors that affect this process, and its consequences.

This study's findings highlight the mechanism of how antecedents influence big data analytics capabilities (BDAC) and subsequently impact innovation performance. The study applies the resource-based view (RBV) and dynamic capabilities view as theoretical frameworks to clarify the mechanism behind enhancing innovation performance. It investigates the antecedents of BDAC constructs, explores the interactions among these constructs, and evaluates their effects on innovation performance. This research enriches the existing literature by revealing how BDAC constructs promote their development, unveiling the dynamics within BDAC, and emphasising the vital role of data variety. Furthermore, it extends the literature by uncovering novel connections between dynamic capabilities and

BDAC and between environmental uncertainty and BDAC, broadening the scope of environmental uncertainty and examining the effects of BDAC constructs on innovation performance. For practitioners, this research suggests that innovation performance can be ultimately improved by enhancing data variety and establishing supportive internal and external environments for BDAC.

Samenvatting

Innovatie is cruciaal voor het stimuleren van economische groei. Het bevordert de ontwikkeling van nieuwe producten en diensten om de economie vooruit te helpen. Aan de andere kant vereist innovatie de toepassing van nieuwe technologieën. Een dergelijke technologie die verschillende sectoren heeft opgeschud is het integreren van big data. Door gebruik te maken van big data kunnen bedrijven waardevolle (consumenten)inzichten ontsluiten om innovatie in verschillende sectoren aan te moedigen. Hoewel big data het potentieel heeft om innovatie en economische groei te stimuleren, is het managen van op big data gebaseerde digitale productinnovatieprojecten uitdagend. Organisaties hebben middelen, kennis en vaardigheden nodig om deze projecten te managen. Bovendien moeten bedrijven zich ook bewust zijn van de privacy- en beveiligingsrisico's die gepaard kunnen gaan met big data-innovatieprojecten. Gezien de vele factoren die van invloed kunnen zijn op big data-innovatieprojecten, is het noodzakelijk om te begrijpen hoe deze projecten worden beïnvloed om ze goed te kunnen managen. Dit doctoraatsonderzoek geeft antwoord op de volgende onderzoeksvraag: Hoe ontwikkelen bedrijven big data-analysevaardigheden voor digitale productinnovatie? Dit onderzoek identificeert de verschillende factoren die van invloed zijn op big data-innovatieprojecten en onthult de mechanismen hoe deze factoren elkaar en digitale productinnovatie beïnvloeden.

Dit onderzoek betreft een multiple-case studie met vier cases om antwoorden op de onderzoeksvraag te verkennen. De case studies hebben vier nieuwe ontwikkelingsprojecten voor digitale diensten geanalyseerd, afkomstig van vier grote bedrijven die actief zijn in verschillende sectoren (namelijk transport, internet, navigatie en cybersecurity) in de afgelopen vijf jaar. In deze cases wordt big data van verschillende bronnen gebruikt als input voor "machine learning" en statistische modellen om digitale productinnovatie mogelijk te maken. Voor het verzamelen van onderzoeksgegevens voor deze studie werd een semi-gestructureerd interviewprotocol gebruikt, met de focus op het proces van het gebruik van big data in innovatieprojecten, de uitdagingen en factoren die dit proces beïnvloeden, en de gevolgen ervan.

De bevindingen van deze studie benadrukken hoe antecedenten "big data analytics capabilities" (BDAC) en vervolgens innovatieprestaties beïnvloeden. De studie past de "resource-based view" (RBV) en de "dynamic capabilities view" toe als theoretische kaders om het mechanisme achter de innovatieprestaties te ontsluiten. Het onderzoekt de antecedenten van BDAC-constructen, verkent de interacties tussen deze constructen, en evalueert hun effecten op innovatieprestaties. Dit

onderzoek verrijkt de bestaande literatuur door te onthullen hoe BDAC-constructen de ontwikkelingvan digitale innovaties bevorderen, de dynamiek binnen BDAC bloot te leggen, en de cruciale rol van data-variëteit te benadrukken. Bovendien breidt het de literatuur uit door nieuwe verbindingen tussen “dynamic capabilities en BDAC” en tussen “omgevingsonzekerheid en BDAC” bloot te leggen. Voor praktijkmensen suggereert dit onderzoek dat innovatieprestaties kunnen worden verbeterd door data-variëteit te vergroten en ondersteunende interne en externe omgevingen voor BDAC te creëren.

CHAPTER 1

Introduction

Innovation is critical in driving economic growth. It fosters novel products and services to move the economy forward (Chesbrough, 2011). Innovation requires new technologies to thrive. One such technology that has revolutionised various industries is big data. By harnessing big data, businesses can unlock valuable insights, enhance decision-making processes, and ultimately encourage innovation across sectors. While big data holds immense potential for driving innovation and economic growth, managing big data-enabled digital product innovation projects is challenging. Organisations need the resources, knowledge, and skills to manage and leverage big data-enabled digital product innovation initiatives. In addition, companies should also be aware of the privacy and security risks associated with these big data innovation projects. Considering all the many factors influencing big data innovation projects, it is necessary to understand how they impact these projects in order to manage them well. This research aims to identify the factors that influence such projects and then shed light on the underlying mechanisms of how these different factors influence each other and affect the use of big data in digital product innovation.

Before exploring the mechanisms behind how these factors influence the use of big data in digital product innovation, it is necessary to explain the context. This chapter first explains the crucial role of innovation for the economy, then brings up the topic of innovation management: innovation process management, the success factors of innovation management, and innovation management seen through the resource-based view. Then the chapter moves on to big data and its role in the innovation process, along with applications in different industries. Apart from the positive role of big data, it also touches upon the societal impacts of big data. The chapter ends with the research motivations.

1.1 Innovation and economy

According to the Organization for Economic Cooperation and Development (OECD) (2005), product innovation is defined as “the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations”. Innovation is a major economic growth force (Rosenberg, 2006). An OECD report (2015) emphasises that innovation contributes significantly to economic growth, often around 50% of total GDP growth, depending on the country’s development level. Innovation brings novelty and variety to the economy. If the flow of novelty or innovation dries up, the economy will settle in a “stationary state” with little or no further growth (Metcalfe, 1998). Innovation is, therefore, crucial to the growth of the economy in the long term (Fagerberg, 2009).

Due to innovation, countries, regions, and firms differ greatly in terms of economic performance (Bilbao-Osorio & Rodríguez-Pose, 2004). Compared to less innovative countries and regions, innovative countries have higher rates of productivity and income (Fagerberg & Godinho, 2006). The same is true for companies. A company that innovates prospers. Companies survive by developing new products and introducing them to the market (Cooper, 2001; Schilling & Hill, 1998). Regardless of the innovative firm's industry, innovation is positively associated with revenue growth (Thornhill, 2006). The speed at which innovations are introduced improves a firm's operational and financial performance, while the quality of innovations contributes to financial gains (Wang & Wang, 2012). Due to these desirable consequences, business leaders strive to foster innovation (Breznitz & Ornston, 2018; Muro & Katz, 2011).

1.2 Innovation management

Companies that innovate effectively thrive. When aiming to foster innovation and thrive, a company's ability to design and develop new products and effectively bring them to market becomes crucial. Successfully introducing an innovation – whether in-house or to the market – requires goal-oriented steering and shaping of innovation activities (Gaubinger, Rabl, Swan & Werani, 2015). These wide-ranging structured and coordinated activities fall under the category of innovation management.

The following section discusses innovation management in more detail, starting from how innovation activities have been organised and managed – how can the innovation process be divided into different stages, and what is the general process? The success factors of innovation management are then presented. Finally, innovation management is explored from the resource-based view.

1.2.1 The different stages of innovation

The previous section introduced the concept of innovation management. This section zooms into the product innovation process and how the process can be divided into different stages.

Product innovation is critical for the growth and profitability of many organisations, but it is also highly risky because only a small number of such innovations make it to success (Crawford, 1979). In the 1980s, most new product success and failure studies focused on the "bigger picture" of strategy, synergy, orientation, and which market and technology to choose (Cooper & Kleinschmidt, 1986). Few studies aimed at investigating what happens during the product development process and whether these activities' strengths and weaknesses impact the outcome of the process (Cooper & Kleinschmiddt, 1986). Thus, Cooper and Kleinschmidt (1986) conducted an empirical study to discover what activities and tasks were undertaken in the new product development projects, how well they went, and the impact of each activity on the project outcomes (Cooper & Kleinschmidt, 1986). The study revealed the activities that happened in the process, whereby the process can be divided into stages, and it built a foundation for the stage-gate model. Since then, many models have been introduced to describe the different stages of innovation for managing the NPD process.

1.2.2 Innovation process stages in this research

Many models describe and guide the innovation process. They have all gone through many iterations. Since models are intended to depict and capture complex phenomena, they often oversimplify certain aspects (Trott, 2012). When it comes down to choosing the appropriate models, different models might fit different projects. Various factors need to be considered, such as the size of the development team, and whether there is limited knowledge of users' needs and preferences. To guide product innovation activities in practice, there is no model that fits all situations (Trott, 2012).

Take three typical models for example – agile, stage gate and design thinking. These models represent different focuses and schools of thought when it comes to managing an innovation process in practice. The agile model could be an easy fit for big data-enabled digital product innovation projects because big data innovation projects can be broken down into smaller ones (Larson & Chang, 2016). Agile is useful in iterating smaller projects, especially in supporting explorations and validation in predictive and prescriptive analysis (Ambler & Lines, 2016). However, innovation projects need structures from a strategic level. Adopting the stage-gate model to organise structures from a strategic level and the agile model to balance productivity and flexibility would make the most out of the two models and best suit the needs of big data-enabled innovation projects (MacCormack & Verganti, 2003). Design thinking is great for ill-defined problems (Brown, 2008). Combining design thinking with big data could uncover users' needs on a bigger scale while also enabling deep understanding when focusing on certain groups (Mortati et al., 2023).

As discussed above, when choosing a suitable model for big data-enabled innovation, the most important consideration is to adjust and combine the original model with big data. It must be kept in mind that in practice, different models are adopted to different extents, with some being adopted strictly and others more loosely (Dunne & Martin, 2006; Sethi & Iqbal, 2008). Thus, in order to describe the different stages of product innovation, this dissertation looks at the main activities or stages that most models have in common: pre-development, development, and commercialisation (Langerak, Hultink & Robben, 2004; Trott, 2012). Pre-development represents the ideation and planning activities carried out before moving on to development. Several activities occur during the pre-development stage, such as strategic planning, business and market analysis, and generating and evaluating ideas for new products. This is the stage where the team gathers and processes customer needs and matches them with potential technical solutions (Dougherty et al., 2000). Development is the main stage of the actual development of the product. This stage focuses on the specifications of the product. Concepts for new products are developed, and prototypes are tested with potential customers. Market research is needed at this stage to establish product goals and determine feature trade-offs. Commercialisation focuses on market launch. This stage involves the release of new product specifications to manufacturing and deciding on launch tactics (Hultink et al., 1998).

In this research, the focus is on the role big data plays in the innovation process. Specifically, this study focuses on using big data in the pre-development and development stages, where the team gathers information on customers' needs, and then designs and develops the product to fulfil the needs with technical solutions (Dougherty et al., 2000). Thus, later in Section 1.6, this dissertation organises the knowledge about the different roles big data can play in the process by dividing it in a generic way: pre-development, development, and commercialisation.

1.2.3 Success factors of innovation management

Moving on from the discussion of generic models of the product innovation process, the dissertation now shifts attention to the success factors in innovation management. While the stages of innovation shed light on the different aspects of the progress of product innovation, the success factors could reveal the underlying reasons why some innovations are effective while others fail.

Is there a "formula" for innovation success? Success cannot be guaranteed by simply copying the "formula" to another company, industry or across countries. However, it is still possible to identify the basic patterns of success in terms of factors or their combinations (Gaubinger et al., 2015).

A group of scholars started research in this field in the 1980s. Their recent study analysed more than 2,000 NPD projects over the past 30 years (Cooper &

Kleinschmidt, 2011). Another group of scholars conducted a meta-analysis of success/failure studies in the past decades (e.g. Szymanski and Henard, 2001). A newly updated meta-analysis of this research stream reported the success factors based on 233 empirical studies on new product success published from 1999 through 2011 (Evanschitzky, Eisend, Calantone & Jiang, 2012). Table 1 presents the most important predictors of new product success according to the results of the most recent meta-analysis (Evanschitzky et al., 2012):

Table 1. The predictors of new product success (based on Evanschitzky et al., 2012)

Predictor of new product success	Definition
Product characteristics	
Product advantage	Superiority or distinctive features, quality or characteristics that make the product outperform its competitors
Product technological sophistication	How do stakeholders such as customers recognise the level of advancement of the product from technological point of view (i.e. often described as high-tech, low-tech)
Strategy characteristics	
Marketing synergy	The compatibility that company's marketing skills match with the ones needed to successfully implement product innovation projects
Technological synergy	Congruency between the existing technological skills of the firm and the technological skills needed to execute a new product initiative successfully
Company resources	The capabilities, skills, and knowledge a company can get access to when it comes to business activities such as product development
Dedicated human resources	A specialised team of people that focuses on supporting the product development projects
Dedicated R&D resources	The capabilities, skills, and knowledge from a company that focuses on supporting research and development
Strategic orientation	The company's overall mindset, focus, or philosophy for strategic planning and decision making
Process characteristics	
Structured approach	Formalised, organised or systematic procedures for product innovation
Predevelopment task proficiency	High level of efficiency and effectiveness in conducting prelaunch activities (e.g. idea generation, idea screening, market research)
Marketing task proficiency	High level of efficiency and effectiveness with which a company conducts its marketing activities
Launch proficiency	High level of efficiency and effectiveness with which a company launches the product/service
Reduced cycle time	Shorten the concept-to-introduction time (i.e. time to market)
Market orientation	The activities company conduct to identify customer needs so that product can be designed and developed to fulfil the needs

Table 1. Continued.

Predictor of new product success	Definition
Cross-functional integration	The degree of collaboration, coordination, and communication among internal functions within a company for product innovation projects
Cross-functional communication	The level of communication among departments or different functions in a product innovation project
Senior management support	Degree of senior management support for a product innovation project
Marketplace characteristics	
Likelihood of competitive response	Degree or likelihood of competitive response received from the market to a new product introduction
Competitive response intensity	High degree or level of competitive response received from the market to a new product introduction
Market potential	Anticipated market size or growth in customer demand
Environmental uncertainty	The degree of complexity in the external environment where the company operates in (e.g. regulatory environment, technology uncertainty)
Organisational characteristics	
Organisational climate	The atmosphere of a company that reflects how day-to-day decisions are made based on the shared values and norms within the company
External relations	The collaboration and coordination between the company and other companies or organisations
Degree of formalisation	Extent to which structured, organised or explicit rules and procedures govern decision-making in the company or project

When these factors are applied to big data-enabled digital product innovation projects, the results indicate that the factors product technological sophistication and technological synergy are closely related to technology and thus might seem more relevant. In fact, all factors could also apply to big data-enabled digital product innovation projects. For example, these projects would involve multiple teams playing different functions. Thus, cross-functional communication would be important to keep teams together to reach their goal of making a successful product. Outside the company, environmental uncertainty would also affect the success of a big data innovation project (e.g. regulation environment and technology uncertainty).

The finding also revealed that the importance of success factors generally declines over time, suggesting that new and more comprehensive theoretical approaches are needed to capture the underlying nature of NPD success factors (Evanschitzky, 2012). The scholars also pointed out a possible explanation for this. Knowledge of these success factors is already widespread among managers (Evanschitzky, 2012). The new technology trend of big data and artificial intelligence might also play a role in changing the success factors of innovation. In addition, innovation goals such as sustainability could redefine the success of product innovations.

Even though the success factors have become well known and their importance has declined over time, some factors might remain relevant, and it might be worthwhile to explore the mechanisms behind the underlying nature of NPD. With big data becoming available for NPD, it would be interesting to find out what difference big data could make to NPD. Looking at the success factors in the table above, these big data-related factors could point out a direction for exploring the mechanism behind using big data for innovation. Considering big data as a technology, the success factors regarding technological advancement and implementation are relevant for this research. For example, in terms of product technological sophistication, which concerns the perceived technological sophistication of the product, big data could enhance the product, turning it into a more sophisticated high-tech product than one without big data. This poses the question of how does big data make a difference? What is the mechanism behind it? For example, does it make a difference because of volume, the large amount of data or the variety of different sources of big data? This research project will explore these questions.

The success factors could yield some clues about how to manage innovation or what factors could affect the innovation process. However, as the success factors keep on changing, there is no consistent best practice for innovation management (Evanschitzky, 2012; Tidd, 2001). The reason behind this phenomenon is that the best practice is contingent on a range of factors (Tidd, 2001). These factors can be organised in different ways; for example, Tidd (2001) categorises these factors into organisation configurations and environmental contingencies. The following section discusses these resources in more detail.

1.2.4 Resources, innovation, and competitive advantage

Countries, regions, and companies that innovate thrive and rely on their ability to use resources and capabilities for this purpose. From a corporate perspective, the foundation of a company is a set of material resources, human skills, relationships, and relevant knowledge (Abernathy & Clark, 1985). Companies rely on these competencies or competitive ingredients to design and develop products that are attractive for the market (Abernathy & Clark, 1985). Grant (1991) articulates the association between a firm's resources and competitive advantage: For a company, the resources are the inputs for its production process. The resources can be financial, physical, human, technological, reputational, and organisational. For example, big data itself is a technological resource that companies can leverage to gain a competitive advantage. Capabilities comprise the capacity of a team of resources to perform some task or activity. Capabilities constitute the main source of competitive advantage. The complexity of the capabilities of a company can build a huge barrier that hinders others from entering the market, thus making it easier to sustain its competitive advantage. For example, big data analytics capabilities (BDAC) comprise a company's ability to capture and analyse data through processes and structures to generate insights. It contains a group of knowledge and skills that companies get access to. However, just like success factors, the competitive advantage also erodes over time, mainly due to depreciation and imitation by rivals. Maintaining competitive advantage requires companies to develop their resource base continually. From this perspective, managing innovation means managing resources that are relevant to product innovation. Thus, this research investigates big data-enabled digital product innovation from a resource-based view, with a central focus on BDAC. The aim is to reveal the antecedents and consequences of developing BDAC. More details are provided in Chapter 2: literature review and theoretical background.

To summarise, innovation is the driving force that keeps the economy moving. Managing innovation is a complex task that involves various activities. The best practice for innovation keeps on changing because it is contingent on different factors, including resources, capabilities, and the external environment. Technology also plays an important role in it. When new technologies become available, they can be used for product innovation purposes as well. One of them is big data.

1.3 Data, data everywhere

The first modern meaning of the word "data" is recorded in 1946 as "transmittable and storable information by which computer operations are performed" (Etymology Dictionary, 1946). The term "big data" was coined in Silicon Valley in the mid-1990s (Diebold, 2012). It describes the exponential growth of digital data generated from various sources and different formats. So, big data has been around for quite a while – what is the situation now and where does big data come from? Currently, 2.5 quintillion bytes (one followed by 18 zeros) of data are created every day (Marr, 2021). These numbers might be difficult to comprehend. Thus, here are some statistics illustrating the amount of data being generated, from search engines to video streaming, communication, transportation, to IoT (Marr, 2021):

- Google currently processes more than 40,000 searches per second (3.5 billion searches per day).

- YouTube users watch 4,146,600 videos per minute.
- Skype users make 154,200 calls per minute.
- Uber riders take 45,788 trips per minute.

The Internet of Things, which connects "smart" devices with each other and collects data of all kinds, is growing rapidly, leaping from 2 billion devices in 2006 to 200 billion devices in 2020. Huge amounts of data are being generated – and the speed is still accelerating.

To summarise, there is an explosion of huge amounts of data of various types: from mobile phones, the internet, and company accounts to geospatial and biomedical data. The amount of new data created, captured, replicated, and consumed annually is expected to double in four years (IDC, 2022). It is a nonstop, growing tidal wave. These data have huge potential, one of which involves combining them with artificial intelligence (AI) to contribute to product innovation. The following sections discuss this in greater depth.

1.4 Big data and AI

Despite the volume of big data, there are ways to harness and leverage its potential for valuable insights and transformative applications. Data generated and collected from various sources, such as IoT devices, can be used in the computer learning process (Pencheva, Esteve & Mikhaylov, 2020). This strongly contributes to the increasing maturity of artificial intelligence (AI) technologies and the viability of AI applications (Pencheva et al., 2020). AI and big data are closely related to each other. One way in which this can be seen is that big data is harvested from IoT physical devices and digital platforms and systems. These data are then fed to an AI model or big data analysis model (Rathore, Shah, Shukla, Bentafat & Bakiras, 2021). From another perspective, AI can be seen to be applied in several ways to facilitate the capturing and structuring of big data and the analysis of big data for key insights (O'Leary, 2013). AI is especially useful in structuring and analysing big data. For example, after an advertisement is published, analysts can access structured transaction data like the time and location of the ad, as well as unstructured data, such as how many tweets mention the ad. Combining these data provides insights into the ad's impact, revealing whether it generated positive or negative reactions from the audience (O'Leary, 2013). Machine learning (ML) is known as a subfield of AI (Cioffi, Travaglioni, Piscitelli, Petrillo & De Felice, 2020). For product innovation, big data collected from IoT sensors is stored in databases as input for ML-based AI to realise automated smart machines for traditional industries such as agriculture, enabling the remote control and monitoring of greenhouse equipment (Misra, Dixit, Al-Mallahi, Bhullar, Upadhyay & Martynenko, 2020). In this research project, big data is also used as input for machine learning models for purposes such as utilising user navigation big data as input for ML algorithms to estimate the number of customers visiting petrol stations. This estimation is one of the main functions of the software as a service product. Another example is the use of network big data for ML algorithms to distinguish normal and abnormal network activities, which is a key function of the cybersecurity system product in Case 4. The following section discusses big data and product innovation in greater depth.

1.5 Big data and innovation

Today, many companies can get access to a large amount of data from various sources, including data collected from inside the organisation, such as transactional data, and outside of the organisation, such as user-generated data (e.g. Zuboff, 2015). Often, data are collected unintentionally and in an unstructured manner, that is, these data were not produced or collected for the same purposes they are eventually used for (Anderson 2008, Constantiou and Kallinikos, 2015).

With such a vast amount of valuable data, many applications are possible, including innovation. However, such data-enabled innovations were not part of the original plan. This phenomenon is observed in this research as well; for example, two out of four of the case companies already had access to huge amounts of these types of data but did not have a plan to use them for innovation projects (e.g. Case 1 & Case 3). It was not until the companies realised the value of big data for product innovation, and the role it could play in it, that they began their big data-related innovation projects. The following section discusses the possible role big data could play in innovation.

1.6 The role of big data in the innovation process

Analysing big data has a huge impact on innovation. By leveraging various sources of big data, companies can gain valuable insights to better understand customers' needs and understand the market. These insights gained from analysing big data are beneficial in all stages of the innovation process. The following part elaborates on the role big data plays in different stages.

Pre-development

In pre-development, big data can be an important knowledge source for market research and product ideation. Companies can gain insights from big data mining to understand customers' needs and preferences and even predict future needs. For example, using big data collected from online customer reviews and Google Trends makes it possible to predict the product features and attributes that will be relevant to consumers in the future (e.g. Liu, Jin, Ji, Harding & Fung, 2013; Yakubu & Kwong, 2021).

Development

Moving on to the development stage, companies can then develop prototypes based on the customers' preferences that were identified in the pre-development stage. During development, big data is useful for prototype testing and development. Often combined with machine learning, digital products can benefit from big data-based A/B testing and greyscale testing (Sales, Botelho, Patikorn & Heffernan, 2018). A/B testing is an experiment often run on webpages or APPs to compare two versions and find out which one performs better. Greyscale testing involves selecting a pilot group of people for trial before a product is officially released.

Big data also makes it possible to personalise products and services, such as by leveraging biological data for personalised medicine (e.g. Kinkorová & Topolčan, 2020; Vidal, Endris, Jozashoori, Karim & Palma, 2019); big data is also useful for classifying body shapes and modelling the relations between human bodies, fashion themes and design factors to offer fashion design recommendations for designers (Dong, Zeng, Koehl & Zhang, 2020). Other examples are well-known personalised video streaming services like YouTube and Netflix (Amatriain, 2013).

Commercialisation

In commercialisation, leveraging big data can increase the effectiveness of new product launches by identifying the relevant market segment and launching at the right time (Del Vecchio et al., 2018; Martelo-Landroguez et al., 2019). For example, by leveraging big data collected from all advertising channels to reach out to micro-segments by running concentrated marketing campaigns, big data can also be of great help in pricing. For example, analysing sales data from thousands of products with different discount levels facilitates making better decisions on product pricing to optimise profit (Baker, Kiewell & Winkler, 2014).

From pre-development and development to commercialisation, big data technology can be applied to all stages of product innovation. How can big data be applied to product innovation in real life? The following section presents some examples across various industries.

1.7 Example applications of big data in innovation

Various industries are using big data to initiate new products and improve performance. Several examples are provided in the following paragraphs in industries including aircraft, biomedical, farming, and retailing.

1.7.1 Big data in aircraft innovation

It is estimated that half a terabyte of data is generated per flight (Hollinger, 2015). This is machine-generated data derived from sensors. Having access to such a huge amount of data, the aircraft manufacturers Boeing and Airbus are considering how to leverage this valuable information (Hollinger, 2015). In the case of bad weather and turbulence, real-time dashboards based on big data may provide pilots with valuable information to adjust navigation accordingly. Innovation can benefit from this rich information, for example, by building parts that last longer. Analysing sensor data brings valuable insights regarding component vulnerability, which could guide the design and development of these parts (Oh, 2017). Big data also makes predictive maintenance possible; the flight data recorder and sensors serve as data sources, and the data are then analysed with statistics to predict when maintenance tasks are needed for which parts (Tool Sense, n.d).

1.7.2 Big data in biomedical innovation

The biomedical big data industry is experiencing rapid growth with advancements in genome sequencing. Genomic data comprise one of the fastest-growing big data types used in research and personalised medicine. By 2025, 100 million to 2 billion human genomes will be sequenced (Stephens, Lee, Faghri, Campbell, Zhai, Efron, Iyer, Schatz, Sinha & Robinson, 2015). These sequence data are stored in approximately two to forty exabytes. YouTube, in comparison, requires only one to two exabytes each year. Analysing such large amounts of sequence data is undoubtedly expensive. Up to 10,000 trillion CPU hours could be required. According to Stephens et al. (2015), personalised medicine is one of the biomedical applications that can be enabled by this amount of data. Prior to personalised medicine, most patients who did not have a particular type or stage of cancer received the same type of treatment, which worked more effectively for some patients than for others. Personalised medicine is enabled by analysing large-scale data. Researchers in this area are developing methods to analyse large-scale data in order to develop solutions that are tailored to individual needs and are thus hypothesised to be more effective. Cancer patients may still receive standard treatments, such as surgery to remove tumours. Doctors can also recommend personalised cancer treatments. The challenge in biomedical big data applications, as in many other fields, is how to integrate various data sources to gain further insight (Li & Chen, 2014).

1.7.3 Big data in farming innovation

Big data has many applications in innovation in the farming industry as well. For example, it could optimise the yields from existing farmlands. Monsanto, a multinational agricultural biotechnology corporation, has released a prescriptive planting system that generates location-specific customer insights (Vance, 2013). The system combines 25 million mapped fields, 150 billion observations of soil and 10 trillion weather simulations with hundreds of thousands of seeds and terabytes of yield data to determine which seed grows best in which field under what conditions (Vance, 2013). Since Monsanto's system was introduced two years ago, farmers have seen their yields increase by around 5 per cent (The Economist, 2014). According to some seed companies, technology solutions like these can help increase corn production from 160 bushels an acre to over 200 bushels per acre (The Economist, 2014).

Big data can also optimise farm equipment. The farming equipment company John Deere integrates sensors into their farming equipment and applies big data to manage them (Talend, n.d.). This can be a lifesaver as it provides users with information about tractors, such as service due dates, and fuel refill alerts. Therefore, equipment utilisation is maximised and equipment lifespan is extended (Talend, n.d.).

1.7.4 Big data in retailing innovation

With the adoption of consumer technology and the development of multi-channel shopping experiences, data have become increasingly critical for retail industries (Mercier, Richards & Shockley, 2013). According to Marr (2021), Walmart is leveraging big data in the grocery industry. Walmart is building a private cloud that can handle 2.5 petabytes of data per hour. The company is able to gain valuable business insights through the combination of various sources of big data. For instance, when looking at newly released products, it examines which products are often purchased together. In terms of predicting product demand at specific locations, for example, the company analyses what would be the best new product to introduce to their customers. The data sources used by Walmart include Twitter, local events, weather, in-store purchases and online clicks, and many other sales, customers, and product-related data.

Another example comes from Ralph Lauren, famous for its Polo Shirt. The fashion company introduced the Polo Tech Shirt (Marr, 2016). Sensors are attached to silver threads inside the high-tech shirt to collect wearers' movement data and health data such as heart rates, steps taken, and calories burned. This big data-enabled product helps not only improve fitness but also prevent injury from "overdoing" workouts.

In summary, big data has the potential to enable models with higher precision in many application areas – and these highly precise models are influencing and transforming business.

1.8 The ethical impacts of big data

Big data has driven the development of digital products and revolutionised many industries. Its ability to process large amounts of information and uncover valuable insights has led to many advancements and benefits. However, like any technology, there are potential risks and ethical impacts associated with big data.

Big data innovation does not always bring positive changes to this world – it can also usher in ethical issues. Big data analytics and artificial intelligence-based innovations raise significant concerns. The most obvious of these are posed by privacy issues and data protection, but those are far from the only ones. Data protection officers and scientists along with industry experts have identified eight groups of core norms and ethical values that are affected by big data applications. These values include privacy protection, equality and non-discrimination, autonomy, controlling one's own identity, transparency, solidarity, contextual integrity, property and copyright (Christen, Blumer, Hauser & Huppenbauer, 2019). The following paragraphs explain them in more detail.

1.8.1 Privacy

One of the main dangers of big data is the potential for privacy violations. Companies collecting and using personal data without the permission of the individuals in question would lead to violations of personal privacy (European Union, 2022). Potential Internet of Things (IoT) consumers are still concerned about IoT smart home devices because they fear the usage of their data (Lippett, 2022), for example, the potential overreach of private information collected by IoT vendors.

Big data-related privacy concern is a very common phenomenon. In fact, in three out of four cases in this study, the practitioner mentioned privacy and regulations. In Cases 1, 3 and 4, considering the privacy of users, the practitioners from the three cases could not use certain data sources, even though the data source could provide insights on a more accurate level. In addition, practitioners from Case 3 emphasised the importance of addressing privacy concerns, as the privacy regulatory measures are becoming stricter.

1.8.2 Equality and non-discrimination

Discrimination refers to treating certain people differently, especially by treating them worse than others based on gender, race, sexuality, etc. (Cambridge Dictionary, 2024a). Non-discrimination refers to the opposite, that is, treating people fairly, and is a basic legal principle (Christen et al., 2019). For example, gender-based discrimination exists in big data innovation. Search engines tend to display advertisements for higher paying jobs to more male users than female users (Lambrecht and Tucker, 2019). Another example is social scoring. Some marketplace lenders use algorithms to analyse personal data such as activities on social media, search behaviour, and devices used when surfing on the internet to decide whether to grant someone a loan (Packin, 2019). This can help some people who are not able to get a loan from classical bank systems to have access to loans but raises concerns about unjustified discrimination. A balance of privacy and discrimination violations and a company's right to prevent credit losses is required (Christen et al., 2019).

1.8.3 Autonomy

Autonomy refers to the ability to make one's own decisions without being controlled by external forces (Cambridge Dictionary, 2024b). The external forces that influence autonomy are often controlled by the central authorities. Big data makes it possible to not only monitor individuals but also influence their behaviour (Kappler, Schrape, Ulbricht & Weyer, 2018). Companies and authorities that manage large-scale sociotechnical systems like transport systems are capable of controlling those systems in real-time. They can do this by accessing real-time data from sensors, machines, and users. The following example illustrates the interplay between decentralised decision-making and central authority (Kappler et al., 2018).

Today, drivers who follow navigation instructions from their cars or mobile APPs transmit large amount of data including position, speed, and destination. These data are used by route guidance systems like Google Maps as input for algorithms to predict future states of the system and provide route guidance to users. In addition, traffic control centres use these data patterns to monitor traffic, predict traffic jams, and reduce CO2 emissions. This reflects centralised decision making. Under this system, the users are still free to follow the suggestions or make their own route choices. However, a central authority seeking to achieve broader societal goals might not always align with the individuals in the system. The drivers want a fast route while the authority seeks to optimise traffic objectives. This reflects the interplay between the two.

In addition, it also raises the question of whether users who get used to this kind of algorithm-based guidance are still able to make their own decisions without the guidance. Before big data, drivers made their route plan before starting the journey. Now, the real-time systems enable real-time planning, which comes very handy. Drivers simply follow the guidance and continuously update the route planning. Users may increasingly become more dependent on them and lose options when visiting new places.

1.8.4 Control the (digital) identity

It is challenging for people to control their own (digital) identity or information about themselves because of big data innovation (Christen et al., 2019). It is normal for people to adapt their digital identity to different situations. For example, a person's profile on a dating website is different from their profile on a business network. However, if big data innovation conducts ubiquitous surveillance of people's digital behaviour, it will be very challenging for a person to control and adjust their own digital identity (Christen et al., 2019).

1.8.5 Transparency

Transparency in business means offering stakeholders such as customers, partners and investors clear information about a company's business processes, products and services so that the stakeholder can make informed decisions (Christen et al., 2019). In terms of big data, transparency means that individuals know what data are collected from them and how they are used. However, this information is often disclosed in lengthy terms and conditions documents that are not comprehensible to laypeople and are in fact rarely read by users (Cakebread, 2017). In addition, the algorithms used for data analysis are sometimes "black boxes" that are hard to understand. This makes it very difficult to test and trust their outcomes (Christen et al., 2019). The good news is that data scientists are trying to break into the black box and understand how it works; for example, in image recognition, the understanding starts from recognising which neurons are identifying which objects in the image (Savage, 2022).

1.8.6 Solidarity

Solidarity refers to mutual support within a community. In modern life, solidarity involves providing financial aid for those in need, as anyone may end up facing hardship in life. Solidarity is rooted in utilising the costs-by-cause principle in all types of insurance. For example, someone who intentionally causes harm may limit the benefit of solidarity from others. With big data innovation, insurance costs-by-cause rules are enlarged. An insurance company could require the customer to adopt a certain diet to decrease certain health risks as a precondition for receiving solidarity benefits such as reimbursements from the insurance. This is also in conflict with making one's own choices. Fortunately, legal barriers have already been set in certain countries. In Switzerland, for example, insurance companies cannot refuse to exclude anyone from basic insurance because of their behaviour (Christen et al., 2019).

1.8.7 Contextual integrity

People expect different levels of integrity depending on the situation they are in; they might accept being treated differently in economic spheres, but not in health, legal or education matters (Lane, Stodden, Bender & Nissenbaum, 2014). For example, when people share personal health data for research purposes, they do so to help others to have medical care. However, if this data is used for economic reasons, such as to maximise profit or customise offer conditions, then it violates contextual integrity, since the original intention for sharing data was different. Big data makes it hard to detect violations of contextual integrity, because it is obtained from various sources, traded by data brokers, and used in complex algorithms or statistical models. It is difficult to keep track of where and how data are used, and when it would be appropriate to share personal information (Christen et al., 2019).

1.8.8 Property and copyrights

The economy has certain moral foundations such as property and copyright, which are also protected by law. However, big data raises debate regarding whether data falls under the legal protections. Using online services often generates or discloses personal data; in some cases, it is one of the sources of revenue for some companies. However, this raises questions about data ownership. Ethically, both data providers and companies that analyse these data should get a fair share of the profit. Companies often have freedom of choice when it comes to investment; they do so when expecting to earn a profit from it. However, that is not the case for users who provide data. They have fewer choices. The business model at the moment is that data providers get to use the related product or service for free, but alternatives may be needed in the future (Christen et al., 2019).

To sum up, the increasing use of smart devices and collection of big data for innovation has significant societal impact. Consumers share personal information to providers in exchange for products and services like real-time recommendations (Kappler et al., 2018). For companies, customer data represent a growing source of competitive advantage, and thus gaining customers' confidence is key. Transparency in data usage and offering fair value in return would build trust, while concealment loses customers (Morey, 2015). For society, the aggregated data can be used to influence individual and collective behaviour. This raise concerns that must be addressed with regulations (Kappler et al., 2018). Leading international companies like Apple and Google have significant power in managing data collection and distribution, and this strongly calls for regulation (Sharon, 2021). Future regulation should combine transnational legal instruments with constant monitoring of the societal effects of big data. Only when we have a clear understanding of the circumstances in which big data leads to undesired effects, will we be able to criticise and prevent them. To achieve this aim, social scientists, data scientists and engineers should join forces (Kappler et al., 2018). When using big data for innovation projects, the practitioners could ask ethical questions such as: “Does our product genuinely contribute to social welfare?” or “Are we potentially causing harm to the user?” or “Are we adequately involving users in the development of this new product?” or even “Do we have the right to use this resource?” These questions might help practitioners better guide and assess their work from an ethical point of view (Fabian & Robert Fabricant, 2014).

To summarise this chapter, big data is everywhere – and it comes from various sources in different formats. It can be used in all stages of the innovation process in multiple industries. Despite its benefits, it also has negative impacts on society. In this research, the focus is on leveraging big data for product innovation. The following section explains this in more detail.

1.9 Research motivation

The ability to generate valuable insights from big data could be an important innovation driver (Xu, Frankwick, & Ramirez, 2016). Accessing and using big data from a variety of sources allows firms to generate new ideas for innovation and better understand the needs of their consumers (Anderson, Potočnik, & Zhou, 2014). Integrating big data into a business can provide significant benefits to a company. The early adopters of big data analytics have gained a competitive advantage over their peers in the industry (Wegener & Sinha, 2013). Companies that use big data analytics are twice as likely to be in the top quartile in financial performance as their market peers; they make decisions five times faster than their competitors and are three times as likely to execute decisions (Pearson & Wegener, 2013). They invest heavily in collecting internal and external big data and apply big data analytics to beat competitors (Marshall et al., 2015). According to Forbes Insights’ strategic research survey of 316 executives worldwide (Forbes Insights, 2015), an increasing number of organisations are realising the value of utilising the vast amounts of data generated by the digital universe. A majority of respondents view big data and analytics as very important methods for achieving a competitive advantage. Approximately two-thirds of the respondents state that big data and analytics initiatives have contributed significantly to revenue growth.

However, in spite of heavy investments and great expectations, many companies have failed to improve their innovation outcomes (e.g. innovation competency) through big data (Johnson, Friend, & Lee, 2017). Some are wondering whether they will be able to gain valuable insights out of heterogeneous data from a variety of sources (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Gartner's report predicts that at least 80% of big data analytics (BDA) projects will fail to deliver business results and will never make it into production (Gartner, 2019). With the help of new technologies, it is now possible to collect more data than ever before. Yet, in every industry and every part of the world, senior leaders are wondering whether they are taking full advantage of the vast amounts of data they possess (LaValle et al., 2011) and how to achieve that (Ransbotham, Kiron & Prentice, 2016). Nearly half of the executives worldwide indicated that data-driven strategies are not universally accepted in their organisations, with the single greatest cultural challenge being adapting and refining a data-driven approach (Forbes Insights, 2015). Approximately 44% of them consider putting big data learning into action to be a major operational challenge.

Current studies in this field have found evidence of a positive relationship between various big data concepts and product innovation. However, the mechanism behind how this positive relationship is built is unclear. Which aspects of the big data concept influence which aspects of product innovation and in which way? This research project aims to fill this gap by exploring what factors influence big data-enabled digital product innovation and how they influence it. Chapter 2, the literature review, will explain the current study and the gap in more detail. Section 1.10 below describes the thesis structure in detail.

1.10 Thesis structure

This thesis consists of 10 chapters. Chapter 1 introduces the topic of this thesis. Chapter 2 presents the literature review and the research questions. Chapter 3 introduces the method. Chapters 4, 5, 6 and 7 present the results of four case studies exploring the answers to the research questions. Chapter 8 presents the results of a cross-case analysis to answer the research questions and build the theoretical framework to illustrate the mechanism. Chapter 9 discusses the major implications of this research. Finally, Chapter 10 presents the conclusion and summarises this research.

Chapter 1 introduces the topic of using big data in digital product innovation. It starts with the importance of innovation and how to manage innovation, then moves on to big data and the role of big data in the innovation process. The chapter ends with the research motivation.

Chapter 2 provides a review of the literature. It starts from the theoretical background of this study, which is the resource-based view and dynamic capabilities view, and then presents the definitions of the core concepts. The main body of the chapter focuses on the empirical research on the antecedents and consequences of using big data in innovation. This is followed by a summary of the research gaps and introduction of the research questions.

Chapter 3 describes the research method. It presents the case study design and interview set up. Then chapter then introduces the case backgrounds for the four cases and describes how the data analysis was conducted with the aim of answering the research questions.

Chapters 4, 5, 6, and 7 present the results of the four case studies. Each chapter starts with a case background. The chapter then presents the main findings of the case: the overview of the conceptual model of the antecedents and consequences of developing BDAC in innovation. This is followed by descriptions of how the result was arrived at from the qualitative data analysis.

Chapter 8 presents the cross-case analysis results. It identifies the internal and external influencing factors of BDAC, influencing factor of data variety, and influencing factors of innovation performance. It is explained how each of the factors has an influence.

Chapter 9 focuses on the discussion of this thesis. It starts with the main findings of the study, accompanied by the resulting theoretical framework. The theoretical and practical implications follow. The chapter ends with limitations and future research.

Chapter 10 is the conclusion chapter. It presents the summary of the previous chapters.

CHAPTER 2

Literature review

2.1 Introduction

The literature review presented in this chapter is a key component of this research project, providing an in-depth exploration of using big data in innovation. The phenomenon can be investigated by analysing the mechanism of developing big data analytics capabilities (BDAC) in product innovation by adopting the combination of a resource-based view (RBV) and dynamic capabilities (DC) as a theoretical lens. This literature review aims to identify a gap in using big data in innovation when seen through said theoretical lens.

The following part of the literature review starts by elaborating on the theoretical foundations adopted for this research, followed by defining the core constructs and context behind developing BDAC in product innovation. Finally, it presents the synthesis of empirical research on the antecedents and consequences of using big data in innovation.

2.2 Theoretical background

When it comes to the relationship between theory and practice, many people refer to a famous quote that is often attributed to Immanuel Kant: "Theory without practice is empty; practice without theory is blind." There is a symbiosis between theory and practice. This is especially true in research studies. Whether explicitly stated or not, such studies are guided by a theoretical perspective (Sadowski, 1993). In business studies and other social studies, theory and practice cannot flourish without each other. Thus, it is essential to build empirical business studies on solid theoretical foundations (McAvoy & Butler, 2018).

One foundation of this research is the firm's dynamic capabilities view (DCV). It is considered an extension of the classical RBV of the firm (Barney, 1991). Dynamic capabilities theory is derived from recognising that companies should build a competitive advantage in a rapidly changing market and technological environment rather than adapt or respond to it (Teece, Pisano & Shuen, 1997). Teece et al. (1997, p. 516) define dynamic capabilities as "the firm's ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments". According to a consensus in the literature, capabilities tend to perform differently in operation and may result in different levels of competitive advantages depending on various internal and external factors (Drnevich & Kriauciunas, 2011; Mikalef et al., 2019).

Since this research aims to identify the core resources that would form an environment that influences the usage of big data in innovation, the resource-based view (RBV) can serve as a suitable underlying theoretical foundation. For this reason, RBV is adopted as the theoretical foundation. According to the resource-based view, companies have access to different types of resources, which can be categorised as tangible resources, intangible resources, and human skills (Grant, 1991). This categorisation of resources has long been used in the IT capability literature (Bharadwaj, 2000). Combining RBV and DCV effectively explains how innovation capabilities are fostered within a company, providing a theoretically grounded connection between a firm's resources and innovation capability (Hadjimanolis, 2000). In addition, although RBV provides a solid foundation for identifying resources, these resources can lie dormant (Wu, Melnyk & Flynn, 2010). Resources do not necessarily reflect what a company can do. Instead, they reflect what a company can access (Größler & Grübner, 2006). Thus, this research project combines RBV with DCV because DCV enables examining the organisational capabilities that direct these resources to achieve competitive performance gains (Mikalef, Pappas, Krogstie & Giannakos, 2018).

2.3 Definitions of the core concepts

By combining RBV and DCV, this research views the phenomenon of using big data for product innovation as a company developing BDAC in the context of changing internal and external environments. Thus, BDAC, big data, and product innovation are the key concepts for this research because they are the elements of the phenomenon under investigation. The following section presents their definitions along with definitions of related concepts. These definitions are applied throughout this dissertation, from the literature review to the findings.

2.3.1 Big data analytics capabilities (BDAC)

The concept of big data analytics capabilities (BDAC) is defined as a company's ability to capture and analyse data through processes and structures to generate insights (Gupta & George, 2016; Mikalef, Boura, Lekakos & Krogstie, 2019; Wamba, Gunasekaran, Akter, Ren, Dubey & Childe, 2017). In this research project, BDAC is the core concept for investigation and analysis. BDAC is a combination of many resources and skills. Figure 2.1 presents the two different layers of constructs of BDAC. The first layer of constructs consists of tangible resources, intangible resources, and human skills. The second layer of constructs comprises the elements belonging to the first layer. For example, basic resources, data and technology belong to tangible resources because they are physical or measurable.

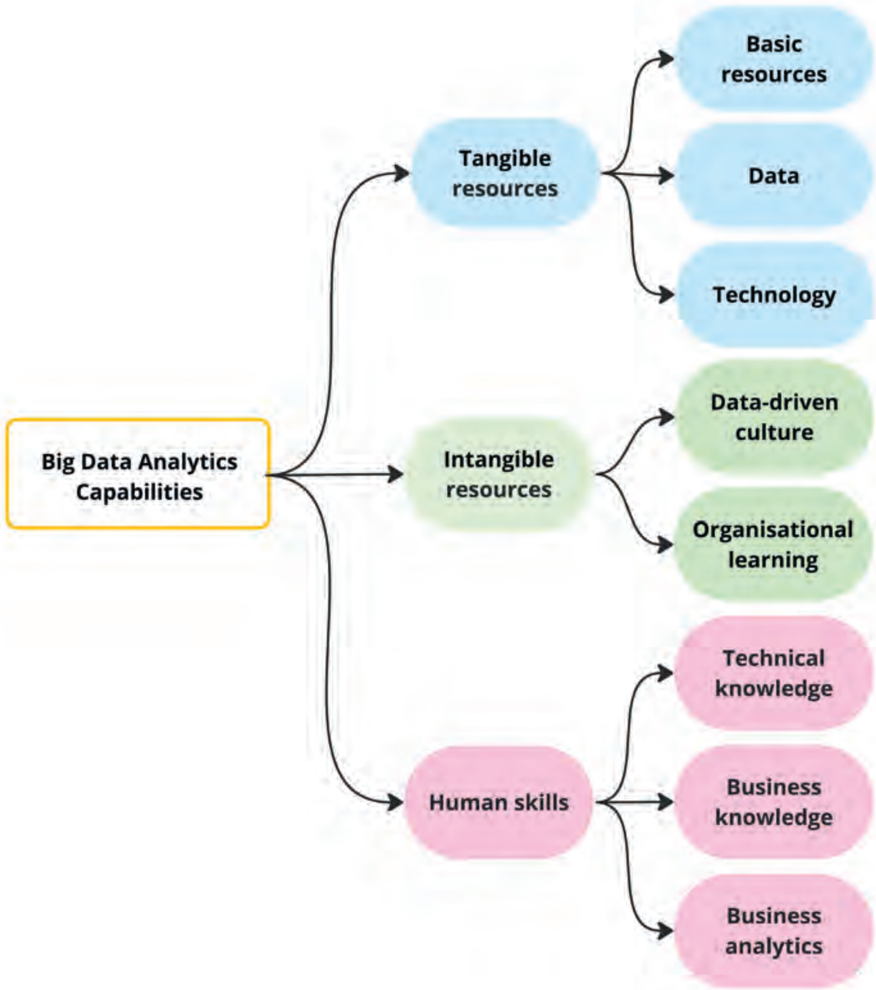


Figure 2.1 BDAC first-layer and second-layer constructs

Table 2.1 presents the definition of BDAC along with its first-layer and second-layer constructs (Mikalef et al., 2019). Among the first-layer constructs, tangible resources are physical or measurable; intangible resources are not physical and as such are more difficult to measure than tangible resources. Human skills highlight the many different work-related skills of employees, including those categorised as technical and managerial skills. The table also presents the definition of the second-layer constructs of BDAC. These definitions are applied in the rest of the dissertation, the research data analysis, findings, and discussion.

Table 2.1 The definition of BDAC and its first- and second-layer constructs (Mikalef, Pappas, Krogstie & Giannakos, 2018)

The definition of BDAC and its constructs	
Big data analytics capabilities (BDAC)	The company’s ability to effectively capture and analyse large and complex data to generate valuable insights, which is achieved by effectively leveraging its large and complex data, appropriate technology, and talent in the relevant field through processes and structures (Gupta & George, 2016; Mikalef et al., 2019; Wamba, Gunasekaran, Akter, Ren, Dubey & Childe, 2017).
Tangible resources	Assets used in a company's operations that have a physical presence and can be measured (Murphy, 2023).
Intangible resources	Assets used in a company's operations that don not have a physical presence in nature and difficult to measure (Murphy, 2023).
Human skills & knowledge	The skills that people have, including business skills, technical skills and business analytics
Basic resources	The essential assets that company need for its operation, including financial support and raw materials, etc.
Data	Information that is transmittable and storable (Etymology Dictionary, 1946).
Technology	The application of scientific knowledge for solving problems or achieve goals; it could be methods, systems, and devices (Collins, 2019).
Organisational learning	The collected endeavours of companies to leverage existing knowledge and explore new knowledge to confront the unpredictable conditions of the market (Teece, 2015).
Data-driven culture	The shared value, norms, and mindset that organisational members should leverage insights derived from data analysis to inform decision-making (McAfee & Brynjolfsson, 2012).
Technical knowledge	The expertise required to manage the technological elements to extract value from big data (Gupta & George, 2016). Eg., Programming and coding, data management, technical infrastructure management, etc.
Business knowledge	Understanding where to make efforts to gain insights from big data and recognising its value (Akter and Wamba, 2016).
Business analytics	The systematic thinking process of analysing data and gaining insights to inform and support decision-making by using quantitative and qualitative computational tools and methods (Power, Heavin, McDermott & Daly, 2018).

Table 2.2 below presents the characteristics or elements of tangible resources, including data, infrastructure, software and IS (Mikalef et al., 2018). Each resource has many different characteristics. For example, volume, variety and velocity can be used to describe data, and data also have other aspects – likewise, so do other tangible resources like infrastructure and software. These characteristics of the resources provide foundations for the further step of this research when analysing data and coding. The characteristics are used for coding purposes if the observed phenomenon matches them. If none of the characteristics match, new elements are proposed.

Table 2.2 Constructs of tangible resources

Tangible resources	Characteristics	Reference
Basic resources	e.g. financial resources	Mikalef, Boura, Lekakos & Krogstie, 2019
Data	accuracy, completeness, confidentiality, currency, reliability, security, timeliness, volume, variety, velocity	Brinkhues et al. (2014), Chae et al. (2014), Erevelles et al. (2016), Gupta and George (2016), Kamioka and Tapanainen (2014), Olszak (2014), Ren et al. (2016), Wamba et al. (2015), Phillips-Wren et al. (2015) and Vidgen et al. (2017)
Technology	e.g. software and hardware	Mikalef, Boura, Lekakos & Krogstie, 2019

Table 2.3 below presents the characteristics or elements of intangible resources, including organisational learning and data-driven culture (Mikalef, Pappas, Krogstie & Giannakos, 2018). Organisational learning include aspects such as generating, retaining, and transferring knowledge. Data-driven culture also has many elements; for example, data-driven culture is sometimes embodied in the support of top management for using data in business operations. The capability of managing big data analytics comes from the capability of managing these elements.

Table 2.3 Constructs of intangible resources

Intangible resources	Characteristics	Reference
Data-driven culture	fact-based culture, learning culture, top management support for data-driven projects, prioritising investments in data analysis	Davenport et al. (2001), Davenport (2013), Gupta and George (2016), Kamioka and Tapanainen (2014), Kiron et al. (2014), Kiron and Shockley (2011), Lamba and Dubey (2015), Olszak (2014), and Mikalef et al. (2017)
Organisational learning	generating, retaining, and transferring knowledge	Argote, 2012

Table 2.4 below presents the characteristics or elements of human skills and knowledge (Mikalef, Pappas, Krogstie & Giannakos, 2018). These include technical knowledge such as programming and non-technical knowledge such as business knowledge, relational knowledge, and business analytics. Business knowledge focuses more on the strategy side, while business analytics refers to a large variety of analytics activities such as predictive modelling and business reporting; relational knowledge is mainly about communication between people and team building in big data projects.

Table 2.4 Constructs of human skills and knowledge (Mikalef, Pappas, Krogstie & Giannakos, 2018)

Human skills and knowledge	Characteristics	Reference
Technical knowledge	Programming and coding, technical infrastructure management, data collection management, structured and unstructured data management	Akter et al. (2016), Elbashir et al. (2013), Garmaki et al. (2016), Gupta and George (2016), Kamioka & Tapanainen (2014), Olszak, (2014), and Mikalef et al. (2017)
Business knowledge	Business strategy, business processes, marketing knowledge, change management	Akter et al. (2016), Erevelles et al. (2016), Elbashir et al. (2013), Gupta and George (2016), Olszak (2014), Garmaki et al. (2016), and Olszak (2014)
Business analytics	Statistical analysis, forecasting, predictive modelling, model management, model optimisation, simulation, business reporting, data visualisation, web analytics, social media analytics, text mining, text analysis, audio analysis, video analysis	Davenport et al. (2001), Cao and Duan (2014b), Chae et al. (2014), Chen et al. (2016), Erevelles et al. (2016), Galbraith (2014), Garmaki et al. (2016), Lamba & Dubey (2015), LaValle et al. (2011), Olszak (2014) and Vidgen et al. (2017)

As can be seen, big data analytics capabilities include a wide variety of capabilities to leverage tangible and intangible resources along with the human skills available to the company. The development of BDAC during innovation projects is a very dynamic process. Any change in the specific elements of BDAC contributes to changes in BDAC as a whole. To further investigate the phenomenon of developing BDAC in innovation projects, it is necessary to understand the concept of innovation and big data. The following section discusses their definitions and elements in detail.

2.3.2 Innovation

Innovation is another key concept for this research, as the aim of using big data is to innovate new products. In both research and practice, *innovation* and *big data* are frequently used buzzwords, but different people understand them differently. To prevent misunderstandings and clearly frame this research project, it is necessary to define what is meant by both innovation and big data.

There are many definitions of innovation. OCED (2015) defines innovation as the implementation of a novel or significantly enhanced product (tangible product or intangible service), a process, a fresh marketing approach, or an innovative organisational strategy in business practices, workplace management, or external interactions.

In this PhD dissertation, the focus is on new or significantly enhanced products or services, specifically digital product innovations. First, let us look at product innovation. In the early 20th century, Schumpeter (1934) defined innovation as the practical implementation of ideas that result in the introduction of new goods or services or improvements in the offering of goods or services. In recent research, innovation is defined as the creation of value by using relevant knowledge and resources for the conversion of an idea into a new product, process, practice or improvements in an existing product, process, or practice (Varadarajan, 2018), developing a new product or enhancing an existing one in order to meet customers' needs in a novel way. Furthermore, in the industry, ISO TC 279 in the standard ISO 56000:2020 defines innovation as "a new or changed entity realising or redistributing value". The common elements in the definitions are newness, value (improvement, enhancement), and implementation or realisation.

Now, let us move on to service innovation. Although the concept of service innovation is widely used, only a few research papers have attempted to define it clearly. These definitions vary greatly among research papers. In the study by Witell (2016), 84 definitions of service innovation were identified. These definitions were categorised into three perspectives, namely, the assimilation perspective, the demarcation perspective, and the service-dominant logic perspective.

From an assimilation perspective, service innovation is considered to constitute a part of product innovation. As an example, Giannopoulou (2014) defines service innovation as a type of product innovation involving the introduction of a service that is new or significantly improved in terms of its characteristics or intended uses (Giannopoulou 2014).

From a demarcation perspective, service innovation can be defined independently as a unique service solution that includes new service concepts, customer interactions, value systems, revenue models, or technological service delivery systems (Breunig 2014).

Furthermore, under the service-dominant logic perspective, service innovation can be defined as the process of combining various resources to create novel resources that are beneficial (i.e. value experiencing) to some actors in a given context (Lusch & Nambisan, 2015). This definition of service innovation thereby sees service innovation as the redeployment of resources. The focus is on the value experienced by the customer instead of the output delivered by the companies.

Judging from these definitions of product and service innovation – under the assimilation perspective, the demarcation perspective, or the service-dominant logic perspective alike – the difference between product innovation and service innovation has become blurry. Service innovation can involve a type of innovation (Giannopoulou, 2014), technological service delivery systems (Breunig 2014), or a redeployment of resources. None of these definitions of service innovation conflict with product innovation.

Thus, in this PhD research project, service and product innovation are not differentiated. To keep consistency, the term product innovation is used here, with a specific focus on digital product innovation. To explain this, the dissertation starts by looking at the trend of digital transformation.

2.3.3 Digital transformation

While innovation propels organisations towards progress, digital transformation helps organisations harness the full potential of digital technologies. As innovation and digital transformation come together, businesses are pushed into a new era of growth and adaptability, where technology becomes a powerful enabler of change and success.

Digital transformation has emerged as an important phenomenon in strategic information systems research (Bharadwaj, El Sawy, Pavlou, Venkatraman, 2013; Matt, Hess & Benlian, 2015) as well as in practice (Tabrizi, Lam, Girard & Irvin, 2019; Westerman, Bonnet & McAfee, 2014). Based on a systematic review and semantic analysis, Vial (2019) conceptualised the definition of digital transformation as "a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies" (p. 121).

Looking at how the rankings of the five largest listed companies by market capitalisation have changed in recent years, it is evident that digital transformation may have disruptive effects on innovation (Bughin & van Zeebroeck, 2017; Cozzolino, Verona & Rothaermel, 2018). It is no secret that companies such as Apple, Microsoft, Amazon, Alphabet (Google), and Facebook (Meta) have leveraged digital technologies to develop a variety of innovative products and services (Appio, Frattini, Petruzzelli & Neirotti, 2021).

A word co-occurrence analysis (Waltman, van Eck & Noyons, 2010) has been conducted on research published regarding the intersection of digital transformation and innovation management during 2017-2020 (Appio et al., 2021). The results indicate that the concept of digital transformation is usually adopted in relationship with a wide range of recently introduced enabling technologies such as AI, industrial internet of things, big data, and smart products (Appio et al., 2021).

2.3.4 Digital innovation

Digital innovation comprises a cluster that is closely related to digital transformation (Appio et al., 2021). This is not surprising, because a major driving force behind digital transformation is digital innovation resulting from and involving experimental and exploratory use of digital technology (Drechsler, Gregory, Wagner & Tumbas, 2020). The content of this cluster revolves around the value of information technology for businesses, with an emphasis on analytical information processing capabilities (Brunswicker & Schecter, 2019; Saldanha, Mithas & Krishnan, 2017). Furthermore, this thematic cluster investigates the types of institutional arrangements that organisations use to embrace digital innovations and how these arrangements affect capabilities, processes, collaborations, and governance (Hinings, Gegenhuber & Greenwood, 2018; Svahn, Mathiassen & Lindgren, 2017; Verstegen, Houkes & Reymen, 2019).

In general, digital innovation refers to the use of digital technology during the process of innovation or as a result of innovation (Nambisan, Lyytinen, Majchrzak & Song, 2017). Digital innovation takes a product-centric perspective and involves combining physical and digital products to form new products (Lee & Berente, 2012; Yoo, Henfridsson & Lyytinen, 2010). Nambisan et al. (2017) conceptualise digital innovation as "the creation of (and consequent change in) market offerings, business processes, or models that result from the use of digital technology" (p. 224). Based on this definition of digital innovation, digital innovation management refers to "the practices, processes, and principles that underlie the effective orchestration of digital innovation" (Nambisan et al., 2017, p. 224). The abovementioned definition of digital innovation covers various innovation outcomes, such as new products, platforms, and services, new customer experiences and other value pathways.

2.3.5 Definition of big data based on its essential characteristics

As a buzzword used by both academics and the industry, the term big data has been defined in many ways by various scholars and practitioners. This section will define big data in the context of this research, along with the characteristics of big data, the technologies behind turning big data into insights, and the insights that could yield value for society.

In reviewing the definitions and main research themes in the big data-related literature, scholars have concluded that the core concept consists of the following four aspects (De Mauro et al., 2016):

- Volume, variety, and velocity are widely accepted characteristics of big data.
- Big data is an information asset by its very nature.
- To make effective use of this information asset, "technology" and "analytical

methods" are needed.

- Big data can be transformed into insights that are valuable to companies and society as a whole.

Combining existing definitions in the field and mainly adopting the definition from De Mauro et al., this research defines big data as:

> A high volume, velocity and variety of information assets that exceed the processability of conventional database software tools that require advanced technology and analytical methods to transform them into value (Beyer & Laney, 2012; De Mauro, et al., 2016; Dumbill, 2013; Manyika et al., 2011; Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012).

The dimensions of big data commonly refer to three characteristics: volume, variety, and velocity (McAfee & Brynjolfsson, 2012). Several authors have added other characteristics, such as veracity (Schroeck et al., 2012) and value (Dijcks, 2013). Volume, variety and velocity are the most commonly mentioned three characteristics, yet there are no universally agreed definitions for them. The definitions are contingent upon the sector, location and size of the firm, and may change over time (Gandomi & Haider, 2015).

Big data volume measures the magnitude of data (Gandomi & Haider, 2015). It is considered impractical to define specific thresholds for big data volume. What is considered "big" data today might not seem so big tomorrow because data storage capacity is increasing and more data can be captured (Gandomi & Haider, 2015). Boeing generates one terabyte of data per flight (Hollinger, 2015), and Facebook is home to 40 billion photos (Cukier, 2010). Data sizes were reported in terabytes (10^{12} bytes) and petabytes (10^{15} bytes) and now in zettabytes (10^{21}bytes) and yottabytes (10^{24} bytes), but in the future, the unit of choice will be brontobytes, which is 10^{27} bytes (Pence, 2014). As the first proposed characteristic, volume is only one aspect of big data. Other characteristics are also important when it comes to defining big data.

Big data variety refers to the different data formats and sources (Kaisler, Armour, Espinosa, and Money, 2013). Big data variety ranges from structured and semi-structured data to unstructured data (Hofacker, Malthouse, & Sultan, 2016; Sagiroglu & Sinanc, 2013). Structured data only accounts for a tiny part of all existing data today, constituting less than 5% (Cukier, 2010). According to Gandomi & Haider (2015), there are three types of data: structured, unstructured, and semi-structured.

Structured data can be stored in a predefined format; examples of structured data are spreadsheets or relational databases. Unstructured data are data that lack the structural organisation required by machines for analysis; thus, they are not stored in traditional relational database management systems. Unstructured data account for about 95% of all data, for example, texts, weblogs, GPS location information, sensor data, graphs, videos, audio data and other online data. Semi-structured data represent a relatively rare type; positioned between structured and unstructured data, there are no strict standards for semi-structured data. Semi-structured data refers to data that contain tags to separate data elements rather than conforming to fixed fields (Sagiroglu & Sinanc, 2013). A typical example of semi-structured data is Extensible Markup Language (XML). It is a textual language for exchanging data on the Web (unstructured text data) that have enough structure, in the form of user-defined data tags, to make them machine-readable, thus making it (XML) semi-structured (Gandomi & Haider, 2015).

Big data velocity refers to the speed at which data are generated, stored, and analysed (Gandomi & Haider, 2015). It also reflects the speed at which data are collected (George, Osinga, Lavie & Scott, 2016). Velocity can be easily associated with the speed data are pouring into most organisations (Dykes, 2017). Data natives such as Google and Facebook are undoubtedly generating data at a fast speed. Every second, Google processes more than 40,000 searches on average; that is 3.5 billion searches per day (Marr, 2021). Every minute, there are 510,000 comments posted and 293,000 statuses updated on Facebook (Marr, 2021). Besides these born-digital companies, retail giant Walmart generates more than one million customer transactions every hour (Cukier, 2010). Traditional processing of data in batches cannot meet the increasing need for speed. Big data technologies enable real-time processing to generate insights from the huge amount of “perishable” data (Gandomi, & Haider, 2015).

Big data veracity refers to the reliability level of certain data sources (Gandomi, & Haider, 2015; Schroeck et al., 2012). “Veracity” has been included due to the importance of authenticity (White, 2012). The “veracity” of data sources represents the reliability and trustworthiness of the sources from which the data are gathered (White, 2012). A high level of data reliability is not only a requirement but also a challenge, and even the best data-cleaning technology cannot remove the uncertainty posed by some kinds of data (Schroeck et al., 2012), such as the uncertainty inherent in, for example, weather conditions, economic factors, and human sentiment and truthfulness. Even so, such data still yield valuable insights. Uncertainty can be managed in many ways, such as by combining different data sources and adopting advanced mathematics (Schroeck et al., 2012).

Big data value refers to the economic value that can be extracted from data, which varies significantly for different data sources (Dijcks, 2013). Value is often hidden in data (Gandomi, & Haider, 2015). High value can be obtained by identifying the valuable data and then transforming and analysing it (Dijcks, 2013). Dijcks (2013) defines big data with characteristics including volume, velocity, variety, and also "value". Value refers to its transactional, strategic, and informational benefits (Wamba et al., 2015; Wixom, Yen & Relich, 2013).

Big data is not necessarily new, and many of the data sources already exist (Gandomi & Haider, 2015). Organisations have been extracting unstructured data from internal sources, such as sensor data, and external sources, such as social media (Gandomi & Haider, 2015). Innovation mainly originates from data management technologies and analytics, which is the next concept that will be introduced here.

Data analytics

> *"Information is the oil of the 21st century, and analytics is the combustion engine."*
> – Peter Sondergaard, Executive Vice President, Gartner Research

Data analytics can be categorised into three types: descriptive, predictive, and prescriptive. This categorisation of analytics types has been widely accepted in both academia and practice (Bekker, 2022; Cote, 2021; Lepenioti, Bousdekis, Apostolou & Mentzas, 2020; Wang, Gunasekaran, Ngai & Papadopoulos, 2016). Descriptive analysis focuses on the current state and mainly answers questions regarding: "What happened?"; diagnostics analytics answers questions such as: "Why did it happen?"; predictive analytics answers: "What will happen"; and finally, prescriptive analytics focuses on "How can we make it happen?" (Deshpande, Sharma & Peddoju, 2019). In this research, the four cases mainly use big data for predictive analytics and sometimes descriptive analytics, and both types of analytics serve the main function of the product.

Descriptive analytics

Descriptive analytics leverage past or current data to investigate and reveal what happened or is happening (Deka, 2014; El Morr & Ali-Hassan, 2019), and possibly prepare the data for the next step of the analysis (Deshpande, Sharma & Peddoju, 2019). The wide range of data helps to acquire a comprehensive review so that managers can make informed strategic business decisions (Frankenfield, 2020). The output of descriptive analytics is a group of indicators of past or current performance. The indicators assist in understanding both success and failure to improve decision-making based on the lessons learnt (Deka, 2014; El Morr & Ali-Hassan, 2019;). One example is analysing monthly income per product group to facilitate deciding on which product categories to focus on (Bekker, 2022). Another example is identifying seasonal sales surges in products (Cote, 2021).

Predictive analytics

Predictive analytics leverage historical data to measure future probabilities, identify patterns and predict trends (Deka, 2014; Deshpande et al., 2019). Such analytics use statistical models and forecast techniques to forecast the future (Deshpande et al., 2019). Common predictive analytics modelling tasks include classification (mainly decision tree-based), clustering, association, divergence detection (finding deviations), and web mining (Deka, 2014). There is an expectation that this will reduce risks, enhance operations, and increase revenue (SAS, n.d.). Due to its positive potential for resolving complex problems and revealing new opportunities, many organisations are optimistic about predictive analytics (Sharma, Sharma, Purohit, Rout & Sharma, 2022). The applications of predictive analytics are very broad (Gandomi et al., 2015). The business applications include risk assessment, sales forecasting, and customer conversion prediction (Gibson, 2018). In this research project, predictive analytics is the dominant type of data analytics adopted by the case study companies to realise the function of their products. For example, weight sensor data are used to predict how crowded a train or metro will be (Case 1).

Prescriptive analytics

The purpose of predictive analytics is to forecast what might happen in the future. Prescriptive analytics takes one step further, aiming to suggest actions that may be taken in the future (El Morr & Ali-Hassan, 2019). A prescriptive analytics model determines the optimal solution or outcome based on the available parameters (Deshpande et al., 2019). The adoption of prescriptive analytics can serve as a benchmark for measuring the analytics maturity of an organisation (Deka, 2014). Applications of prescriptive analytics include, for example, evaluating and innovating in business operations and achieving business goals while balancing constraints (Deka, 2014).

The above section focuses on the definition of the key concepts of this research project. While using big data for innovation is not an entirely new phenomenon, the following section discusses the current empirical studies investigating this phenomenon. It identifies several gaps in the current field.

2.4 Empirical research on the antecedents and consequences of using big data in innovation

This doctoral project answers the main research question: How do companies develop big data analytics capabilities in digital product innovation? Before conducting empirical investigations, this research will first explore what empirical studies in the field have already discovered. The existing empirical studies can be separated into quantitative and qualitative studies within the context of using big data in innovation. The quantitative studies examine the relationships between the antecedents, the big data concepts and product innovation, testing whether the relationship is positive, negative, or not prominent. The literature is separated by different big data concepts – BDAC, big data characteristics, and applying big data analytics. The qualitative studies explore how the relationship between big data usage and innovation has developed and explain the formation of the relationship. Figure 2.2 below presents the structure of empirical research on the antecedents and consequences of using big data in product innovation.

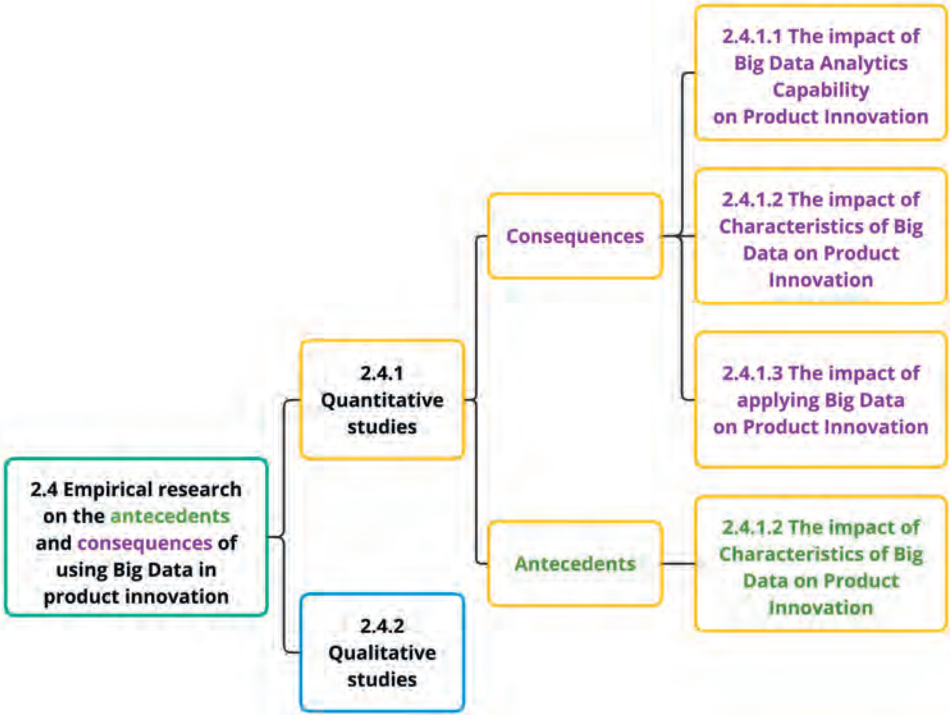


Figure 2.2 The structure of empirical research on the antecedents and consequences of using big data in product innovation

2

2.4.1 Quantitative studies

Current empirical studies on the antecedents and consequences are reviewed in this section. More studies focus on consequences than antecedents. This section first discusses the current findings about consequences because they dominate the field, and then the findings about antecedents follow.

The existing studies investigate using big data in innovation through various theoretical lenses, for example, the resource-based view, dynamic capabilities view, and organisational learning. Based on these different theoretical lenses, different terminologies or concepts are used: (1) big data analytics capabilities, (2) characteristics of big data, and (3) applying big data. The meanings and constructs of these concepts differ from each other, which makes it difficult to compile the results of the corresponding empirical studies. Thus, this review separates the literature streams by the terminologies used in these studies. Figure 2.2 presents the structure.

Current quantitative studies on the consequences of using big data in product innovation

Despite using different terminologies and theoretical foundations, the current studies show that the relationship between big data and its consequences on innovation is positive. To examine this relationship, most academic researchers follow a quantitative approach at the company level. These studies examine the relationship between different big data concepts and their effect on different aspects of innovation.

2.4.1.1 The impact of big data analytics capabilities on product innovation

BDAC is a very inclusive concept. Rooted in the resource-based view, BDAC is defined as the company's ability to capture and analyse data to generate insights by effectively leveraging its data, technology and talent through processes and structures (Gupta & George, 2016; Mikalef, Boura, Lekakos & Krogstie, 2019; Wamba, Gunasekaran, Akter, Ren, Dubey & Childe, 2017). BDAC is formed when a company effectively deploys its big data-related tangible, intangible resources and human skills (Mikalef et al., 2019). Tables 2.1, 2.2, 2.3, and 2.4 in Section 2.2 above present the definitions of BDAC and its first layer and second layer constructs.

Table 2.5 below summarises the studies that examine the relationship between big data analytics capabilities (BDAC) and innovation. One study found that BDAC positively and indirectly affected incremental and radical innovation capabilities (Mikalef et al., 2019). A firm's dynamic capabilities mediate the effect. A later study in 2020 found that BDAC's positive effect on incremental and radical innovation capabilities was indirect, mediated by information governance (Mikalef, Boura, Lekakos & Krogstie, 2020). Thus, when a company strengthens its BDAC through mediators, it can boost its radical and incremental innovation capabilities. The effect is mediated by information governance or dynamic capabilities, depending on the aspect taken to examine the phenomenon. Research also found that BDAC positively and significantly impacts new product development (NPD) (Dubey, Bryde, Graham, Foropon, Kumari & Gupta, 2021), specifically (a) increasing the introduction of new products, (b) expanding product range, (c) facilitating entry into new markets, and (d) fostering entry into new technology fields.

Take a closer look at the studies in Table 2.5 below. According to Mikalef et al. (2019), BDAC can enhance innovation capabilities and the effect is mediated by dynamic capabilities. In addition to the mediator, there are also different moderators influencing the effect of BDAC on innovation. Table 2.6 presents the definitions of the mediators, moderators, and the outcomes in this study. The effect of BDAC on innovation is mediated by a group of capabilities called dynamic capabilities. These capabilities are aimed at adapting to rapidly changing environments. Apart from mediators, there are also three moderators that cover the three aspects of environmental uncertainties, namely, dynamism, heterogeneity, and hostility. Dynamism focuses on uncertainty from the market, heterogeneity focuses on uncertainty from customers, and hostility focuses on uncertainty from competitors. The effect of big data on innovation is measured by innovation capability, which distinguishes between radical and incremental innovation capabilities. While *incremental* refers to the ability to expand existing products, *radical* refers to aiming at exploring new ones. The results suggest that the effect of BDAC on dynamic capability is enhanced in environments with high heterogeneity. This means that under complex and changing market conditions, BDAC can be built by managers to sense and transform the way their businesses operate to adapt to the situation. Under high dynamism, the effect of dynamic capabilities on incremental innovation capabilities is accelerated.

Table 2.5 Quantitative studies examining the relationship between big data analytics capabilities (BDAC) and innovation

Quantitative studies examining the relationship between big data analytics capabilities (BDAC) and innovation			
Study	Mikalef, Boura, Lekakos & Krogstie (2019)	Mikalef, Boura, Lekakos & Krogstie (2020)	Dubey, Bryde, Graham, Foropon, Kumari & Gupta (2021)
Title	Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment	The role of information governance in big data analytics-driven innovation	The role of alliance management, big data analytics and information visibility on new-product development capabilities
Big data aspect	Big data analytics capabilities (BDAC): 1. Tangible sources 2. Human skills 3. Intangible sources	Big data analytics capabilities (BDAC): 1. Tangible sources 2. Human skills 3. Intangible sources	Big data analytics capabilities (BDAC) (Gupta & George, 2016): 1. Tangible resources 2. Human skills 3. Technical skills 4. Data-driven culture 5. Organisational learning
Antecedent	N/A	N/A	Alliance management capability (AMC)
Mediator/ Moderator	Mediator: dynamic capabilities Moderator: dynamism; heterogeneity;	Moderator: information governance	Information visibility (IV)
Consequence measurement	1. incremental innovation capability 2. radical innovation capability	1. incremental innovation capability 2. radical innovation capability	New product development (NPD) market performance & financial performance

Table 2.5 Continued.

Quantitative studies examining the relationship between big data analytics capabilities (BDAC) and innovation			
Finding	1. BDAC can lead to enhanced incremental and radical innovation capabilities by affecting the underlying processes of a firm's dynamic capabilities. 2. dynamic capabilities fully mediate the effect of BDAC on incremental and radical innovation capabilities. 3. under high heterogeneity, the effect of BDAC on dynamic capabilities is amplified. 4. when environmental dynamism is increased, the effect of dynamic capabilities on incremental innovation capabilities is accelerated.	1. BDAC has a positive effect on incremental and radical innovation capabilities. 2. information governance may not have any substantial influence in amplifying insight into incremental innovative capabilities, but it plays an important part in accelerating the formation of a firm's radical innovative capabilities. 3. under conditions of high environmental dynamism, the effects of information governance and BDACs are significantly amplified.	1. AMC has a positive and significant impact on NPD and BDAC. 2. BDAC has a positive and significant impact on NPD. 3. the function and role of IV positively moderates the association between AMC and NPD; and between BDAC and NPD. 4. NPD has a positive and significant effect on the market performance and financial performance.
Data collection	Completed questionnaires from 175 CEOs and IT managers working in multiple industries from Greece.	Completed questionnaires from 175 CEOs and IT managers working in multiple industries from Greece.	Completed questionnaires from information, R&D and operation managers in 219 firms from the Indian auto components sector.
Which level	Company	Company	Company
Theoretical base	Resource-based view (RBV) Dynamic capabilities view	Resource-based view (RBV)	Dynamic capabilities view

Table 2.6 Definitions of mediators, moderators, and outcomes in the first stream of literature

Definitions of mediators, moderators, and outcomes in the first stream of literature	
Mediator: Dynamic capabilities	Ability to integrate, build, and reconfigure internal and external competencies in response or adapt to rapid changes in the environment. It includes the sensing, coordinating, learning, integrating, and reconfiguring processes.
Moderator: Dynamism	"The rate and unpredictability of environmental change" (e.g. product obsolescence, technology change, competitors' behaviour) (Newkirk & Lederer, 2006, p.482).
Moderator: Heterogeneity	"Complexity and diversity of external factors", such as "diversity in customers' buying habits, diversity in the nature of competition, and diversity in product lines." (Newkirk & Lederer, 2006, p.483).
Moderator: Hostility	"The availability of resources and the degree of competition in the external environment." (Newkirk & Lederer, 2006, P.483).
Outcome: Incremental innovation capability	The firm's ability to expand its existing product lines or services.
Outcome: Radical innovation capability	The firm's ability to explore new product lines or services.

In summary, BDAC is a very complex concept that has first-order and second-order constructs, with the first being tangible, intangible resources and human skills, and the second being the elements of the first-order constructs. There is an agreement in the existing literature that BDAC has a direct positive effect on innovation, either innovation capability or performance. However, a number of questions remain unclear regarding the effects of the specific constructs of BDAC on using big data in innovation. For example, how does data-driven culture influence the usage of big data in innovation, and how does technology influence the usage of big data in innovation? This research project will explore these questions.

2.4.1.2 The impact of characteristics of big data on product innovation

While some studies focus on the complex concept of BDAC, others take a different perspective on the phenomenon and focus on the characteristics of big data instead (Johnson, Friend & Lee, 2017; Ghasemaghaei & Calic, 2019; Ghasemaghaei & Calic, 2020). The characteristics of big data include volume, variety and velocity (the 3Vs); efficiency and efficacy are the consequences of the impact of the 3Vs. Table 2.7 presents the definitions of the concepts mentioned in this stream of literature.

Table 2.7 Definitions of concepts in the second stream of literature

Definitions of mediators, moderators, and outcomes in the second stream of literature	
Volume	"Big data volume measures the magnitude of data available to an organisation" (Johnson et al., 2017, p.644).
Variety	Big data variety is a measure of the richness of the data representation (Kaisler, Armour, Espinosa, & Money, 2013) available to the firm, making big data even more expansive (Sagiroglu & Sinanc, 2013).
Velocity	Big data velocity describes the speed at which the firm processes and analyses customer data.
Efficacy	"The degree of success of an innovation" (Alegre & Chiva, 2008, p. 317).
Efficiency	"The effort made to achieve that degree of success" (Alegre & Chiva, 2008, p.317).

Table 2.8 below summarises these studies that examine the effect of the 3Vs on innovation. The following paragraph explains the findings from these studies in detail. Johnson et al. (2017) examined the relationship between the 3Vs (volume, variety, velocity) and new product revenue. The findings indicate that big data facilitates new product revenue, but the effect is moderated by customer turbulence, which reflects the changing demands of customers. The effect of customer turbulence attenuates the effect of big data volume on NPR but accentuates the effect of big data velocity on NPR and has no moderating effect on big data variety and NPR. This means that the speed of processing data needs to be increased in order to keep up with the constantly changing demands of customers. The insights gained from analysing a high volume of data would be hurt most by changing demands. However, the benefit of using a greater variety of data to generate valuable insights would not be harmed by changing demands from customers. The study highlights the importance of data variety and velocity in the process of generating valuable insights reflecting customer demands.

Similarly, Ghasemaghaei & Calic (2019) also focus on how big data can bring valuable insights and further investigate how these insights then influence company competency. The results show that all big data characteristics have a significant impact on data-driven insights, and data-driven insights fully mediate the effect of big data on innovation competency (Ghasemaghaei & Calic, 2019). One year later, the researchers tested the direct impact of the 3Vs on innovation performance (Ghasemaghaei & Calic, 2020). In their study, innovation performance is measured by efficiency and efficacy. Efficiency measures the time and effort expended to achieve a degree of success (Alegre & Chiva, 2008). Efficacy measures the extent to which an innovation is successful (Alegre & Chiva, 2008). The results show that big data variety and velocity significantly impact innovation performance (efficacy & efficiency), while volume does not significantly impact innovation performance (Ghasemaghaei & Calic, 2020). This result is in line with the previous study by the

Table 2.8 Quantitative studies examining the relationship between the characteristics of big data and innovation

Quantitative studies examining the relationship between the characteristics of big data and innovation.			
Author	Johnson, Friend & Lee (2017)	Ghasemaghaei & Calic (2019)	Ghasemaghaei & Calic (2020)
Title	Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process	Does big data enhance firm innovation competency? The mediating role of data-driven insights	Assessing the impact of big data on firm innovation performance: Big data is not always better data
Big data aspect	The 3Vs of big data usage: 1. volume 2. variety 3. velocity	Big data characteristics: 1. volume 2. variety 3. velocity 4. veracity	Big data main characteristics: 1. volume 2. variety 3. velocity
Antecedent	Exploration/exploitation orientations	N/A	N/A
Mediator/ Moderator	Mediator: Customer turbulence	Mediator: Descriptive/predictive/prescriptive data-driven insight	N/A
Consequence Measurement	New product revenue (NPR) at the business unit level.	Innovation competency: 1. exploitation competency 2. exploration competency	Innovation performance: 1. efficiency 2. efficacy
Findings	1. An exploration orientation has a positive effect on all three dimensions of a firm's big data usage – volume, variety, velocity – while an exploitation orientation has no effect on any of them (discussion in 2.3.1.4). 2. Customer turbulence <u>attenuates</u> (-) the effect of big data volume on NPR but accentuates (+) the effect of big data velocity on NPR, and has no moderating effect on big data variety and NPR.	1. Big data variety, velocity, and veracity have a significant impact on all types of data-driven insight, including descriptive, predictive, and prescriptive insights. On the other hand, the volume of big data does not have a significant impact on any type of data-driven insight. 2. Descriptive and predictive insight marginally impacts exploitation competency but does not have a significant impact on exploration competency. 3. Regarding the impact of the characteristics on innovation competency, the result shows that data-driven insight fully mediates the effect of big data on innovation competency.	1. Big data variety and velocity significantly impact innovation performance (efficacy & efficiency). <u>2. Volume does NOT significantly impact innovation performance.</u> 3. Compared to other firms, firms that have not been careful to quickly integrate large amounts of various types of data cannot considerably increase their financial returns. 4. To successfully increase consumer satisfaction, firms need to not only integrate large amounts of data, but also process different types of data (e.g. social media, images, and pictures) in real time.

Table 2.8 Continued.

Quantitative studies examining the relationship between the characteristics of big data and innovation.			
Method	261 surveys of managers reporting on their business unit's NPD processes and big data usage, in multiple industries.	280 surveys are collected from executive managers, vice presidents, middle-level managers and business unit or department managers in multiple industries answering questions regarding the impact of big data on firm outcomes.	Survey of 239 top and middle-level managers in the US in manufacturing, services, utilities and finance.
Which level	Business unit level	Company	Company
Theoretical base	Dynamic (market) capabilities theory	Gestalt insight learning theory; organisational learning theory	Organisational learning theory (grounded in RBV); resource-based view (RBV)

same scholars (Ghasemaghaei & Calic, 2019) in which they tested the impacts of big data characteristics on data-driven insights. Big data variety, velocity, and veracity have a significant impact on all types of data-driven insight, while volume does not have a significant impact on any type of data-driven insight.

The results from this stream of literature indicate that to generate data-driven insights and obtain value from them, companies need to use various sources of data in a time-efficient way (Johnson et al. (2017); Ghasemaghaei & Calic, 2019; Ghasemaghaei & Calic, 2020); simply increasing the amount of data from the same source is not helpful. The studies also open an interesting gap – the further exploration of why big data variety and velocity matter for innovation performance, whereas volume does not. To understand this, it is necessary to know how data variety and velocity influence innovation performance. For example, regarding data variety, in which way does the variety of data make a difference? Regarding velocity, which elements of velocity matter? Is that data collection or processing? How does speeding up these activities contribute to better innovation outcomes?

2.4.1.3 The impact of applying big data on product innovation

While the previous two groups of scholars focus on BDAC and big data characteristics, another group of researchers has explored the relationship between using big data analytics and its effect on product innovation outcomes. Multiple studies focus on big data usage and its impact on different aspects of innovation consequences. Table 2.9 summarises the findings from these studies, focusing on three different consequence measurements, each with one study as an example. The studies are all based on more than 100 questionnaires collected from practitioners from multiple industries. The study suggests that the use of big data analytics is associated with a higher innovation propensity and intensity (Niebel, Rasel & Viete, 2019). As companies use big data in their business operations, they are more inclined to innovate new products and also generate more sales with new products and services. The usage of big data analytics (BDA) is boosting innovation performance through the mediator of agility, and the data-driven culture and team sophistication could moderate this effect. The finding is valid regardless of industry and firm size (ZareRavasan, 2021). Effective use of data aggregation and analysis tools has also been found to directly affect new product success (Shirazi, Tseng, Adegbite, Hajli & Rouhani, 2021).

Table 2.9 Quantitative studies examining the relationship between applying big data and innovation

Quantitative studies examining the relationship between applying big data and innovation			
Author	Niebel, Rasel & Viete (2019)	ZareRavasan (2021)	Shirazi, Tseng, Adegbite, Hajli & Rouhani (2021)
Title	Big data – big gains? Understanding the link between big data analytics and innovation	Boosting innovation performance through big data analytics: An empirical investigation on the role of firm agility	New product success through big data analytics: empirical evidence from Iran
Big data aspect	Application of big data Analytics	Big data usage	effective use of data aggregation/analysis tools
Antecedent	n/a	n/a	effective use of data aggregation/analysis tools
Mediator/ Moderator	n/a	Mediator: sensing agility, decision making agility, acting agility. Moderator: data-driven culture, BDA team sophistication	(CAG) customer agility, (NPS) new product success
Consequence measurement	1. innovation propensity 2. innovation intensity	Innovation performance	New product success (NPS)
Findings	1. The use of big data analytics is associated with a higher innovation propensity and a higher innovation intensity. 2. firms' general ICT intensity measured by the share of employees working predominantly with PCs is not significantly related to innovation propensity.	1. BDA use is positively associated with sensing agility and innovation performance. 2. Sensing agility is positively associated with decision-making agility. 3. Decision-making agility is positively associated with acting agility. 4. Acting agility is positively associated with innovation performance. 5. BDA team sophistication moderates the link between BDA use and sensing agility. 6. Data-driven culture moderates the link between sensing agility and decision-making agility.	1. Effective use of data aggregation tools has a direct effect on CAG. 2. Effective use of data aggregation tools is directly linked to NPS. 3. Effective use of data analysis tools has a direct effect on CAG. 4. The effective use of data analysis tools has a direct effect on NPS. 5. CAG has a direct effect on NPS. 6. PIP is directly linked to NPS.

Table 2.9 Continued.

Quantitative studies examining the relationship between applying big data and innovation			
Method	The board of management or the head of the IT department from 4,400 (manufacturing and services) firms were interviewed about their ICT usage. The respondents are from industries (%): manufacture 51.88, retail 5.84, wholesale 4.77, services 37.52.	185 valid questionnaires returned from chief information officers (CIOs) from a wide range of industries in Iran.	120 valid questionnaires returned from decision makers from top 500 Iranian companies from various industries.
Which level	Company	Company	Company
Theoretical base	Knowledge production framework	Dynamic capabilities (DC) theory; firm agility	Resource-based theory (RBT); capability-building view

Other study also supports the claim that big data facilitates product innovation, in terms of product newness and the amount and speed of introducing new products (Tunc-Abubakar, Kalkan & Abubakar, 2022). Big data analysis (BDA), along with traditional marketing analysis (TMA), and big data system quality (BDSQ) are significant determinants of new product development (NPD) success (Aljumah, Nusei & Alam, 2021). To sum up, using big data analytics has a positive impact on fostering a company's intention to innovate, to innovate more often, and to innovate faster; it boosts innovation performance and new product success, for example, in terms of increasing sales of new products.

2.4.1.4 The impact of antecedents on using big data

In terms of antecedents of big data, there are fewer studies in the current literature. Some of the studies mentioned above also examined the antecedents. Johnson, Friend & Lee (2017) examined the effect of company orientations on big data usage. They found that an exploration orientation has a positive effect on all three dimensions of a firm's big data usage – volume, variety, and velocity. In contrast, an exploitation orientation does not affect any of them (Johnson et al., 2017). This means that firms searching for new knowledge to foster radical innovation would use a greater number of sources of data to generate insights quicker for radical innovation compared with firms that aim at exploiting existing knowledge and skills for incremental innovation.

The other study came from Kwon, Lee & Shin (2014). The scholars examined the antecedents of the acquisition intention of big data analytics. Their result suggests that the firm's competence in maintaining the quality of corporate data can positively affect its intention to engage in big data analytics. In addition, if a company perceives that using external data sources will yield high benefits, then it has a greater possibility to adopt big data analytics in the future.

Another study mentioned above is by Mikalef et al. (2019). According to this study, different resources (tangible and intangible) are seen as the construct of BDAC rather than as antecedents existing separately from BDAC. However, there are different layers of elements that build up BDAC. The relationships between each factor among the elements remain unclear.

To sum up, within the literature exploring antecedents of using big data in innovation, the focus is scattered. Some have suggested that big data usage is driven by exploring new knowledge and skills rather than by exploiting existing ones (Johnson et al., 2017). Some have revealed that if companies have competence in maintaining the quality of their data, they would then intend to use big data analytics. There is a gap in exploring more details about which kinds of knowledge, skills, competence, or resources are related to big data usage and in which ways. In addition, if there are relationships among these resources, how do these factors influence each other?

2.4.2 Qualitative studies

The studies mentioned above examine the positive relationship between big data and innovation through quantitative research methods. To dive deeper into exploring what kind of positive relationship big data has with innovation, other studies offer various perspectives and use qualitative methods.

In the current literature, the line between product and service is already either very blurry or does not exist (Lusch & Nambisan, 2015). Thus, while this section focuses on service innovation, the term is considered to be synonymous with product innovation. Several qualitative studies have investigated service innovation with service-dominant logic (Lehrer, Wieneke, vom Brocke, Jung & Seidel, 2018; Troilo, De Luca & Guenzi, 2017; Chester Goduscheit & Faullant, 2018). Table 2.10 below presents these studies. Under service-dominant logic (SDL), the distinction between product and service is obsolete, and product innovations are also service innovations (Lusch & Nambisan, 2015). The focus is on the value generated from the use of the product-service offerings (Chester Goduscheit et al., 2018).

The first study in this table comes from Lehrer et al. (2018). It focuses on service individualisation and adopts the theoretical lens of combining service-dominant logic together with materiality and affordances. This study presents a theoretical model explaining that the material features of big data analytics technologies (including data sourcing, storage, analytics and exploitation features) afford service automation and human-material service practices, which then enable service innovation.

The second study is from Chester Goduscheit & Faullant (2018); it also adopts the theoretical lens of SDL and investigates how different configurations of conditions could yield a radical service outcome in B2B manufacturing SMEs in Denmark. The finding suggests that while digitalisation is the core condition for radical service innovation, it is not sufficient. The different configurations of conditions could enable radical innovation, with some serving as the core condition, combined with other peripheral conditions. These conditions include a network of actors (collaboration within a broader value chain), resource density (mobilising sufficient resources to create value) and resource integration (engagement with customers in co-creation of services).

Finally, Troilo et al., (2017) focus instead on service innovation aspects and data-rich environments. These service innovation aspects include service concept innovation, customer experience innovation, and service process innovation. Their study explored the process through which data-rich environments link to service innovation in companies. The research revealed that in a data-rich environment, technological enablers could pave the way for service innovation through data density processes. The data-rich environment is made up of internal and external data-rich environments. The external data-rich environment includes volume, variety, and velocity. The internal data-rich environment includes data management and analytics technologies. The technological enablers are made up of data management and analytics technologies. The data density processes include pattern spotting, real-time decisions, and synergistic exploration.

To sum up, current qualitative studies have taken two different angles to explore the mechanism of service innovation (service concept innovation, customer experience innovation, and service process innovation), either from the angle of the four main axioms of SDL (actor value network, resource liquefaction, resource density, resource integration) (Chester et al., 2018) or the perspective of data management and analytics technologies (Troilo et al., 2017). In addition, the researchers also proposed a framework to describe the mechanism behind how big data analytics can bring the outcome of service individualisation from data features through the mediator of actions. Within this stream of literature, there is a gap in exploring how different resources interact with each other and the mechanism behind them to reach service or product innovation. In addition to the current study on service individualisation, a gap exists in exploring another aspect of service or product innovation, such as the mechanism behind influencing performance by big data.

Looking in greater depth at specific types of big data, the current literature illustrates the usage of social media and use phase data in new product development. Bashir, Papamichail & Malik (2017) investigated the use of social media as a source of information for NPD. The researchers used a case study to illustrate the usage. Their findings reveal that social media tools and online discussion forums play a significant role in obtaining customers' information for NPD projects for purposes such as better sensing the potential opportunities for their products in new markets or enhancing their existing products. Social media data help firms to know how their products and their competitors' products are perceived in different markets and which aspects of the products require further improvement. However, social media data are used informally, as side information, whereas companies rely more on their trusted and dedicated research and development (R&D) functions to support NPD.

2

Use phase data also contribute, having strong potential in the idea-generation process. Use phase data are generated during the use phase by the product itself (e.g. sensors) and related services (e.g. maintenance, repair, mobile applications). According to the study by Wilberg et al. (2017), which focused on the mechanical and plant engineering industry, an analysis of use phase data can reveal the needs and preferences of customers regarding product features and usage scenarios for products. It can complement assumptions and experience during conceptual design, reducing over-engineering, understanding demand peaks, and improving simulations. By replacing data generated in the test laboratory, the use phase data can also speed up endurance tests.

To sum up, current studies regarding specific data sources are focused on two types of data: social media data and use phase data. Social media data are used to explore the direction of product design. These data offer insights into how customers perceive their products and competitors. However, such data are only used as side information, and the main input for product development comes from R&D. The use phase data are similar to social media data usage in product design; the data yield insights into conceptual design to understand the usage and improve endurance. There is a gap in exploring other sources of data apart from social media and other types of use phase data from different industries.

Table 2.10 Qualitative studies examining the relationship between applying big data and innovation

Qualitative studies examining the relationship between applying big data and innovation			
Author	Lehrer, Wieneke, vom Brocke, Jung & Seidel (2018)	Chester Goduscheit & Faullant, 2018	Troilo, De Luca & Guenzi (2017)
Title	How big data analytics enables service innovation: Materiality, affordance, and the individualization of service	Paths toward radical service innovation in manufacturing companies—A service-dominant logic perspective	Linking data-rich environments with service innovation in incumbent firms: A conceptual framework and research propositions
Big data aspect	Big data analytics	Data-rich environment, digitalisation	Data-rich environment (i.e. the 3 Vs of big data, data management and analytics technologies)
Antecedent	Data sourcing & storage features. Data analytics features. Data exploitation features.	The four main axioms of SDL: Actor value network, resource liquefaction, resource density, resource integration.	Data management and analytics technologies
Mediator/ Moderator	Mediator: • Automation of customer-sensitive service actions • Human-material customer-sensitive service practices	n/a	Mediator: pattern spotting, real-time decisioning, synergistic exploration Moderator: organisational enablers (e.g. agile processes)
Consequence measurement	Service individualisation	1. Service concept innovation 2. Customer experience innovation 3. Service process innovation	1. Service concept innovation 2. Customer experience innovation 3. Service process innovation
Findings	The research concludes that big data analytics (BDA) have potential for service innovation and identified factors that are relevant to new value propositions. The research identified two key roles of BDA in service innovation, automation of customer-sensitive service provision and human-material customer-sensitive service practices and highlights how these are grounded in material features of BDA.	The study proposed a research model that explains how different configurations of conditions could bring about a radical new service as an outcome. Among those configurations, some conditions play important role, and others play secondary role. Digitalisation is the core condition, configurated with several peripheral conditions (actor value network, resource liquefaction, resource density and resource integration) to enable radical service innovations.	The research result supports the view that data management and analytics technologies enhance service innovation in incumbent firms. The effect is mediated by data density processes (pattern spotting, real-time decisioning, synergistic exploration) and enhanced by organisational enablers (e.g. agile processes, top management support).

Table 2.10 Continued.

Qualitative studies examining the relationship between applying big data and innovation			
Method	Exploratory multiple case studies in four private sector B2C firms in insurance, banking, telecommunications, and e-commerce industries from Switzerland, Austria, and Germany; 30 interviews were conducted in order to develop a model to explain how big data analytics enables service innovation.	A multiple case study approach: the data come from interviews and secondary materials from 24 B2B manufacturing SMEs in Denmark. Data analysis through a fuzzy set qualitative comparative analysis.	40 semi-structured interviews in seven large “non-digital native” firms in Italy and the UK in the service, electricity, gas, petrol stations, car sharing, postal services, telecommunications, banking, logistics, and insurance sectors.
Which level	Company	Company	Company
Theoretical base	Service-dominant (S-D) logic. Materiality and affordances	Service innovation. Service-dominant (S-D) logic	Service innovation. Data-rich environments

2.5 Summary of research gaps

2

Although the use of big data in product innovation has been known in the industry for quite some time, the topic of exploring the antecedents and consequences of using big data for product innovation is still in its early stages. Scholars have mainly explored the effect of big data and a few covered antecedents. Their merits notwithstanding, there remain unresolved gaps and questions to reach a more comprehensive understanding of the topic.

First, many current studies have examined the relationship between big data and product innovation. There is an agreement in the existing literature that using big data (BDAC, 3Vs, and big data usage) has a positive impact on fostering innovation, boosting innovation performance and new product success (e.g. Dubey et al., 2021; Ghasemaghaei & Calic, 2020; ZareRavasan, 2021). Most studies focus on the consequences of using big data, but a gap remains in the literature exploring both the antecedents and consequences of using big data. Furthermore, there is also a gap in exploring the relationship of antecedents and consequences in more detail, which would involve investigating which kinds of knowledge, skills, competence, or resources are related to big data usage in which ways. For example, what is the effect of the constructs of BDAC on using big data in innovation? What is the impact of data-driven culture on using big data in innovation? In addition, if there are relationships among these resources, then how do these factors influence each other? For example, what is the effect of organisational learning on human skills? Regarding the characteristics of big data, there are unanswered questions as to why variety and velocity have a significant impact on innovation performance, but volume does not. Regarding the effect consequences, there is a gap in exploring a deeper level of detail and finding out how elements of BDAC impact product innovation performance.

Second, as most studies explore the relationship between big data and product innovation in general, there is a lack of research focusing on digital products. This gap should be filled, as digital products account for a big portion of the products being developed, and most of them are developed using various sources of data.

Moreover, there is a lack of reflection on the project level. The environmental issues that are encountered vary from project to project even within the same company, and involve resources such as human resources, funding, and team culture. When observation focuses on a smaller unit such as the project level, the mechanism behind it can become clearer. Thus, there is a need for more project-level research from a qualitative viewpoint.

By filling in these gaps, the knowledge in the field will become deeper and more detailed. For example, the effect of big data on innovation can be down to the relationships between the elements of big data from the scope of the innovation projects. Thus, granular details would make the mechanism behind it more comprehensive.

2.6 Research questions

This research aims to fill in the gaps in the extant literature by using a multiple case study design (Yin, 2017) from resource based view and dynamic capabilities view to explore the mechanism behind developing BDAC in innovation projects in practice and with which results. Specifically, it addresses one main research question with four sub-questions:

RQ: How do companies develop big data analytics capabilities in digital product innovation?

Sub-RQ1: What are the antecedents of developing big data analytics capabilities in digital product innovation at the project level?

Sub-RQ2: How do the identified antecedents influence the development of big data analytics capabilities in digital product innovation at the project level?

Sub-RQ3: What are the consequences of developing big data analytics capabilities in digital product innovation at the project level?

Sub-RQ4: How do big data analytics capabilities influence its consequences in digital product innovation at the project level?

Figure 2.3 Conceptual model illustrating the research questions of this study.

Figure 2.3 visualises the research questions of this study. Guided by these research questions in this research project, this study aims to develop a theoretical framework. The framework links company characteristics, environmental factors, big data analytics capabilities constructs and performance consequences. The following Chapter 3 explains the method adopted in this study.

CHAPTER 3

Methods

3.1 Introduction

This doctoral research explores the phenomenon of using big data in product innovation projects and aims to answer the main research question: "How do companies use big data in digital product innovation?" The research investigation focuses on the antecedents and consequences of using big data in innovation projects, highlighting the relationship between them to unveil the underlying mechanism. Thus, this research project adopts the inductive research approach to explore the phenomenon. It relies on rich qualitative data – interview transcripts collected from four cases in different industries for analysis. The aim is to build a theoretical framework that explains the abovementioned mechanism. This chapter explains this approach in detail and the rationale for using it.

3.2 The approach to research

According to Creswell & Creswell (2018), the broad research approach is the plan or proposal to conduct research. The three components of the broad research approach are philosophy, research designs, and specific research methods.

3.2.1 Philosophical worldviews

The philosophical worldview, also known as the paradigm, can be seen as a general philosophical view of the world and the nature of research that a researcher brings to the study (Creswell & Creswell, 2018). Schwandt (2007, p. 87) describes epistemology as "the study of the nature of knowledge and justification". There are many types of epistemologies. This study assumes a relativist epistemological orientation. Relativism is "the doctrine that denies that there are universal truths" (Schwandt, 2007, p. 87). By applying a theoretical lens, the way the perspectives of different participants can be captured across cases depends on the theory in use (Yin, 2013).

3.2.2 Case studies

According to Yin (2013, p.16), "a case study is an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident". The research topic of using big data in innovation is still at an exploratory stage. Thus, the qualitative research method is adopted as it is appropriate for exploring and understanding new research topics (Creswell & Creswell, 2018; Silverman, 2008). The case study method was adopted to investigate the research questions as this is an empirical inquiry investigating a contemporary phenomenon in depth (Yin, 2013). Using multiple case studies enables the identification of company characteristics and environmental aspects related to the development and consequences of big data analytics capabilities in digital product innovation processes. Case study research is especially useful for "how" and "why" questions (Yin, 2013); it aligns with one of the research questions, "How do the identified antecedents influence big data usage in digital product development and its consequences at the project level?"

This doctoral research project contains a multiple-case study with four cases. The case studies analysed four new digital service development projects from four big companies operating in different industries (i.e. transportation, internet, navigation, and cybersecurity) in the past five years. Data from various sources were used in machine learning and statistical models to do predictive analysis to enable the main function of the product (see Tables 9 and 10 for overviews of the cases and interview participants, respectively). The semi-structured interview protocol focused on the process of using big data in innovation projects, the challenges and factors that affect this process, and its consequences.

3.2.3 Sample

This research project selected four cases that are all big data-enabled digital product development projects in Europe and Asia in the past five years. These projects were initiated by four companies in the public transport, news applications (app), navigation, and cybersecurity industries. These cases offered a rich empirical setting for this research project for three main reasons. First, the companies could access a vast amount of data from various data sources, and the data were generated at a fast speed, which aligns with the 3Vs of big data (volume, velocity, variety). Second, the projects are digital product development projects initiated in recent years. Third, the cases are selected from Europe and Asia because factors such as privacy regulations, company cultures, and customer preferences are different, which could influence the usage of big data in innovation projects. Those differences highlight the influences of the factors. For example, the differences in privacy regulation could further reveal the importance of privacy in influencing the data sources used in the project. Access was gained via four contact persons, each of whom worked as a manager in the case companies. The managers then connected the researcher with the interviewees. The interviewees are people who were involved in different stages of the product development projects.

In general, the interviewees were all involved in the product innovation project, so they are in a position to provide information about the project that others would not know. Each participant shares information about the project based on their role in it. The participants might not know all the details of the project. For example, the managers know how the project was initiated, the product owner knows quite a lot about the design of the product, while the developers know how the algorithm works; together they build a more complete picture of the project. On the other hand, different participants' answers validate each other regarding the product description, process, and activities. These answers about activities and what happened in the innovation process could help this research in analysing and identifying factors that influence the innovation outcomes.

Table 3.1 Overview of case companies

	Company A	Company B	Company C	Company D
Area	Europe	Asia	Asia	Asia
Industry	Public transport	News app	Navigation	Cybersecurity
Data source	Multiple, mainly weight data	Multiple, mainly internet content	Multiple, mainly user navigation data	Multiple, mainly network data
Big data innovation project	Big data is the input for statistical and artificial intelligence models to predict the estimated crowdedness.	Big data is the primary input to train algorithms to make suggestions for content creators for inspiration and personalise the browsing experience for users.	Big data is the input for machine learning algorithms to estimate the number of customers visiting the petrol station.	Big data is the input for machine learning algorithms to distinguish normal and abnormal network activities.

3.2.4 Semi-structured interviews

The data were collected through semi-structured interviews; the data for the first case were collected in spring 2020 and for the second, third and fourth cases from May 2021 to November 2021. The interview protocol (see Appendix) focused on using big data in innovation projects, the factors that affect this process, and the results/consequences. Interviews were performed online (due to the coronavirus pandemic), opening with the introduction of the research and then proceeding on to the themes of the interview protocol. The interview guide was derived from innovation management literature and survey questions from a previous study (Mikalef et al., 2019). The interview guide was also tested and improved through several pilot interviews with practitioners at a bike-sharing, consultant, and grocery delivery company.

The interview was semi-structured, and thus all questions or their alternatives were asked during the 16 interviews. Some questions were more relevant to some interviewees. People taking care of hands-on tasks in projects could answer with a deeper level of detail, whereas managers who were mainly involved in the initiating phase would provide information from a different perspective. This diversity provides information about the project from different angles.

Sometimes questions were modified during the interviews. For example, when a participant did not think that their innovation project is divided by stages when answering "how do you divide this innovation project in different stages?", a different question was asked: "Could you briefly describe the innovation process for this project?"

Follow-up questions were asked when necessary to encourage more accurate and thorough responses. Examples of such questions include "what did you mean by 'the term'", or "could you give me an example?" Follow-up questions were also used to encourage an answer; for instance, if the interviewee had little to say in response to "what challenges to using and creating value from big data do you see?", then a follow-up question was asked, such as: "what do you spend a lot of time on during the process?"

Multiple interviews were conducted for each case, aiming to reach theoretical saturation. Interviews were recorded and transcribed for analysis. Each interview lasted between 47 and 180 minutes, most typically between 90-100 minutes. Some interviews are longer than average because the interviewee kindly explained the technical part of the project to enable the researcher to understand their work; it takes time to explain technology to someone who does not come from a computer science and Internet of Things background. Extra time was spent on explaining which sensors are used, how these sensors work, and in rough terms how the insights from different sensors are integrated in the algorithms and models, etc.

Among the four cases, people believed that their products had been improved continuously since the innovation process began. To understand the participants' opinion about BDAC, we needed to focus on the influencing factors because together they made up BDAC. Potential influencing factors were identified through questions such as "what do we need in order to improve the product?" Several common factors were mentioned by most participants; a number of different ones were mentioned depending on the role people played in the project and what they focused on.

For triangulation, desk research was conducted with publicly available materials and product introduction slides. For example, the product details could be found on the websites and the apps. Some of the product launch information was mentioned in news releases. Product introduction slides designed to be shared with clients were also shared with the researcher.

The following section introduces the four cases, including project descriptions, interviewee descriptions and their involvement in the projects.

3

Table 3.2. Summary of interview participants

Company	Interview number	Participant's position	Company	Interview number	Participant's position
A	1	Programme manager A	C	1	Solution architect
	2	Programme coordinator		2	Domain expert
	3	Programme manager B		3	Data analyst
	4	Product owner		4	Product manager
B	1	Product manager A	D	1	Developer A
	2	Data scientist		2	Product manager
	3	Product manager B		3	Developer B
	4	Content operations specialist			
	5	Developer			

3.2.5 Four cases – introduction

Case 1 is a software development project in the transportation industry. The software product predicts the crowdedness of trains for passengers when they take public transportation. A large company initiated the project in the transportation industry in a European country. The company saw significant potential in big data analysis for product innovation and had already conducted several projects previously where big data was used in the innovation process. In this project, multiple data sources were used as input to build a model to predict crowdedness during the development process. This new service is now available on the market. Four people participated in the interviews, including two programme managers, a coordinator of data & analytics, and an innovation project leader who was also the project owner. All participants were involved in different stages of the project, from ideation and development to implementation, and were highly knowledgeable about the activities involved in the project. Programme manager A managed several analytics teams and oversaw the project after launch. Programme manager B was involved in project establishment, including obtaining funding from the company, and building up the team, etc. The programme coordinator (consultant) participated in the ideation stage of the project and continues to coordinate the project. The product owner actively participated in the project, including building the models and APIs for the product.

Case 2 is a mobile browser app development project. The mobile browser is designed to display personalised web content for mobile phone users. A large company in the internet industry in an Asian country initiated this project due to changes in the mobile browser market, which has shifted from information searching to information feeding and from mainstream media and publications to new media. During the development project, user profiles and browsing content data are collected and analysed to provide personalised content for users. Meanwhile, content creation suggestions are provided for content creators based on the same information.

To gather information about the development of this project, five people participated in the interviews: two product managers, a data scientist, a developer, and a content operations specialist. They were all involved in the development and implementation of the project and were highly knowledgeable about the project's activities. Product manager A participated in the service design of the news app that serves the app users. The app content is provided by content creators. The creators use the news content platform to get inspiration and upload their content. Product manager B participated in the service design of the news content platform that aims to serve content creators. The content operations specialist monitors the product performance indicators and aims to implement improvements, as the product is continuously updated after launch. The data scientist and developer participated in building and iterating the models.

Case 3 is a Software as a Service (SaaS) development project. The product provides operation indicators to assist petrol retailers in their daily operations. It was launched by one of the leading navigation providers in an Asian country. The company offers maps, navigation devices, and apps for businesses and individual customers. The company collected and analysed navigation data during this development project to generate indicators for SaaS products. Petrol retailers refer to the indicators and insights from the SaaS product to facilitate their daily operational decision-making. To understand the development process of this project, the researchers interviewed a domain expert, a product manager, and a data analyst. All three interviewees played essential roles in the project's planning, development, and implementation. This is a SaaS product with many traffic-related indicators as the product's functions. The data scientist participated in building models for these indicators. The domain expert helped design the product to fit the petrol industry and participated in meetings with potential clients to understand user needs, for example. The solution architect was involved in the early ideation and design phase of the project, had meetings with clients to understand user needs, then translated the needs into product design. The product manager had user meetings and designed the product from a detailed level, from designing the indicators to visual design.

Case 4 is a cybersecurity system development project. A cybersecurity company launched this project in an Asian country. The product is a security system that protects clients' business systems. Historical cybersecurity operation data are collected to train machine learning models to identify threats more effectively. Three people were interviewed: two developers and a product manager. They were all involved in the development or implementation of the project. Developer A participated in developing the big data platform. The big data platform aggregates various sources of data on a large scale on the cloud by combining data management hardware and software (Rice, 2023). The product manager designed the product – mainly the functions and features based on the big data platform. Product developer B built and improved the algorithms for the product and participated in user meetings to understand user needs.

3.2.6 Data analysis

All interviews were transcribed and coded according to recommended practices, first per case and subsequently across the four cases. In within-case analysis, the researcher did data analysis by considering the existing terms and relationships investigated in the existing literature and applying the theoretical lens. The second part is cross-case analysis. In cross-case analysis, the same method applies, and the existing literature is more fully discussed and compared, with iterative confrontation between the coded data and existing literature.

The data analysis process went through several stages to serve the purpose of answering the research questions. In stage one, initial open coding, or first order coding, was conducted to break down qualitative data in transcripts into discrete parts, labelled with codes and group codes into categories (Corbin & Strauss, 2007). For example, "privacy & regulation" and "company culture" were two factors that were subsequently grouped into the "environmental uncertainty" and "intangible resources" categories, respectively. The second stage was axial coding, or second order coding, which sought similarities and differences among the codes, linked categories, and connected data (Corbin & Strauss, 2007). The third stage mainly involved iterating between emerging data and categories and the relevant literature to identify whether the findings have precedents or discovered any new concepts (Gioia, Corley & Hamilton, 2012).

The table below presents four examples of first and second order codes for RQs in this research. First order coding aims to arrive at answers to RQ1 and RQ2, respectively, "What are the antecedents of using big data in digital product development at the project level?"; "What are the consequences of using big data in digital product development at the project level?" Second order coding aims to obtain answers to RQ3, "How do the identified antecedents influence big data usage in digital product development and its consequences at the project level?" The codes also depend greatly on the context of the quote and the theoretical lens applied for this study. When applying the theoretical lens, the elements of big data analytics capabilities, dynamic capabilities and environmental uncertainties are used as a checklist for codes. If none of the elements fit, then new elements emerge. Due to the limited space in the table, only the quote has been extracted, but not the context. Within case analysis in Chapters 4, 5, 6, and 7 and cross-case analysis in Chapter 8 will explain these findings in more detail.

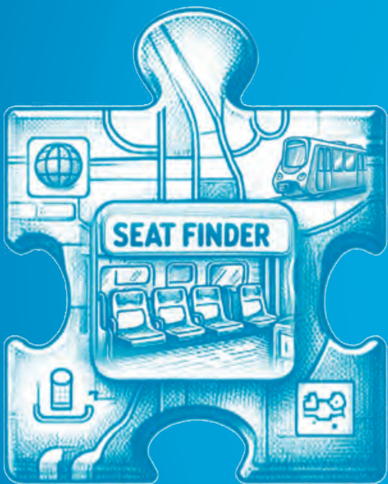
Table 3.3 Selected empirical examples of first order and second order codes

Empirical examples	First order codes	Second order codes
"It's not just that I want to be data-driven; here's your data, like terabytes of data, good luck! That doesn't make you a data-driven company. You have to get information from all of your data. And to get this information out of your data requires a lot of work for each decision you want to take. You have to get all your data ready. You have to build your information products, and then you can become a truly data-driven company. It takes time." (product owner, Case 1)	Data-driven culture or company culture (intangible resources). A construct of BDAC (according to the theoretical lens applied to this research).	The company culture could facilitate development of BDAC.
"There was a dramatic drop in advertising revenue, which hugely impacted our income. Our business shrank, even if the number of users grew. We don't have income, so it was a pure expense. Operations and human resources are all costs, right? The whole business had to be shrunken after income dropped." (product manager A, Case 2)	Financial resources (tangible resources). Uncertainty of the actions of customers. Project stopped.	Financial resources and uncertainty of the actions of customers could influence BDAC.
"Because of the coronavirus, the pandemic has led to a decline in sales, and the decline in sales has caused their [petrol retailers] competition to become more intense. Everyone will compete for a small number of users. Once their sales decline, they need tools. Whereas in a rising market, they just need to expand on their own. They don't need to refine their operations." (domain expert, Case 3)	Hostility (environmental uncertainty).	Under hostility, companies look for new technology to help them compete, in this case big data.
"When clients use our product, we have a discussion. If they don't want to give us their data for further training, that is fine. But we encourage clients to bring us their own data to update our algorithms, there are all kinds of situations [...] Especially in large enterprises, which are strict about this, some data cannot leave their intranet." (developer B, Case 4)	Privacy (privacy & regulations). Data variety.	Privacy is restricting data variety.

The general analytic strategy for this multiple case study is close to grounded theory (Corbin & Strauss, 2007). This study worked the data from the "ground up" during the process. It identified patterns suggested by the data and then extracted insights from them. The difference is that it applies the theoretical lens from the resource-based view of the firm. The within-case analysis mainly extracted the linkage between factors (internal and external) and the development of big data analytics capabilities (BDAC), and the connection between different aspects of BDAC and innovation performance.

The cross-case synthesis in this study juxtaposed the data from several interviewees to search for a similar linkage between certain factors and aspects of big data usage. When encountering factors that emerge from one case but not another, the issue was analysed to come up with possible reasons for this.

3


SEAT FINDER

CHAPTER 4

Case Study 1 –
Big data in transportation

4.1 Introduction

This chapter presents the first case study. The case involves a company in the field of public transport, which uses big data in the development phase of a digital product development project. Big data sources collected from sensors are used as inputs for statistical and machine learning models, enabling the prediction of crowd levels to assist customers to find available seats more easily. To understand the project and collect data for research, four people who were involved in the project participated in one-on-one interviews. The case analysis identifies factors and relationships by adopting the theoretical lens of the resource-based view (RBV) and dynamic capabilities (DC). This chapter starts with the case background, then follows the findings with the framework as an overview, followed by quotes and explanations.

4.2 Case background

This digital product development project was initiated by a transportation operator. Big data sources from sensors are used as input for models to predict how crowded is the train arriving to the station. The crowdedness is down to coach level. Based on this information, passengers can decide if they should get this coming train or metro, or the next one. They can also decide which coach to take, the one on the front or the back, based on the crowdedness level. The project has three origins: First, to innovate on the commercial side. Second, to improve efficiency in transporting more travellers with full use of available seats. Third, harness the company's fast generating and aggregating data. This digital product development process has three stages: First, the idea is generated and selected. By leveraging existing data, it is technically feasible to generate insights regarding crowdedness. The company can then decide on whether to continue the product development based on its alignment with current business priorities and concerns. Second, develop and apply the service on a small scale on a limited number of lines. Third, scale up the service to many lines. Big data was used during the second and third stages of the development process. There are two main motivations for the company to use big data. First, people see the potential of big data and how it can make the company more data-driven. The second motivation is to thrive in competition in this fast-changing world. In the development process, over a billion data points are generated daily from various data sources; these data were used in the project as input for models to predict the level of crowdedness. The company used several analytical applications, including statistics and machine learning models, and then validated the crowdedness prediction with headcounts. During the development and implementation process, different cross-function teams collaborated, with each being responsible for different parts of the project, such as data collection, data analysis and app development. The service was first implemented on a few lines for the pilot test. After positive results, it was extended to cover more lines. During the Covid pandemic, the service was suspended for some time due to the very low occupancy level of the trains. A registration system was introduced to help understand travel patterns and make health-related predictions. Shortly after, the crowdedness prediction service was restored and resumed operations.

4.3 Findings

The findings suggest that in Case 1, there are many factors affecting the development of big data analytics capabilities (BDAC) in product innovation. These factors belong to the categories of tangible resources, intangible resources, human skills, dynamic capabilities, and environmental uncertainty. On the other hand, developing BDAC has an impact on innovation performance.

Figure 1 gives a general overview of the conceptual model of antecedents and consequences of developing BDAC in innovation in Case 1. The findings revealed several tangible resources including data variety, data reliability and data velocity. They also uncovered one intangible resource: organisational culture. Case 1 also highlighted human skills and knowledge including communication skills, team building and technical knowledge. Coordinating as one element of dynamic capabilities could also influence BDAC. Dynamism, corporate social responsibility, and privacy regulations also build an environment that has an impact on developing BDAC. Finally, developing BDAC has an impact on innovation performance. Product performance and project performance are the measurements of innovation performance.

In the following paragraphs, the association between antecedents and developing BDAC, and between BDAC and its consequences is explained in detail. Big data variety as an element of BDAC stood out from the findings, as several antecedents influenced BDAC by having an impact on big data variety. Thus, the findings are organised by the influencing factors of BDAC, big data variety and innovation performance.

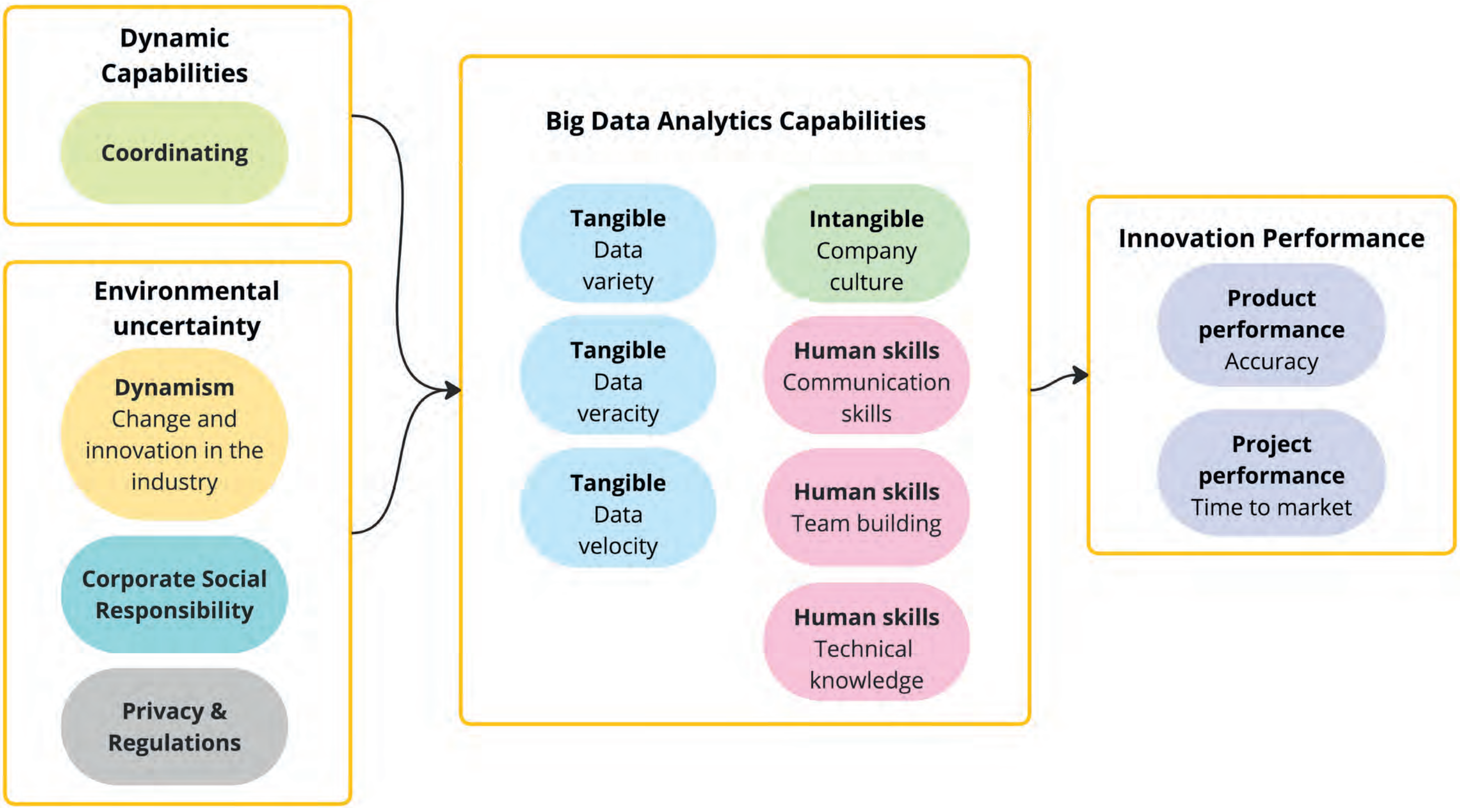


Figure 4.1 Conceptual model of antecedents and consequences of big data usage in innovation in Case 1. See Figure A1 in Appendix for a more detailed conceptual model.

4.3.1 Organisational culture & developing BDAC

Intangible resources – Organisational culture (data-driven culture)

Organisational culture refers to a set of values, beliefs and behaviours shared among employees of an organisation (Deshpande & Webster, 1989; Ravasi & Schultz, 2006; Belias & Koustelios, 2014). Data-driven is a trait of organisational culture. A data-driven company tends to use big data to make decisions in the innovation process with insight from data rather than instinct or personal experience. As the company uses big data analytics, it gradually develops big data analytics capabilities. Thus, a data-driven culture is identified as an influencing factor for big data analytics capabilities in Case 1. This is a journey that takes time for the company to complete.

> "... Well, there is a difference between how data-driven they wanted to be and how data-driven they are… the understanding of 'it needs to be data-driven' is getting better and better with everyone. But we don't have all the data available, so that also takes time. And you have to build your systems to get those data in the right form to make decisions. So, you have to make the right dashboard and predictions. And that also costs time. It's not just that I want to be data-driven; here's your data, like terabytes of data, good luck! That doesn't make you a data-driven company. You have to get information from all of your data. And to get this information out of your data requires a lot of work for each decision you want to take. You have to get all your data ready. You have to build your information products, and then you can become a truly data-driven company. It takes time." (product owner, Case 1)

The premise of being data-driven is a mindset that depends on whether people in the company want it to become data-driven. Of course, being data-driven is not going to change anything unless the company has access to the necessary data, which is the first step. That said, merely having the data is not enough. It also takes time to extract and generate insights from those data. Then, the next step is to use the insights generated from data to support decision-making. When that has been done, the company will become more data-driven. Thus, people in a data-driven company would be more willing to use big data and develop big data analytics capabilities. So, a data-driven company culture facilitates big data analytics capabilities.

In addition, several aspects of improving data awareness and becoming more data-driven were identified from interview data, including dissemination by data advocators and management support. First, data advocators such as data scientists can encourage people to use data in business, which helps build a more data-driven company. On the other hand, it is also important that the data-driven initiatives have support from the top. Management support is key to implementing the initiatives and allocating human resources and data resources for them.

> "I think that having data scientists who can spread the word is not the only thing that will help, because there's noise everywhere saying that you have to use data … the top of the top must also be involved, so the strategic choice of [the company] to become data-driven will also help to get everyone moving, and becoming more data-aware, and makes it easier to more use data. So, you have to reach out to all these people from both top-down and bottom-up." (programme coordinator, Case 1)

4.3.2 Dynamism & developing BDAC

Environmental uncertainty – dynamism – change and innovation in the industry

This research adopts the definition of dynamism as the rate of change and innovation in the industry as well as the uncertainty or unpredictability of the actions of competitors and customers (Lawrence and Lorsch, 1967; Thompson, 1967; Burns and Stalker, 1961). In Case 1, respondents considered big data as one of the primary sources to help the company stay competitive in the transportation industry. For a public transport operator, alternative transport solution providers, such as autonomous driving vehicle manufacturers, are competitors in the fast-changing environment. What would these competitors do and how would customers of transportation react to the fast-changing and unpredictable environment? In reaction to the dynamism described above, the company believed it needed new technology and innovation to survive. Thus, change and innovation in the industry facilitate the development of the BDAC of a company, as indicated in the quote below:

> "… as a society, there's a lot of stuff going on, of course, and [due to] the development of big data, data solutions are growing very fast, and developing very fast. The risk of a disruptive innovation like self-driving cars is a very real risk for us. So, despite our risk-averse nature, we have to be open to these kinds of innovations and the use of big data to stay competitive in this very fast-changing world, so it's a difficult balance to, on the one hand, just be a reliable transport operator that customers can use and rely on, with vehicles that run on time. On the other hand, [we have to] be more agile and flexible and able to adjust to new solutions that become available." (programme coordinator, Case 1)

4.3.3 Corporate Social Responsibility, BDAC & project performance

Environmental uncertainty – social responsibility

Project performance – time to market

As a public transportation provider, the company has a big responsibility to society. Every new product or service released to the market must be very reliable and not cause chaos in the public transport system. This makes the company very cautious about using new technologies in innovation projects. A big data-enabled innovation must be a proven product before it is released to the market. Any mistake would destroy the company's reputation because the public counts on it to provide reliable public transport. So, social responsibility poses resistance to using big data for innovation, and as a result it takes longer to release such a product to market. According to the European Commission, corporate social responsibility (CSR) refers to the responsibility of enterprises for their impact on society. In academia, CSR can be defined as "treating the stakeholders of the firm ethically or in a responsible manner" (Hopkins, 2005, p. 214). CSR can be seen as both an ethical stance without any expectation of getting rewards and a business strategy of serving the needs of stakeholders to have a greater chance of gaining rewards (Wan-Jan, 2006). In Case 1, depending on the reaction of the stakeholders – public transport users – the innovation might either be accepted or not, making the environment of using big data in innovation uncertain. As the programme manager mentioned, "If we want to use a data solution, it has to prove itself before we can use it." CSR thus makes the company more cautious in using big data. It takes time to prove that the data solution is feasible. This could delay the decision to use big data and initiate this project – and would then also mean that it will take longer for the product to reach the users. Thus, social responsibility is negatively associated with developing BDAC and further with project performance – time to market.

> "… we need to be very careful about making no big mistakes. So, regarding big data, we are very much looking at security, privacy, that kind of stuff, because if we make a mistake there, the whole news (media) will look at us, all the newspapers will look at us and at what's happening there. So that makes us quite careful in that case. And we have a function in society, we move people from A to B, but they mostly travel there because they need to for work or other important things … So, if we want to innovate or if we want to use a data solution, it has to prove itself before we can use it. Whereas in another field, you can try something earlier, and if it fails, well, it fails, and then you try something else. If we fail, then everybody complains about it. So, we are quite strict about it." (programme manager A, Case 1)

4.3.4 Technical knowledge & data variety

Company characteristics – human skills – technical knowledge – data management
Tangible resources – data – variety

Technical knowledge refers to knowledge about technical elements (Akter et al., 2016), for example, data collection/integration and data management. Based on the interview data, data management has influenced data variety. How data are saved or stored affects how easy it is to extract these data for analysis, that is, data collection. When a data source is not stored centrally, the time and effort required for data collection means that it is more challenging or even difficult to use such sources than centrally stored data sources. Thus, data management is associated with data collection and data variety. The quote below exemplifies this:

4

> "Yes, we have identified about 14 or 15 different data sources that might be very useful for us, but it's very difficult to extract data from those systems and into our central data lake … and many of them [data storage systems] are old technology and are stored physically in the vehicle itself. So, building a central model will be very difficult … For example, somebody has to go physically in the vehicle to retrieve a certain type of data. It's not all stored centrally. Only the last generations of vehicles have enabled access to data from our headquarters." (product owner, Case 1)

4.3.5 Team building & data variety

Human skills knowledge – team building
Tangible resources – data – variety

A value chain refers to the series of activities required to take a product from ideation through different stages of development and production, and then deliver it to customers, until final disposal after usage (Kaplinsky & Morris, 2000). In Case 1, managing different function teams in one big project would be very helpful in aligning these teams. For example, changes in customer requirements might result in adjustments to data collection, and thus it is more efficient to manage the data collection team as part of a big project along with other teams in the value chain to efficiently collect data. So, team building could benefit data variety. The following quote exemplifies this:

> "So, the whole chain up to the app should be part of one programme or project because then you have control over everything, and you

> know that what you are building is being used. And the same is true of the back end; we would like to have the sensor data collection team be part of the programme because changes in the requirements upfront from the customers might result in changes in the sensor data collection team. So, when you have the whole chain in one project, steering everything will be a lot easier because you reduce your external dependencies." (product owner, Case 1)

4.3.6 Privacy & data variety

Environmental uncertainty – privacy & regulations
Tangible resources – data – variety

In Case 1, the regulatory environment and privacy issues influence the data source that can be used. Some data sources may be useful in providing vital insights for the project but could not be used due to GDPR. So, privacy and regulations reduced the data sources (data variety) that could be used in the innovation project. Here are the relevant quotes from the participants:

> "And we have some technology that has a privacy aspect – we have cameras on board, and we'd like to use a counting system based on the camera images, but that has a very big legal impact because of GDPR, so I'm not sure if we will ever use it." (product owner, Case 1)

> "So, yeah, Covid is very, very bad, I understand. But is it that bad that we want to track everybody on a personal level to see where they are going? It's really a privacy issue." (programme manager B, Case 1)

4.3.7 Coordinating, data variety, technical knowledge, BDAC, business analytics & product performance

Dynamic capabilities – coordinating
Human skills – technical knowledge
Tangible resources – data – variety
Product performance – accuracy

Coordinating (collaboration with external stakeholders) emerged from the interview data and was associated with developing BDAC (data variety). In the meanwhile, technical knowledge (IT architecture) emerged from the interview data and was associated with developing BDAC (data collection and data analysis). Coordinating is needed for efficiently gathering and processing data because some of the data sources are provided by a third-party supplier. On the other hand, the IT architecture needs to be designed and built in such a way that it supports the demand for real-time or near-real-time data. The two company characteristics influence data collection and near-real-time prediction (business analytics). When prediction gets closer to real-time functionality, the accuracy of the product increases, which represents the product performance. The quotes below exemplify this:

> "We also use data sources from other companies … for example, from company X, who is in charge of the infrastructure, we got a lot of data." (product owner, Case 1).

> "That has to do with the ... really has to do with the speed, the time the data is collected until the moment that data reaches our data platform ... you have to improve the speed of processing in this whole chain of data collection and things for real-time availability. And it's just not there at the moment. And it has to do with, I think, technical architecture, how the system was built, and that when it was built, like 20 or 10 years ago, we didn't foresee this real-time need for data." (programme coordinator, Case 1)

4.3.8 Communication skills & team building, data variety & project performance

Human skills – communication skills & team building
Tangible resources – data – variety
Innovation – project performance – time to market

In Case 1, different groups worked together in the data collection process. While one team is collecting new data sources, another team is adjusting the data model based on the new data source. So, the priorities set for the data collection team play an essential role in influencing the speed and timing of data collection and analysis. When the task is not their top priority, it takes longer (time to market) to get the data. In this scenario, different teams need to work together as one cross-functional team. The key is to maintain good communication and coordination in a timely manner to ensure a seamless fit between data collection and analysis work. See the quote below as an illustration:

> "For example, when the app isn't part of our project or a programme, and we add new data, we have no way to influence that they have the

priority to add it to the app. And you want to have that, so you can say, ok, the app will be changed in time, we will adapt our models in time, and everything will be released at the same moment when adding new values to our customer ... However the priority of the sensor data collection team is set by the sensor data collection team, and we have to try to get on that list and get the priority for our project.

Q: So, in this case, if they don't put it on priority, how do you communicate with them?

A: Lots of talks. Oh, it's, err, networking and talking, showing everybody that what we are building will deliver value to the customers and our organisation. And from that, you try to get everybody to add it to the systems. But it's not centrally managed." (product owner, Case 1).

4.3.9 Data variety & product performance

Tangible resources – data variety
Innovation performance – product performance

The findings suggest that data variety improves product performance, as can be seen, for example, in this quote from a respondent who was asked about any improvements that can be made to the project:

"Always, you saw the picture with all the white circles. There are so many data sources that we can use to improve this, yes." (product owner, Case 1).

Through analysing data sources, the companies could gain insights that are useful for building products. Using different data sources (data variety) can provide diverse insights that enrich the products. Sometimes, insights gained from different sources regarding the same topic or information could also be triangulated with each other. Thus, data variety could improve product performance.

4.3.10 Data reliability & product performance

Tangible resources – data – veracity
Innovation performance – product performance
Human skills - Business analytics - real-time analytics

Another aspect involves analysing data in a shorter timeframe. More reliable data combined with real-time analytics increase product accuracy. The data pattern changes over time, and thus data collected in a short timeframe or in real time more reliably reflects the insights data could bring. Real-time analytics could make it possible to analyse reliable data in real time with minimum delay, thereby improving predictions. Since prediction is the main function of the product, product performance is enhanced. The following quote illustrates this:

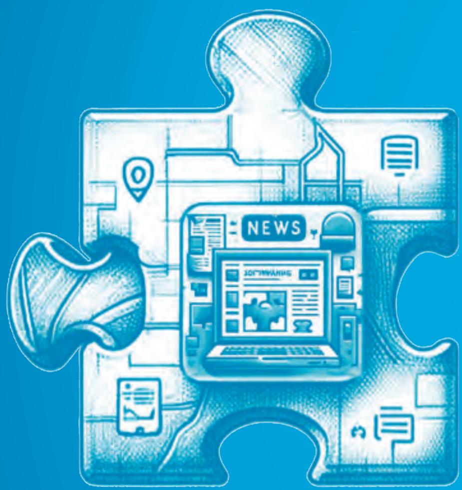
> "And it's also that the current model runs predictions like a few days, or maybe a week in advance – I'm not sure which day, but it's once a week, and the predictions are then all sent, they're fixed. But I'd like to improve the service by using more short-term data and making it more real-time prediction … it's a more accurate prediction if you make it run on shorter notice ... Yeah, but real-time is quite difficult when you are talking about predictions …" (programme coordinator, Case 1)

4.3.11 Data velocity & project performance

Tangible resources – data – velocity
Innovation performance – project performance

Based on the interview data, velocity also affects project performance – time to market. Combining the definitions, velocity measures the speed at which data are generated, collected, stored, and analysed (Gandomi & Haider, 2015; George et al., 2016). When the speed of data collection and preparation increases, the service can be made available sooner to the customers. This shortens time to market, thereby increasing project performance. As a participant states:

> "Speed – we want to move to a single data platform to increase speed. We have now already spent three years building this data platform, and my dream is to have these data available for my data management and also for advanced analytics teams. We are not there yet. We haven't finished. It's still at the start. So, my improvement for them is they got much faster access to the data, and their time to market will be much better." (programme manager, Case 1)


NEWS

CHAPTER 5

Case Study 2 –
Big data in a news app

5.1 Introduction

This chapter presents the second case study. The case comes from a company that offers a newsfeed app. Big data was used in the development phase of this newsfeed app project. The data are used as input for models to categorise news content and make suggestions for both content creators and app users. To understand the project and collect data for this research, five people who were involved in the project participated in one-on-one interviews. The case analysis identifies factors and relationships by adopting the theoretical lens of the resource-based view (RBV) and dynamic capabilities (DC). This chapter starts with the case background and then follows the findings with the framework as an overview. The findings are supported by quotes and explanations.

5.2 Case 2 background

Case 2 is a news app development project. The news app is designed to display personalised news content for mobile phone users. A large company initiated the project in the internet industry in an Asian country. At that time in the market, there was a shift in news content from mass media to personalised newsfeeds, aiming to prevent homogenisation, as all media platforms were offering highly similar news. Content is transitioning from official and high-profile media sources to new media, from professional reporters and editors to non-official content creators who provide alternative content outside of the mass media. The format also shifted from picture- and text-dominant to short video-dominant content. Readers no longer rely merely on searching and browsing; the newsfeed flow is tailored to their liking. To achieve this, data and algorithms are needed. From the users' side, the content is personalised to match their preferences. That content is used as data to feed machine learning algorithms for categorisation, filtering out inappropriate information and matching with users' preferences to make news suggestions. App usage data such as browsing history are collected and analysed to understand user preferences.

On the other hand, the company also guides content creators and suggests hot topics and current and future trends. Machine learning and big data analysis categorise and analyse content to determine its popularity. Five participants were interviewed to gather information about this project's development, including two product managers, a data scientist, a developer, and content operations specialist. They were all involved in the development and implementation of the project and were highly knowledgeable about the project's activities.

5.3 Case 2 findings

The findings from Case 2 suggest that there are two groups of antecedents associated with big data usage in innovation: company characteristics and environmental factors. Company characteristics refer to the features of a company. In Case 2, three company characteristics are identified: company culture, collaboration with external stakeholders and agile product development. Two environmental factors are identified – external shock and technology development. The consequence of big data usage in Case 2 is identified as innovation performance, specifically product performance.

The antecedents and consequences are associated with several constructs of big data usage, including data source, data management and near-real-time processing. Data source refers to where the data that are being used originate from. Data management refers to the management of data resources. Near-real-time data processing is used when information does not need to be processed immediately but a fast speed is still required, such as processing time in minutes. In the following paragraphs, the association between antecedents and big data usage and between big data usage and consequences is explained in detail. Figure 5.1 gives a general overview of the conceptual model of antecedents and consequences of using big data in innovation in Case 2.

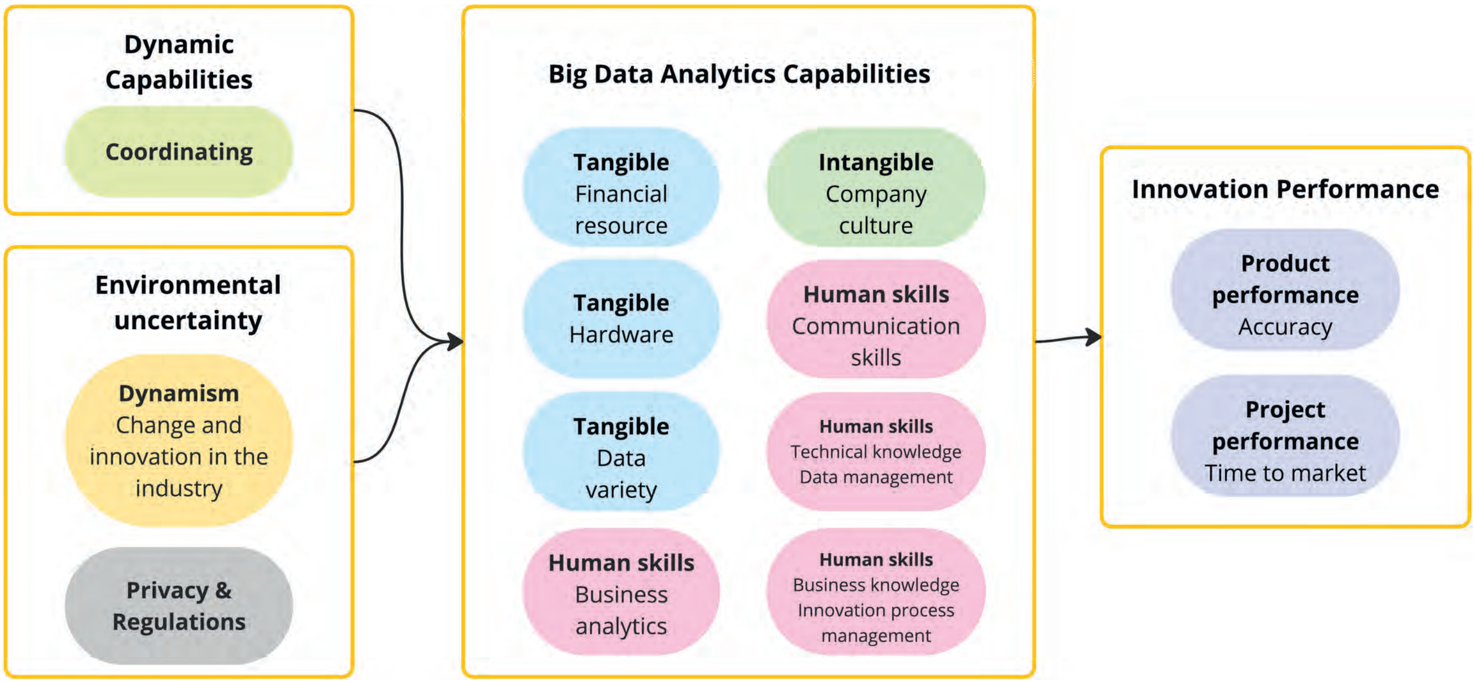


Figure 5.1 Conceptual model of antecedents and consequences of big data usage in innovation in Case 2. See Figure A2 in Appendix for a more detailed conceptual model.

5.3.1 Company culture & developing BDAC

Intangible sources – organisational culture
Human skills – technical knowledge – data management

An open organisational culture can potentially facilitate the development of big data. When people work towards the same goal – that is, focusing on the product – communication becomes easy. In Case 2, as the following quote illustrates, this open working culture specifically makes people feel more comfortable with accessing and using data. Thus, organisational culture is associated with developing BDAC, especially data management.

> "The company culture is simple. Everyone is doing things for the product. No one would say, 'This data is mine. You can't touch it.' I don't think there is any problem with communication, so I don't need to overthink 'people' problems. [About data management:] Just apply for permission. Whoever uses what data, as long as we keep the record [in the system], then it's all fine." (product manager, Case 2)

5.3.2 Basic resources & developing BDAC

Tangible resources – basic resources – financial resources

Companies invent products to make profits – and profits can also serve as the financial support influencing the existence of the product as well. During the outbreak of the pandemic, the company experienced a considerable impact. Advertisement income fell dramatically because the clients who paid for advertisements experienced shrinkage in their business. As a result, the project had to be stopped, thereby hindering the development of BDAC.

> "There was a dramatic drop in advertising revenue, which hugely impacted our income. Our business shrank, even if the number of users grew. We don't have income, so it was a pure expense. Operations and human resources are all costs, right? The whole business had to be shrunken after income dropped." (product manager A, Case 2)

5.3.3 Coordinating, communication skills & data variety

Dynamic capabilities – coordinating
Human skills – communication skills
Tangible resources – data variety

In Case 2, coordinating and communication skills are essential for collecting a variety of data sources because it is vital for the company to collaborate with content providers. In the past, the news app industry mainly relied on mass media institutions. All news apps collaborate with the same groups of media agencies, and thus receive similar content. Due to the need to stand out from the crowd, the company shifted its content towards emerging new media. New media and mass media content are the different sources of big data that have been categorised and matched with user preferences to provide personalised content. Thus, data variety also emerges here as a factor. To increase data variety, collaboration with these stakeholders is vital, which requires coordinating capability and communication skills.

> "In the beginning, I mainly cooperated with media institutions for content. For example, if I cooperated with these ten institutes, the content I got was the content on the official websites of these ten organisations. Our competitors also cooperated with these ten institutions. As a result, the homogeneity issue was very serious. So, we invited some other content creators and let them create different content, and we even talked about exclusiveness. In fact, all new apps like us talk about exclusive copyright content with content creators."
> (product manager B, Case 2)

5.3.4 Innovation process management & data management

Human skills – business knowledge – innovation process management
Human skills – technical knowledge – data management

Agile product development is a method to manage and guide innovation process management. According to the categories of resources introduced by a review on this topic (Mikalef, 2018), the resource type it fits into is "business knowledge". In agile product development, teams build products with short iterations that enable continuous feedback and rapid improvement. In Case 2, the product development team implemented agile product development with short iterations to quickly test with users and then decide whether to proceed with the prototype or not. During this period, considering the limited availability of time and especially human resources, attention has been focused only on the activities that the team thought necessary.

However, data management was neglected during this fast product development cycle, with data governance being implemented only when something went wrong. Thus, innovation process management (business knowledge) is associated with data management (technical knowledge).

> “When we start any product, we don’t have so many people, especially in the internet industry. We usually start small and quickly, launch online, and see users’ feedback, then decide whether to continue or not. So, in the beginning, we did it as quickly as possible until it developed to a certain stage. Problems often came out with data management (e.g. which format to store data to make it consistent and compatible for calculation), and then we started to manage it.” (product manager B, Case 2).

5.3.5 Data variety, business analytics & developing BDAC, product performance

Tangible resources – data variety
Human skills – business analytics
Innovation performance – product performance

5

Data and business analytics set the foundation for the project. In Case 2, the data sources comprise the content created by individual new media and traditional media institutions; data also come from the user’s daily usage of the app, such as browsing history. These data set the foundation for further business analytics to enable content categorisation, searching suggestions and personalised content feeding. Thus, business analytics, together with data, enabled the development of BDAC and good product performance.

> “Without big data, many things cannot be done. For example, the intelligent search mentioned earlier, how can it be intelligent without big data? Intelligence is the result of analysing a series of data, right? What can you analyse if you don’t have data and simply rely on an algorithm? I believe big data will gradually become a basic ability, and everyone will use it.” (Developer, Case 2)

5.3.6 Hardware & product performance

Tangible resources – hardware
Innovation – product performance

Hardware development, such as chip iterations, improves computer calculation speed, leading to cascading benefits, such as faster data processing, a stabler system, and potentially better business analytics that benefit product performance. The following quote illustrates this:

> "For example, if Qualcomm and Nvidia can produce better chips, and Intel can produce faster chips, it will have an impact on us. They can make our hardware environment better and help us improve our computing power. Improving our computing power will bring a series of chain reactions." (data scientist, Case 2)

5

5.3.7 Business analytics & product performance

Human skills – business analytics
Innovation performance – product performance

When content and customer preference data have been processed more frequently, and the recommendation algorithm (business analytics) has been run more often with increasing frequency, from once per day to once per hour to once per minute, the app can harness real-time analysis. Accordingly, the user's landing page on the new app would be refreshed for users who are always interested in breaking news and content to stay informed. Thus, business analytics (specifically, real-time analysis) is associated with product performance.

> "And then there is the recommendation algorithm. In fact, we can deal with the data closer to real-time, supported by this more real-time recommendation algorithm ... For example, you processed the data once per hour, and now you do it once per minute ... The result is a substantial improvement. For example, in the past, the algorithm generated a personalised user page that might be calculated one day in advance. Now, it might be that you get a new page every time you swipe it or open the app. The user experience will definitely be better than before." (data scientist, Case 2)

5.3.8 Data management & project performance

Human skills – technical knowledge – data management
Innovation performance – project performance

When data are first generated and collected, different people save them in different formats and different data systems. Saving data in this way makes data governance and consistent *data management* more complex, and people need to put in a lot of work and time. This issue slows down the project (time to market). Thus, data management is associated with project performance. As the product manager from Case 2 said:

> "You save the data in accordance with how the business wants to use them. In fact, there is no universal method that could quickly help any product complete the data processing. So, this is where I feel a headache now. Even if we have our data centre, there would still be a step of data management that would consume a lot of human resources." (product manager B, Case 2)


SAAS

CHAPTER 6

Case Study 3 – Big data in navigation

6.1 Introduction

This chapter presents the third case study, which concerns a Software as a Service (or SaaS) product development project. During this development project, the company collects and analyses navigation data to generate indicators in its SaaS product to assist petrol retailers in operational decision-making. Through analysing the data sources, the SaaS product presents information indicating the traffic situation near petrol stations in certain areas of the city. For research data collection, three people who were directly involved in the project were interviewed. In the case analysis, factors and relationships were identified by adopting the theoretical lens of the resource-based view (RBV) and dynamic capabilities (DC). This chapter starts with the case background, then follows the findings with the framework as an overview, followed by quotes and explanations.

6.2 Case background

Case 3 is a Software as a Service (SaaS) development project. During the pandemic, traffic came almost to a standstill due to the lockdown, and petrol stations struggled. When the lockdown was relaxed, petrol stations faced strong competition. To survive, companies need market information regarding the operation of competitors, for example, which stations are popular in a certain area and how many customers they have. To find out such information about the operations of competitors, petrol stations used to send people to observe them, such as by manually counting the number of visitors.

Now with the help of big data, petrol stations no longer need to send people to watch competitors but can use the big data-enabled product instead. The product provides operation indicators to assist petrol retailers in their daily operations.

It was initiated by one of the leading navigation providers in an Asian country. The company offers maps, navigation devices, and apps for businesses and individual customers. During this development project, the company collects and analyses navigation – or more precisely, origin-destination – data to generate indicators in the SaaS products to assist petrol retailers in their operational decision-making. Origin-destination data are also known as “flow” data, which describe the movement of people from one location to another (Office for National Statistics, 2022). Through analysing the data sources, petrol companies can get access to insights such as the traffic flow around the petrol stations in a certain area.

Initiated in October 2020, the project was relatively new, and a small group of four people were closely involved. To understand the development process of this project, a domain expert, a product manager, and a data analyst were interviewed. All three interviewees played essential roles in the project's planning, development, and implementation. Figure 6.1 below presents the main findings from this case. The following paragraphs explain them in detail.

6.3 Case 3 findings

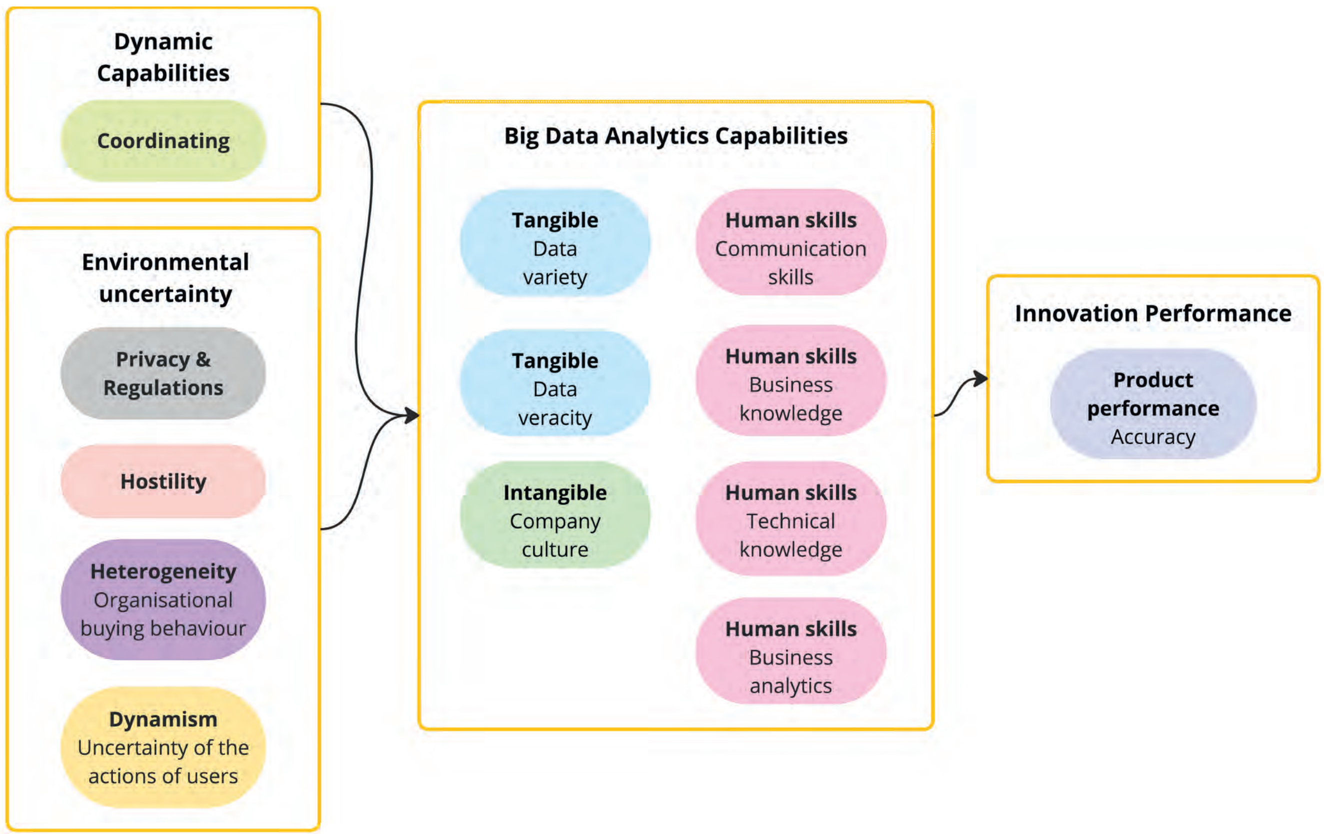


Figure 6.1 Conceptual model of antecedents and consequences of big data usage in innovation in Case 3. See Figure A3 in Appendix for a more detailed conceptual model.

6.3.1 Data, business analytics & BDAC

Tangible resources – data
Human skills – business analytics

In Case 3, the product development project got started because big data as a technology is ready for application in industry. This technology enabled a product that could improve operation for the clients – petrol suppliers. Big data as a technology is ready when both data as tangible resources and business analytics such as statistical analysis and predictive modelling are ready. This indicates that data and business analytics enable the development of BDAC.

> "It is because of big data that we could do the work of reducing costs and improving efficiency in various industries. Our strategic focus this year is energy, which is the oil industry. We could carry out data-driven strategic guidance [for clients] through big data-enabled methods." (product manager, Case 3)

6.3.2 Coordinating, communication skills, data variety , human skills, BDAC & business analytics

Dynamic capabilities – coordinating
Human skills – business analytics
Human skills – communication skills
Human skills - business knowledge
Human skills - technical knowledge
Tangible – data variety

Coordinating is a dynamic capability that focuses on orchestrating and deploying tasks and resources and synchronising with stakeholders (Mikalef & Pateli, 2017). Coordinating with stakeholders is not limited to different functions within the organisation but also extends to business partners. Coordinating also often emerged together with communication skills. In Case 3, coordinating capability and communication skills come to the fore, as collaborating with external stakeholders is vital for model validation and collecting data sources. For example, the company uses navigation data as input for a machine learning model to predict the number of visitors to a petrol station. So, collaborating with stakeholders (coordinating) is key for accessing data sources (variety) and improving data analytics (business analytics). Petrol station operation data, such as sales transaction figures, are vital for validating the model. Petrol stations do not share such transaction data with the navigation company; instead, the output from the model can be shared with the petrol stations to check whether the figures align with their sales data. Data variety is achieved not only with coordination but also with communication skills, because the company needs to communicate well with the petrol stations. Thus, coordinating and communication skills facilitate data variety, which then improves the quality of business analytics. In addition, the company worked with bench marking clients for initiating the product development project. This requires technical knowledge to verify the model, and business knowledge to do market and user research. From initiating the project, company also start developing its BDAC, thus both technical and business knowledge facilitates the development of BDAC. The quote below exemplifies this:

> "So, in this process, I have always said that the support of benchmarking clients is very important. Really, I think there are two core points for expanding the application of big data to clients. First, you must have data, and second, you must have true values and true values are validated when working well with clients ... we give answers with data, the customer tells us whether it is right or wrong, and then we come back to verify our model, and then constantly revise our model to come to a conclusion. We need to do verification like this in different scenes, different user profiling, and different labels in order to reach the final conclusion." (solution architect, Case 3)

6.3.3 Privacy & data variety

Environmental uncertainty – privacy & regulations
Tangible resources – data variety

The project of Case 3 was carried out in an Asian country where privacy is less of a concern and not as strictly regulated as in European countries. Potentially, this might change; the trend is that the regulation might get stricter. When this happens, certain data sources (data variety) might not be available for developing new products and services, which means there will be less data variety. Thus, privacy & regulations influence data variety. As the data scientist from Case 3 mentioned:

> "There might be some negative impacts coming from the social environment, that is, the privacy law and the protection of privacy; in fact, there are certain risks in products like ours that are based on user data. Whether this method can get support from the government remains to be seen – there is no problem now, but will it be restricted in the future, or will it be illegal? These kinds of factors exist." (data scientist, Case 3)

6.3.4 Hostility & developing BDAC

Environmental uncertainty – hostility

In 2020, the company encountered an external shock that brought competitiveness to the market. Due to COVID-19, there were fewer people on the road, and thus fewer petrol refills. Market need declined and competitiveness increased. This is environmental uncertainty caused by hostility. It represents the degree of threat posed to the company by the intensity of the competition and the downswings and upswings in the firm's principal industry (Miller and Friesen, 1978). Under this hostile environment, petrol supply companies look for new tools or technologies to inform their operation decisions, and in this case, big data is that technology. Companies develop BDAC when adopting big data for innovation. Thus, this case illustrates how hostility could facilitate the development of BDAC.

> "Because of the coronavirus, the pandemic has led to a decline in sales, and the decline in sales has caused their [petrol retailers] competition to become more intense. Everyone will compete for a small number of users. Once their sales decline, they need tools. Whereas in a rising market, they just need to expand on their own. They don't need to refine their operations." (domain expert, Case 3)

6.3.5 Heterogeneity & developing BDAC

Environmental uncertainty – heterogeneity - organisational buying behaviour

In Case 3, the company did not initially set out to develop a SaaS product for sale, but thought that the clients might need something like it. Through constantly communicating with business clients regarding their needs, and after completing many iterations, the product in its current form was developed. The drive to make the product came from realising that there was a need for the product – and not just from one client, but many. The product got initiated, designed, and developed thanks to the needs of these clients, and these organisations are willing to pay for the product once it is available to the market. This phenomenon is relevant to a variety of organisational buying behaviour (heterogeneity). Thus, heterogeneity is associated with developing BDAC.

> "We started to touch on the customer's intentions in December because it was not in the map product before. At that time, we estimated that the customer needed this product ... the first [consideration] is whether the needs are universal, and the second is whether there are any special

> needs that we have not seen. In recent years, it has been constantly changing through communication with users." (domain expert, Case 3)

As in all markets, some early adopters are the first ones to embrace new products. When several early adopters started using the big data-enabled SaaS product, they became the benchmarks for others to follow. Big data is not implemented until these clients start the projects.

> "This thing [the big data project] was originally relatively new, and not many people dared to use it. So, we need to build a benchmark from petrol retailers. We need to show that we are advanced, and we have the experience to attract more clients. For example, how much the sales of certain petrol stations have increased after using our products. To build benchmark cases and show that it goes well and has value." (domain expert, Case 3)

6.3.6 Company culture, organisational buying behaviour & BDAC

6

Before the existence of the product, most of the potential clients were not originally data-driven. In the past, most petrol stations used a very traditional method to get more information about the operation states of their competitors. They used to send people to observe other petrol stations. Now, they are becoming more data-driven – they purchase and use the product. The big data-enabled product would help those petrol stations understand their operations better. Thus, big data improves product performance. Data-driven culture influences the client's (organisation's) buying behaviour, and organisational buying behaviour facilitates BDAC development because it is all for the benefit of the clients. Big data enables the product and makes it possible to predict the potential number of customers at petrol stations. Furthermore, those clients who were more data-driven are using data from other partners for input and even building their own data platform. Still, adding data sources from the case company would help those clients who have their own data platform, because data sources provide different insights that could help the company gain more information about traffic.

> "… It used to be done in the traditional way. He [the client] went to the competitor's station and observed the scene. Using this method was very inefficient, so this could help them change their way of thinking. Now, some petrol stations have done a good job. For example, in some provinces and cities, petrol retailers and their business partners do not just use our products – they have their own big data platforms

> and internal systems that are accumulating data. They have already transformed from a more traditional way of operation. In fact, in addition to their own data, they also need our data, so a big data project was the best entry point for us to help our clients." (data scientist, Case 3)

Customers who are more data-driven and digitalised are more inclined to adopt big data-enabled products, thus facilitating the development of BDAC. Sometimes, the data-driven nature of their operations is evident in the way they give user feedback to companies.

> "From the breadth of the feedback and the depth of their thinking, it can be seen that the client has a relatively higher degree of informatisation than others, and the leaders are more open-minded. They are willing to accept big data." (data scientist, Case 3)

6.3.7 Dynamism & BDAC

Environmental uncertainty – dynamism – uncertainty of the actions of users

Whether the company can collect enough data sources (data variety) as input for the product or not depends mainly on the number of users of the app and their behaviour patterns. Data are available mainly in two ways (data variety): First, from using the navigation app and navigating to or driving past a petrol station. Second, from using other apps with the company's software development kits for location or navigation near the petrol station. However, such usage by users is unstable. Dynamism refers to the rate of change and innovation in the industry as well as the uncertainty or unpredictability of the actions of competitors and customers (Lawrence and Lorsch, 1967; Thompson, 1967; Burns and Stalker, 1961). So, dynamism (*uncertainty of the actions of users*) is associated with data variety.

> "Before you arrive at the petrol station, or when you are at the petrol station, you have used your mobile phone, and there is this app on this mobile phone. It calls our service or installs our brand's SDK, or it uses our map software itself, and then we can collect the user's data. But if he doesn't use a mobile phone or doesn't install our app and SDK, we won't be able to get his data. The biggest challenge now is that this is not a stable probability or ratio." (solution architect, Case 3)

6.3.8 Business analytics & product performance

Human skills – business analytics – algorithm development
Innovation performance – product performance – prediction accuracy

As the algorithm is the core technical part enabling the product, improvements to it result in the improvement of the product. Practitioners are continuously iterating the algorithm to aim at a prediction result that is as close to reality as possible. Algorithm development is part of data analytics, one of the human skills. Thus, data analytics is associated with product performance.

> "In terms of accuracy, the product is constantly iterated. We upgrade the algorithm and its technology, that is, the cutting-edge technologies we have to ensure that my algorithms are not biased or that the deviations are getting smaller and smaller [...] Basically, I am constantly iterating to get closer and closer to reality." (product manager, Case 3)

6.3.9 Data variety & product performance

Tangible resources – data variety
Innovation performance – product performance

Previously, petrol retailers have relied on their private domain data for operational decision-making. Navigation data could provide insights from a different perspective, for example, information on where most drivers are going. To help petrol stations to make better pricing decisions, it is not enough to merely have navigation data – extra price data are needed. So, the data sources complement each other. The product will not provide granular insights if it relies solely on either petrol stations or navigation data as a data source. So, data variety is positively associated with product performance.

> "How can I help him? I will give him more data and let him combine my data to make a more comprehensive decision. The decision can be at different levels of granularity. I'd love to tell you that your price must be reduced by two cents to be competitive. But I'm sorry, I don't have price data, so I can only tell you that it is better to make the price more competitive. I will provide you with suggestions based on where my data can be provided. If it needs to go deeper, additional information (or data) is necessary to do so." (domain expert, Case 3)

Data volume

Based on the analysis in the previous section, data-driven customers and the number of app users contribute to data volume. However, the study did not provide any evidence to support an association between data volume and innovation performance.

6.3.10 Data veracity & product performance (prediction accuracy)

Tangible resources – data – data veracity
Innovation performance – product performance

Data reliability is the data characteristic that emerged from Case 3. The data sources, such as the navigation route and road traffic information, are the input for the algorithms that enable the product, and set the foundation for the indicators in the final SaaS product. Small errors can lead to inaccurate prediction outputs. Thus, the reliability of these data is very important for the quality of the final product.

> "To put it bluntly, regarding the data sources used in this project, the bottom layer consists of these basic capabilities, our location service capabilities, our navigation planning capabilities, route algorithms, and the accuracy of road traffic information collection. These are all the underlying data that need to be tested. When the accuracy of our underlying data has reached a certain percentage after years of testing, then we will have the confidence to apply big data to the following scenarios." (data scientist, Case 3)


ALERTS

CHAPTER 7

Case Study 4 –
Big data in cybersecurity

7.1 Introduction

This chapter presents the fourth case study, which concerns a cybersecurity product development project. During this development project, the company collects and analyses cybersecurity operation data for models to learn and distinguish the patterns between normal operations and attacks or threats. For research data collection, three people who were directly involved in the project were interviewed. In the case analysis, factors and relationships are identified by adopting the theoretical lens of the resource-based view (RBV) and dynamic capabilities (DC). This chapter starts with the case background, then follows the findings with the framework as an overview, followed by quotes and explanations.

7.2 Case background

Case D is a cybersecurity product development project. The product can learn from previous cyberattack patterns to recognise attacks and potential threats to warn people. The product development project was launched by a cybersecurity company in an Asian country in 2016, following the trends in the market of using a new technology to enhance cybersecurity – big data. On the one hand, competitors started to try and implement new technology like big data analysis and machine learning to build cybersecurity systems. On the other, many big companies and institutes were demanding up-to-date cybersecurity systems. The project was initiated to address these issues with the leadership of top management from the company.

The product is a security system that protects clients' business systems. Historical cybersecurity operation data are collected to train machine learning models to recognise the pattern of threats and attacks and identify them better. For research data collection, three people were interviewed, including two developers and a product manager. They were all involved in the development or implementation of the project.

7.3 Case 4 findings

This section focuses on the internal and external influencing factors of BDAC along with its consequences identified from this research. Figure 7.1 below presents the main findings from this case. The following paragraphs explain them in detail.

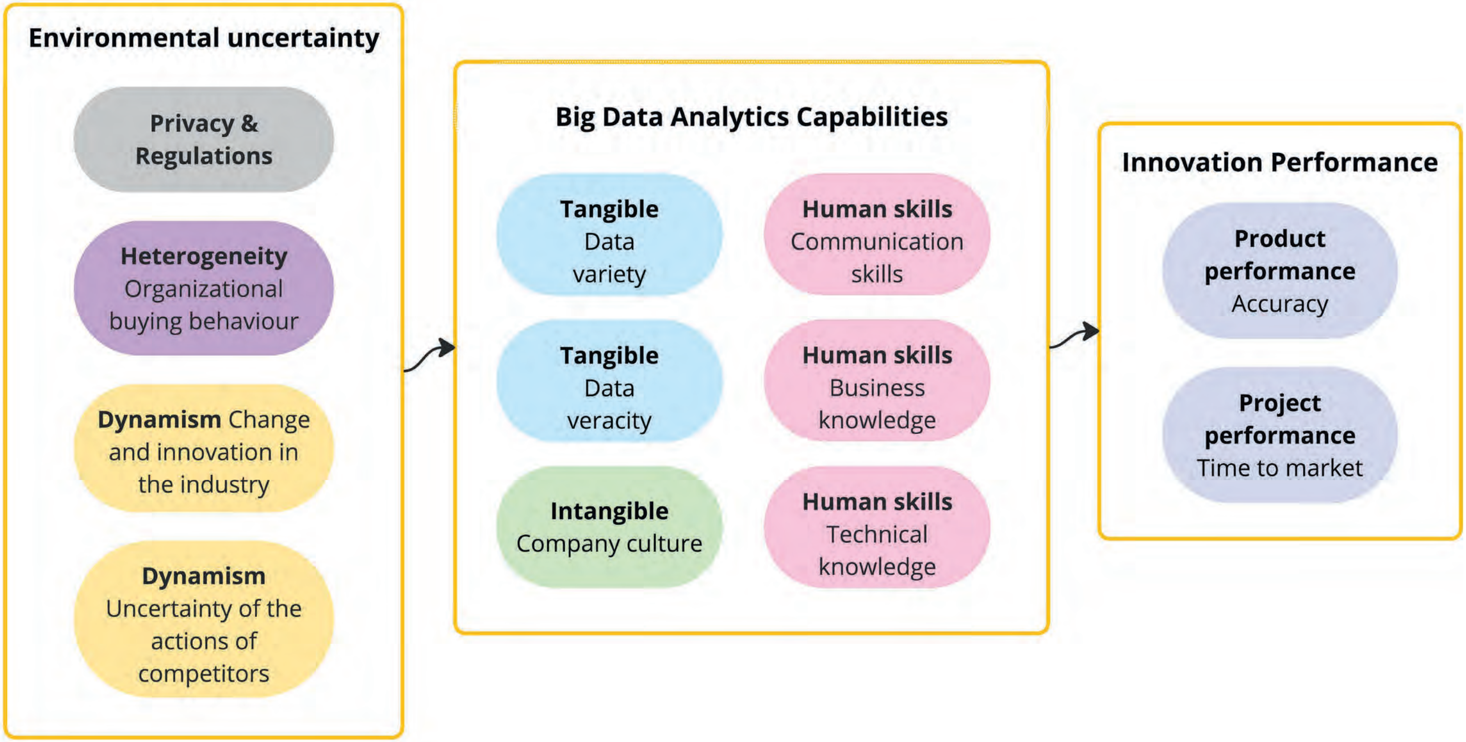


Figure 7.1 Conceptual model of antecedents and consequences of big data usage in innovation in Case 4. See Figure A4 in Appendix for a more detailed conceptual model.

7.3.1 Communication skills, technical knowledge & project performance

Human skills – relational knowledge – communication skills
Human skills – technical knowledge – technical infrastructure management
Human skills - business knowledge - market & user research
Innovation performance – project performance

During the development of the big data platform, it is vital to engage in communication with developers concerning detailed alignments, such as the processing rules and requirements for log data. Depending on the rules and requirements, developers can develop different levels of data storage volume and processing speeds. The rules and requirements for data processing are dependent on market trends and needs. For example, the data (log volume) storage scale in the market reflects the possible scale requirements of the company. Thus, *communication skills* are associated with technical infrastructure management. It is a time-consuming process, and is thus also associated with *project performance* – time to market. In addition, the business and technical knowledge together facilitates product performance.

> "The time-consuming part is that the log processing requirements we need here should be given to the developers of the database in advance, such as which fields we need, the length of the fields that our application requires; what kind of rules and requirements; and after we came back from market research, the size of the general log volume storage scale of the big data platform. All of these should be given to the developers who are building the platform. At least they can have a concept in their mind […] What is the final use scenario in this market? How big is it? There must be usage scenarios for this type of typical customer. Then, what are the requirements and specifications of this type of customer usage scenario? Market research must be done in advance." (product manager, Case 4)

7.3.2 Technical knowledge & product performance

Human skills – technical knowledge – technical infrastructure management
Innovation performance – product performance

The architect is an individual who designs information technology solutions (including IT architectures) and services for organisations. By designing the architecture, the IT architects play a very important role in technical infrastructure management – no matter how hard one tries, no code can save a bad (IT) architecture. Suitable (IT) architecture and good quality coding ultimately improve the quality of the product. So, *technical infrastructure management* is positively associated with *product performance*.

> "First of all, the architect must have a good understanding of the current architecture of the big data platform. He must choose an optimal architecture. If the architecture is wrong, then the following code will not be good for optimising the final performance, right? Design at the architectural level is very important. The second one is about our buddies who write the code, the quality of the code, and the code needs to be optimised to make the performance better." (product manager, Case 4)

7.3.3 Technical knowledge, business knowledge & product performance

Human skills – technical knowledge
Human skills – business knowledge
Innovation performance – product performance

Defining algorithm rules calls for not only technical knowledge but also business knowledge. One example is when dealing with brute force attacks, which involve trying all possible keys to defeat an encryption system. If no rules are defined for cybersecurity alarms, the attacker could theoretically try all possible keys and break into the system. However, if one defines overly strict rules for triggering the alarm, it may go off when normal users forget their password; while if the rules are too loose, the system could be attacked without anyone noticing. So, the cybersecurity expert needs to understand the business context of the security product. On the other hand, technical knowledge is also necessary for realising this task. The best practice comes from the cybersecurity expert's experience; they can better define the algorithm rules based on their knowledge of both technology and business, thereby improving product performance.

7

> "We have this platform, and we can write rules now. What should these rules look like? For example, if I enter the password three times, it will give an alarm, or if I enter the password five times, it will give an alarm. Take brute force cracking as an example. Your username is 'admin', and the password is '123456'. Then I will enter admin. Then I will enter a password randomly. I tried three times. Do you want to trigger the alarm? If you want to do that, then some people may run into a false positive when a real user enters the wrong passwords, right? There is another possibility. If I set the alarm for five errors,

then I will get 100 computers. I tried four times on each computer. That's still a brute force attack." (developer A, Case 4)

7.3.4 Data variety, technical knowledge & product performance

Tangible resources – data variety & completeness
Human skills – technical knowledge
Innovation performance – product performance

Defining the appropriate rules for cyberattacks is associated with the variety of data sources available and cybersecurity experts' technical knowledge. Data sources yielding different types of operation data can offer great insights to inspire the experts when they define the rules – for instance, user login scenario data and especially data that could reveal the various types of processes of brute force attacks. However, when data sources are incomplete or not available, it is necessary to rely on the knowledge that the cybersecurity experts have. The experts gain knowledge through trial and error. Thus, both data variety and cybersecurity experts' technical knowledge are associated with product performance.

> "This is statistics about an intrusion scenario. For example, I contacted more than 30,000 customers and found out that these users mostly entered their passwords accurately. You can just show an alarm after entering the wrong passwords three times. However, some users are not the same. But if someone can get the data from all users from all sources in our country in the scenario of logging in to the system, and get all the attack data – that is, the process records of brute force attack logging – then I can actually analyse it better, but I don't have that much information. You can't have so much data, so you must try your best using this process, and no one can give you an answer, that's the feeling." (developer A, Case 4)

7.3.5 Business knowledge, technical knowledge company culture & developing BDAC

Human skills – business knowledge
Human skills – technical knowledge
Intangible resources – company culture

The intention of using big data started with a gut feeling from the decision maker, so there was top management support (company culture). This was not just a question of business knowledge, but also decision-making instinct – the decision maker believed in implementing the technology. Then, the decision maker discussed its technical viability with technical experts and evaluated whether it would be possible to implement the technology in cybersecurity, especially in intrusion detection. On the one hand, this concerned the viability of the IT solution – a data platform that could store, process, and analyse big data – and on the other, the viability of training the machine learning model and extracting features. So, company culture along with both business knowledge and technology knowledge is associated with developing BDAC.

> "First, our leader, or technical decision maker, believed that this technology could be implemented now. Maybe he was not an expert in big data or machine learning algorithms, but he knew that there was this concept. Then he looked for people who understand the technology behind it. On the one hand, this involved building the big data platform. For that, we had to first find people who have got the know-how to discuss this matter. On the other hand, can anyone train this algorithm to extract features? We found a PhD at that time who participated in the early design phase. This process started from a business impulse that led us to find an expert in the corresponding technical field to have a small meeting and evaluate it." (developer A, Case 4)

7.3.6 Heterogeneity & developing BDAC

Environmental uncertainty – heterogeneity – organisational buying behaviour

In Case 4, because the product is based on big data with purchasing on a pay-as-you-go basis, buying the product is then associated with big data usage. *Heterogeneity* refers to the complexity and diversity of external factors, including the variety of customer buying habits. In this case, the customer is an organisation, and thus heterogeneity is embodied in *organisational buying behaviour*.

Three factors are identified from this case that are associated with *organisational buying behaviour*, namely, sufficient human resources, sufficient budgets, and the impact of cybersecurity on the clients.

7.3.6.1 Sufficient human resources & organisational buying behaviour

After the concept of "situation awareness" and its application in cybersecurity had been introduced to the market, many companies have started following the trends to implement it. It is vital for those client companies to have sufficient human resources to implement such large-scale projects for “situation awareness”. Different functions must work together and align with the supplier – to this end, a complete team is required, with functions such as bidding, procurement, safety, and operation. This is the reason why all clients for this project case are big companies – only such companies have the human resources to support implementation. Thus, clients with *sufficient human resources* are positively related to *organisational buying behaviour*, which is then associated with big data usage.

> “After the concept of situational awareness came out, every security company analysed their detection methods. Those cybersecurity companies found that if they wanted to implement this concept, the scale of the product would be very large. Only these large enterprises have the energy to accept this product. They have personnel, such as a technical team, a security team, an operation, and maintenance team, as well as related project managers, bidding personnel, and procurement departments. They are all available and have a relatively complete system.” (product manager, Case 4)

7.3.6.2 Sufficient budgets & organisational buying behaviour

In the case of a “situation awareness” cybersecurity project, the budget is one of the factors that can determine whether the project can be implemented from an early stage. Large budgets are needed because the product is huge and expensive to buy and implement. So, clients with big budgets are positively associated with *organisational buying behaviour*, which is associated with big data usage in innovation.

> “And they have a safety budget every year and are willing to spend money. And it is true that their business is very important, and they cannot afford to lose even a little bit of security. Then at this point, they have the ability and budget to set up projects of this size. So, they naturally become the so-called first battlefield of various security companies, a very important market for competition.” (product manager, Case 4)

7.3.6.3 The importance of the project for the client companies & organisational buying behaviour

Another related antecedent of acceptance by a client is how important cybersecurity is to that client. Companies can only spend a big amount if the project involves something with a big impact. The clients in this case are all big companies or organisations that play important roles in society, the economy, and the country. For these companies, small cybersecurity problems can lead to huge consequences for them or the country. So, these clients set up big budgets for cybersecurity and can afford large-scale projects like situation awareness projects. Thus, the importance of the project for the clients is positively associated with *organisational buying behaviour*, which is associated with big data usage in innovation.

> "These companies are not short of money. Their business is actually very important. If problems are discovered, they may lose their jobs. So, the budget for security is actually relatively loose. They are willing to even say that the first phase is not completed. Let's do a part of it first and see what effect it has. This is also possible. If they have this concept, they can actually start this kind of project. For such an enterprise that plays an important role in its country, their investment in security is actually very large, and if they find it useful, they don't care about the cost. A few million is not too much for them." (product manager, Case 4)

7.3.7 Dynamism (change and innovation in the industry) and developing BDAC

Environmental uncertainty – dynamism – change and innovation in the industry

Dynamism refers to the rate of change and innovation in the industry as well as the uncertainty or unpredictability of the actions of competitors and customers (Lawrence and Lorsch, 1967; Thompson, 1967; Burns and Stalker, 1961). The development of big data technology can be seen as one of the changes and innovations in the industry. It emerged as the origin of the project from Case 4 because it enabled association analysis. Ten years ago, when there was no big data, it was impossible to do association analysis for intrusion detection. Now, as technology has evolved, it makes it possible to realise association analysis and carry out the new project.

> "Ten years ago, when I was making this product, there was no big data technology; all database storage at this time relied on this traditional relational database. When we wanted to do this kind of association analysis, there was no way to achieve it. [...] Ten years later, with the technology of big data and the technology of Hadoop, we can realise this concept in order to discover new attack behaviours for defence. The trend of this new technology must have some promotion effect on this product, and also on technical attack and defence." (product manager, Case 4)

The other aspect of industry changes and innovation comes from the development of cloud technology. Before cloud technology existed, companies had to invest in multiple servers to run database calculations. Now, with cloud technology, they can deal with a greater volume of data, theoretically with no upper limits. So, the cloud as a technology solution enabled processing a higher volume of data and thus increased product performance.

> "To do this kind of big data calculation in the past, you may have had to buy a few – or even more than a dozen – servers. Now, using the cloud, the capacity can be unlimited or theoretically expanded all the time, so the computing power is stronger, the scope of data collection is also wider, and the amount of data is larger." (developer B, Case 4)

7.3.8 Dynamism (uncertainty of the actions of competitors) & developing BDAC

Environmental uncertainty – dynamism – uncertainty of the actions of competitors

Dynamism sometimes comes from the unpredictability or *uncertainty of the actions of competitors,* which also emerged in Case 4. There was a trend in the industry of adopting big data in cybersecurity product development; the company followed its competitors and started to use big data usage in innovation. Thus, the aspect of *dynamism* in terms of the *unpredictability of competitors* is also associated with *developing BDAC*.

> "It has originated from several aspects. When the project was established, we were inspired by the trends or the voices from the industry, from domestic and foreign manufacturers." (developer B, Case 4)

7.3.9 Data veracity & product performance

Tangible resources – data – veracity
Innovation performance – product performance

When the company first started building an algorithm, it used historical data as samples to train the model. However, the accuracy of the product was not at a satisfactory level after being implemented in the system. To address this, data generated from the current network system have been used to optimise the algorithm, and the result has improved. Data collected from the current network can better represent the current situation than historical data. This indicates that *data reliability* is associated with the accuracy of the product (*product performance*).

> “The false positive rate depends on whether the sample distribution is abundant or not. When we first started, there was a difference in the false positive and false negative rates when the algorithm was trained from the samples and when it was deployed and trained in a live network. The false positive rate and the false negative rate are still there and not low enough. Actually, when we collect certain data in the live network and then use this data to continue optimising the algorithm, the two indicators improve.” (developer B, Case 4)

7.3.10 Privacy regulations, data variety & product performance

Environmental uncertainty – policy and regulations
Tangible resources – data – variety
Innovation performance – product performance

In Case 4, the client’s data are used as the input to optimise the algorithm. So, access to these data sources is crucial, but sharing the data is not always possible. Some companies will not share their network operation data outside their internal network due to privacy protection concerns. The client’s data privacy has an influence on which data sources can be used and limiting the usable data sources can have an impact on algorithm quality. Data sources represent data variety, and algorithm quality affects product performance. Thus, *policy and regulations* influence the variety of data which then has an impact on *product performance.*

> “When clients use our product, we have a discussion. If they don’t want to give us their data for further training, that is fine. But we encourage our clients to bring us their own data to update our algorithms, there are all kinds of situations […] Especially in large enterprises, which are strict about this, some data cannot leave their intranet.” (developer B, Case 4)

CHAPTER 8

Cross-case analysis

8.1 Introduction

The cross-case analysis is the next step in the project, combining the resource-based view (RBV) and the four cases presented in Chapters 4, 5, 6, and 7 above. All interviews were first transcribed and then coded according to recommended practices across the four cases. The data analysis process has undergone several stages to serve the purpose of answering the research questions. In the first stage, the first order coding is conducted. The first order coding focuses on labelling the small pieces of qualitative data with codes and then grouping codes into categories (Corbin & Strauss, 2007). The categories are derived from the theoretical foundation of this research – dynamic capabilities (DC). For example, "uncertainty of the actions of competitors" and "change and innovation in the industry" were two factors that were grouped into the "dynamism" category, which belongs to "environmental uncertainty". Further, the second order coding is done in the second stage to seek similarities and differences between the codes, link categories, and connect data (Corbin & Strauss, 2007). The third stage mainly involved iterating between emerging data and categories and the relevant literature to identify whether the findings have precedents or discovered any new concepts (Gioia, Corley & Hamilton, 2012).

This chapter presents the findings from the cross-case analysis and looks at the company's environment both internally and externally. From the internal perspective, companies own various resources that can be categorised as tangible resources, intangible resources, and human skills (Grant, 1991). From the external perspective, looking outside of the organisation, the environment surrounding the organisation is full of uncertainty, ranging from dynamism and heterogeneity to hostility. All these factors could influence each other and build an environment that influences the use of big data in digital product innovation projects.

The findings are structured by the influencing factors of different dependent variables. Most of the factors are shared between at least two cases, and some emerged from three or even all four cases. Some factors only emerged in one case; they are discussed here because of their unique characteristics. The following section starts with explaining how the internal factors influence big data analytics capabilities (BDAC) and then explains how the external factors influence BDAC, how these factors influence data variety, and finally how these factors influence innovation performance. Factors are classified and defined, and the relationships between factors are explained, followed by propositions that describe the relationship between the dependent and independent variables. Finally, the differences between the cases in terms of the factors are compared. The propositions summarise the findings that are shared between at least two cases.

The relationships that stand out from only one case are discussed in this chapter but are not followed by a proposition.

8.2 Internal influencing factors of BDAC

This section focuses on the internal influencing factors of BDAC identified from this research, which include financial resources, hardware, company culture, technical knowledge, and business knowledge. These five factors belong to three categories: tangible resources, intangible resources, and human skills & knowledge. The relationships regarding the internal influencing factors of BDAC are shown in Figure 8.1 below, which highlights these factors and relationships. Factors and relationships that only emerged from one case are marked with * in the figure. Constructs within the square are the elements of BDAC. For example, tangible, intangible and human skills are the first-layer elements of BDAC. Under the first layer of elements, there are many second-layer elements as well. These constructs or elements are derived from the theoretical lens combining the resource-based view (RBV) and dynamic capabilities (DC) and observation from interview data in this research. These factors are both the elements of BDAC and the antecedents that can influence BDAC.

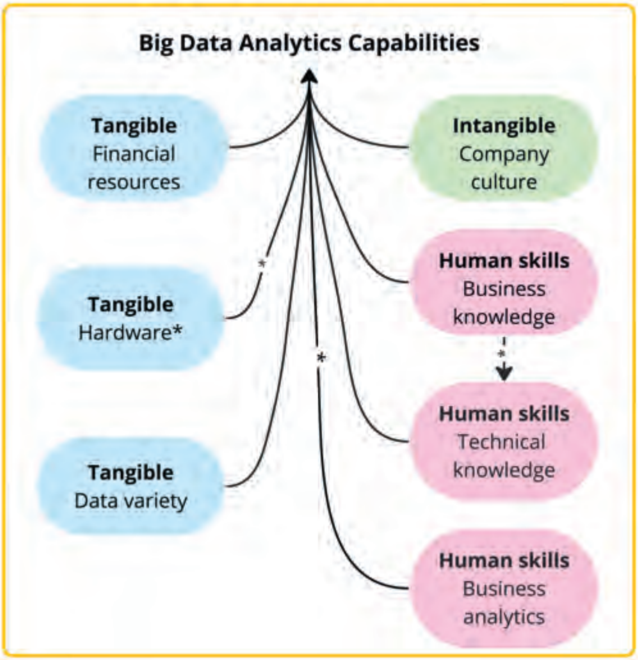


Figure 8.1 The relationships regarding the internal influencing factors of BDAC. *Presents factors and relationships identified from only one case.

8

8.2.1 Tangible resources & BDAC

Tangible resources are constructs of big data analytics capabilities. The tangible resources are resources that can be seen, touched and quantified (Kennedy, 2020). Basic resources (financial), hardware, and data are the tangible resources identified from this cross-case study.

8.2.1.2 Tangible resources – financial resources & BDAC

Financial resources comprise a basic resource required in every project, whether it uses big data or not. For big data projects, financial resources are very important – they cover the cost of data collection and data analysis. In this research, and especially in Case 2, financial resources emerge because when the resources are no longer available, the whole project stops (see quote below in the table). This is a real-life demonstration of comparing the consequences of a project with and without financial resources. Similarly, in Case 1, the project does not exist until funding has been granted, as financial resources are necessary for initiating the project. This indicates that financial resources are essential for developing BDAC.

Financial resources emerged in Cases 1 & 2 but did not stand out in Cases 3 & 4. The possible reason is that this depends on the financial decisions made by companies. Some companies might be stricter on funding projects, while some are less. Also, funding depends on the importance of the projects for the company. Important and promising product innovation projects are more likely to get funding.

Basic resources such as financial resources set the foundation for big data innovation projects. Without them, it would be impossible to operate such projects. This research indicates that they are indispensable resources, although financial resources stood out from research data collected from some cases but not others. Proposition 1 below reflects this relationship.

8

Proposition 1: Basic resources such as financial resources set the foundations and could facilitate the development of big data analytics capabilities.

8.2.1.3 Tangible resources – hardware & BDAC

Hardware is a tangible resource that sets the foundation for big data analytics. From data collection to data processing and analysis, every step of big data analytics relies on the hardware behind it. Table 8.1 below includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences. Like the data scientist in Case 2 mentioned (in Chapter 5), advanced chips with cutting-edge technology could potentially facilitate the development of big data analytics capabilities by increasing computing power. The quote in the table below illustrates that better hardware improves computing power. Big data and analytics skills from BDAC rely on computing power. Thus, hardware facilitates BDAC.

Table 8.1 Overview of financial resources and hardware

Tangible resources		
Construct	**Financial resources**	**Hardware**
Definition	The funds and assets that support the company's operations.	The machines and electronic parts in a computer or other electronic system (Hornby, 2020).
Proof quotes	*"There was a dramatic drop in advertising revenue, which hugely impacted our income. Our business shrank, even if the number of users grew. We don't have income, so it was a pure expense. Operations and human resources are all costs, right? The whole business had to be shrunken after income dropped." (product manager A, Case 2).*	*"For example, if Qualcomm and Nvidia can produce better chips, and if Intel can produce faster chips, it will have an impact on us. They can make our hardware environment better and help us improve our computing power. Improving our computing power will bring a series of chain reactions." (data scientist, Case 2)*
Link to the extant literature	Valuable, rare, inimitable, and non-substitutable (VRIN) resources can contribute to sustainable competitive advantages (Amit and Schoemaker, 1993; Barney, 1991). Intangible resources are considered more VRIN (Bruneel, Ratinho, Clarysse & Groen, 2012; Gimmon & Levie, 2010).	
Differences from the literature	The study provides empirical evidence illustrating the importance of financial resources on a project level; although they are not VRIN, such resources can determine the existence of a project.	The study provides empirical evidence illustrating the importance of hardware in terms of facilitating the development of BDAC.

Hardware as a component of a tangible resource only clearly stands out in Case 2, but not the other cases. However, this does not mean that hardware is not necessary in Cases 1, 3 and 4. Perhaps because it is something that sets the foundations when it exists, people do not notice the support that hardware provides, but implementing big data innovation projects would be impossible without it.

8.2.1.4 Tangible resources – data & BDAC

Data also comprise a tangible resource and an element of BDAC. This research identified three aspects of data, namely, data variety, velocity, and reliability. The cross-case analysis findings regarding data mainly concern their influence on innovation performance. Thus, data are discussed in the "influencing factors of product performance" part of this chapter.

8.2.2 Intangible resources & BDAC

Intangible resources are other constructs of big data analytics capabilities parallel to tangible resources. Intangible resources are resources that are relatively difficult to see, touch and quantify (Kennedy, 2020). Company culture is one of the intangible resources identified in three out of four of the cases.

8.2.2.1 Intangible resources – company culture & BDAC

Company culture is the shared ethos of the company. Company culture is viewed as the widely shared assumptions and values within an organisation that give rise to typical behaviour patterns (Gordon, 1991). Table 8.2 below includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences.

Table 8.2 Overview of company culture

Intangible resources	
Construct	**Company culture**
Definition	Widely shared assumptions and values within an organisation that give rise to typical behaviour patterns (Gordon, 1991).
Proof quotes	*"The previous operations were very extensive. Everyone did what they thought was right without relying on any theoretical foundation, and then just let the market verify it. Thanks to big data, we can analyse before making decisions … Now, we have slowly become more and more inseparable from the ability of big data in terms of development, function, or R&D. We will no longer implement some requirements without a data foundation. We need to get the foundation of data before we can do such a thing." (developer, Case 2)*
Link to the extant literature	Data-driven business model (Hartmann, Zaki, Feldmann & Neely, 2014).
Differences from the literature	The literature on business models sees being data-driven as a key resource to rely on. The findings revealed that companies do not rely solely on being data-driven as a resource. Companies that are data-driven would rely more on data usage, and as they do so, they became more data-driven in culture. So, harnessing data usage and data-driven culture reinforce each other.

Company culture emerged from the interview data. Take data-driven culture as an example. Companies with a data-driven culture use data as a reference when making decisions. To quote one of the respondents: *"Our whole society is generating so much data every day, every second, and it's rationalised our decisions when you use data to make your decisions. Instead of just having a feeling that it would be wise to do something, we are trying as a company – or going on a journey, actually – to become more data-driven, so every solution should involve, should use data to build on, actually" (programme coordinator, Case 1).* The company saw big data as a resource to rationalise decision-making. With more and more data becoming available for use, companies are increasingly aware of their value in decision-making. Data-driven decision-making is preferred over decisions based on experience. Companies become more data-driven and use more big data, thus facilitating the development of BDAC.

This finding emerged from Cases 1, 2 and 4. In Case 1, the development and functional decision-making depend on data, which reflects the influence of data-driven culture on the company's development of BDAC. In addition, the project was initiated thanks to top-management support, which is another aspect of company culture. Case 2 shows that an open culture can form an environment where data can be shared and used freely without being hindered by personal issues. Case 4 shows us that top management support and decision-making instinct are associated with initiating a project to harness big data, and the company develops BDAC as the project proceeds. All three cases suggest that company culture can facilitate the development of BDAC.

In Case 3, company culture did not stand out from the interview data as an antecedent. A possible reason for this is that the company uses data as the foundation of their business. The company's main business is navigation, relying heavily on various location data sources. The company is rooted in data, and has a data-driven culture as the norm, and thus its company culture did not stand out from the findings.

Existing literature considers one type of company culture, data-driven culture, as a construct of BDAC (Mikalef et al., 2019). The current research extends the findings to company culture in general, including data-driven culture, open culture, top-management support, etc. In addition to data-driven culture, other different aspects of company culture also facilitate BDAC. Proposition 2 reflects this relationship.

> **Proposition 2: Company culture facilitates the development of big data analytics capabilities.**

8.2.3 Human skills and knowledge & BDAC

A company's ability to utilise big data technology for innovation is highly dependent on its human skills and knowledge. This research identified two elements of human skills and knowledge: business knowledge and technical knowledge. It also discovered how business knowledge and technical knowledge influence each other and BDAC.

8.2.3.1 Human skills and knowledge – business knowledge and technical knowledge & BDAC

Technical knowledge and business knowledge are both elements of the human skills and knowledge category. Table 8.3 below includes their definitions and proof quotes from this research, then links them with extant literature and discusses the differences.

Table 8.3 Overview of technical knowledge & business knowledge

Human skills & knowledge	
Construct	**Technical knowledge & business knowledge**
Definition	**Technical knowledge:** The know-hows required to handle the technological components to extract value from big data (Gupta & George, 2016). **Business knowledge:** Recognising the value of big data and understanding where to make efforts to gain insights (Akter and Wamba, 2016).
Proof quotes	"First, our leader, or technical decision maker, believed that this technology could be implemented now. Maybe he was not an expert in big data or machine learning algorithms, but he knew that there was this concept. Then he looked for people who understand the technology behind it. On the one hand, this involved building the big data platform. For that, we had to first find people who have got the know-how to discuss this matter. On the other hand, can anyone train this algorithm to extract features? We found a PhD at that time who participated in the early design phase. This process started from a business impulse that led us to find an expert in the corresponding technical field to have a small meeting and evaluate it." (developer A, Case 4)
Link to the extant literature	Human skills (Mikalef et al., 2019)
Differences from the literature	The research identified the contribution of technical and business knowledge to BDAC and delves deeper into how they contribute to BDAC. The findings suggest that technical expertise and business knowledge often appear together in the initiation of big data projects and are equally valuable in assessing feasibility.

8

When developing big data analytics capabilities, both technical and business knowledge are necessary. In Case 4, the big data-enabled innovation project started from the decision maker's instinct – the person believed that it was feasible to apply big data in the cybersecurity project. This instinct came from the decision-maker's experience and knowledge in business. On the other hand, technical knowledge also contributed to it. An expert with relevant technical knowledge was involved in evaluating the feasibility from the technical point of view in terms of training the machine learning model and extracting features. The quote in the table above from Case 4 illustrates this. Proposition 3 reflects this phenomenon.

Proposition 3: Both technical and business knowledge are needed in initiating big data analytics capabilities development.

Technical knowledge comes up in all the cases, in different ways, for example, as data management (Case 1) or algorithm design skills (Case 4). Technical knowledge is a construct of human skills (Mikalef et al., 2019). The research follows the same theoretical lens and recognises it as part of human skills. It contributes to the field by finding that it is often human skills and intangible resources together that make the difference in initiating big data analytics capabilities.

Technical knowledge and business knowledge are resources under the category of human skills as part of BDAC (Mikalef, et al., 2019). The current research took a step further and illustrated how they contribute to BDAC. It was discovered that technical knowledge and business knowledge often show up together when initiating big data projects. Both types of knowledge are useful in evaluating feasibility. This phenomenon is observed across Cases 1, 3 and 4 but is not prominent in Case 2. In Case 2, the company was a player that was a later entrant in the market. The business and technical feasibility of the product had already been proven to work by other companies. The company then did not have to pay too much attention to early feasibility evaluation, and thus business and technical knowledge did not stand out in this case. Stress and competition from the market are stronger forces pushing the company to do the same.

In addition, business knowledge as a factor stood out in Case 2, and the focus was on innovation process management. The following part discusses the findings in more detail.

8.2.3.2 Business knowledge – innovation process management and technical knowledge – data management

According to the classification applied in this research (Mikalef et al., 2019), innovation process management falls under the category of business knowledge. In a fast-paced industry such as the internet industry, innovation processes often move very quickly. Although not all companies with short iterations implement efficient and effective agile processes, most implement fast innovation process management schemes. In Case 2, the company mainly provides a news app to its market, which could be categorised as belonging to the internet industry. In Case 2, the innovation process runs short iterations, quickly tests prototypes with users and then decides whether to proceed with the version or not (see quote in Chapter 5). During the process, considering the fast pace driven by competitors in the industry, attention has only been put on the necessary and urgent aspects. Things like data management, which format and way to store the data were not priorities for the innovation team in the early few iterations. Later on, the team realised that the inconvenience of data management consequences could be influenced by the fast pace, pushing innovation process management. Thus, *innovation process management* could influence *data management (technical knowledge)*, and it could have a negative effect if the management plan does not consider *data management* from a long-term perspective.

> "When we start any product, we don't have so many people, especially in the internet industry, we usually start small and quickly, launch online, and see users' feedback, then decide whether to continue or not. So, in the beginning, we did it as quickly as possible, until it developed to a certain stage. Problems often came out with data management, and then we started to manage it." (product manager B, Case 2)

This phenomenon did not seem to stand out in the other three cases; a possible reason is that in the internet industry, innovation iteration is exceptionally fast-paced, thus making the effect on data management more noticeable. The other cases involve the fields of transportation, navigation and cybersecurity, in which the pace could also be fast but perhaps not as fast as in the internet industry.

8.2.4 Summary – internal influencing factors of BDAC

To summarise, this research identified several BDAC elements that could influence the development of BDAC. These elements can be put under the categories of tangible resources, intangible resources, and human skills. Tangible resources – data, financial resources, and hardware – set the foundations and could facilitate the development of BDAC. Intangible resources – company culture – facilitate the development of big data analytics capabilities. Human skills & knowledge – both technical and business knowledge – are needed in initiating big data analytics capabilities development. In addition, *innovation process management* could influence *data management (technical knowledge)*, and it could have a negative effect if the management plan does not consider *data management* from a long-term perspective.

8.3 External influencing factors of BDAC

This research has identified four factors that influence BDAC externally, which are heterogeneity, dynamism, hostility, and privacy & regulations. Figure 8.2 below illustrates these factors and relationships. Factors and relationships that only emerged from one case are marked with * in the figure.

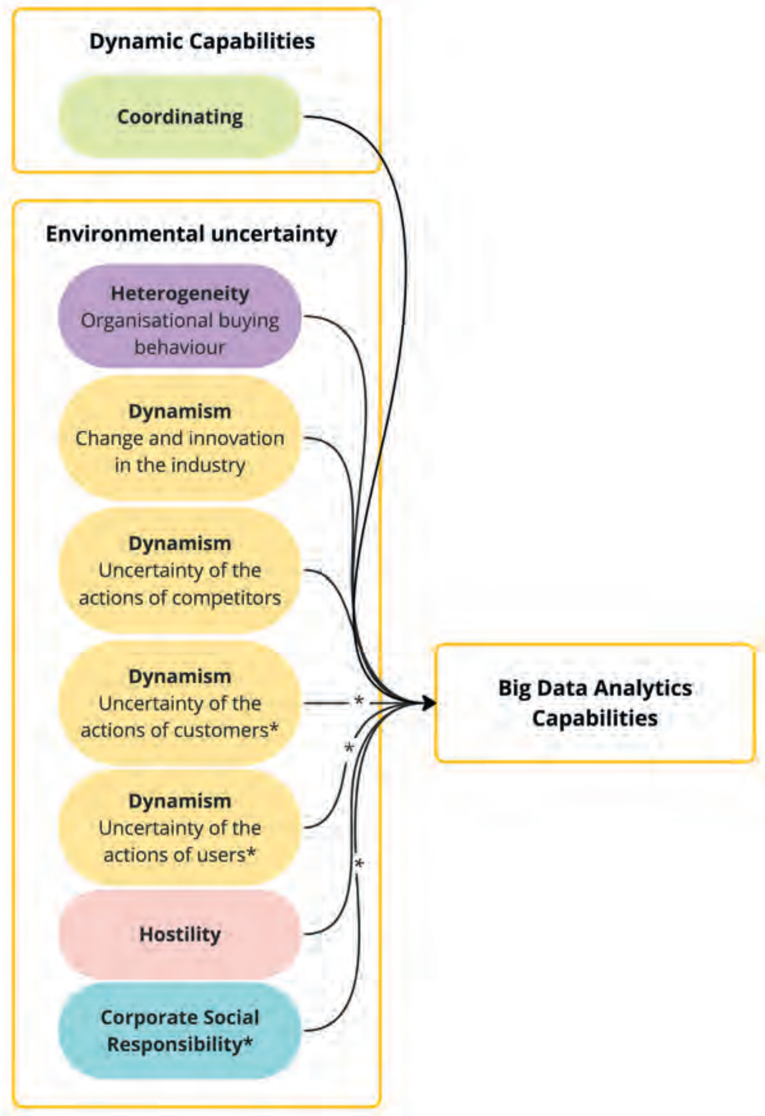


Figure 8.2 The relationships regarding the external influencing factors of BDAC.
*Presents factors and relationships identified from only one case.

8.3.1 Heterogeneity – organisational buying behaviour & BDAC

Heterogeneity is an aspect of environmental uncertainty; this research adopted the definition of heterogeneity from Newkirk and Lederer (2006): "Heterogeneity is the complexity and diversity of external factors. It has been operationalized in terms of diversity in customers' buying habits, diversity in the nature of competition, and diversity in product lines" (p. 483). Based on this definition, organisational buying behaviour (OBB) is part of heterogeneity, because it concerns the decision-making process whereby an organisation purchases products for its operations. This research identified OBB as an element or construct of heterogeneity and discovered how it has influenced BDAC. Table 8.4 below includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences. Regarding the other aspects of heterogeneity, diversity in product lines and competition did not emerge from this research probably because the unit of study is a single product development project from each company, and thus this research does not consider product lines.

Table 8.4 Organisational buying behaviour

Environmental uncertainty – heterogeneity – organisational buying behaviour	
Construct	Organisational buying behaviour (OBB)
Definition	A decision-making process carried out by individuals in an organisation to purchase the products and services needed for their operations (Webster & Wind, 1972). This research uses OBB to represent the buying behaviour of the clients of case study companies.
Proof quotes	*"First, they are more interested in this new technology and more willing to delve into it – they are willing to try new things. This is an obvious characteristic of our customers. Second, customers must have experience in this area. Most of the customers who done deep research on safety have used a lot of traditional safety equipment and have their own unique views and opinions. If the new product can impress them, they are still willing to give it a try. Then the customer must have the budget and money in this area, and this must be discussed clearly with the customer." (developer B, Case 4)*
Link to the extant literature	Customer adoption in mobile payment (Oliveira, Thomas, Baptista & Campos, 2016).
Differences from the literature	In mobile payment, individual customer adoption is influenced by perceived technology security. The study explored the buying habits of business customers, and discovered that apart from openness for innovation, their habits are influenced by perceived technology value and budget.

Organisational buying behaviour (OBB) plays a vital role in developing big data analytics capabilities. This study regards OBB as the buying behaviour of the clients of case study companies. In the two B2B cases, organisational buying behaviour facilitates the development of BDAC, which is illustrated in detail in the following paragraphs.

This research identified different aspects of organisational buying behaviour from the clients of the case study companies. Big data-enabled products or services are often customised for clients, and their buying behaviour is often associated with the service provider developing BDAC in big data innovation projects. First, customers who would purchase big data-enabled innovation are willing to embrace newly developed technologies. Second, the customers are experts in the field who have the experience and knowledge required to understand the value of big data-enabled innovation. Finally, it is essential to have a budget for the new technology, as the developer summarised with a quote in Table 8.4 – Organisational buying behaviour. When clients become more data-driven, they are open to embracing more data sources and volume: *"It used to be done in the traditional way. He [the client] went to the competitor's station and observed the scene. Using this method was very inefficient, so this could help them change their way of thinking. Now, some petrol stations have done a good job. For example, in some provinces and cities, petrol retailers and their business partners do not just use our products – they have their own big data platforms and internal systems that are accumulating data. They have already transformed from a more traditional way of operation. In fact, in addition to their data, they also need our data, so a big data project was the best entry point for us to help our clients"* (Case 3). The clients were already open to the new technology; they had their data platforms and aggregated data; as they were getting more data-driven, they needed external data sources, so the case company got the opportunity to get involved and provide the innovation service. Proposition 4 reflects this phenomenon.

Proposition 4: Heterogeneity (organisational buying behaviour) is positively associated with developing big data analytics capabilities, because customers who are open to new technologies, are experts in the field, or have a budget for the new technology are willing to purchase big data-enabled products or services.

The factor regarding buying behaviour emerged from the two B2B cases (3 & 4) but not from the B2C ones. A possible reason may be that although customers can provide a data source for both B2B and B2C projects, it is easier to attract B2C customers to use the product, thus providing useful data. That said, this is more challenging in the case of B2B customers – trust needs to be built and the product needs to prove itself.

In the literature, a relevant related topic is big data analysis acquisition. Current literature regarding antecedents of big data analysis acquisition mainly focuses on aspects within the organisation (Kwon, Lee & Shin, 2014). The findings supplement the literature by looking outside of the organisation and introducing the role of the customer. This research provides empirical evidence showing the positive association between organisational buying behaviour and developing BDAC.

8.3.2 Dynamism & BDAC

Dynamism is a variable of environmental uncertainty. Dynamism refers to "the rate and unpredictability of environmental change" (Newkirk, Lederer, 2006, p. 482). To better operationalise this concept, this research adopted the more specific definition: "the rate of change and innovation in the industry as well as the uncertainty or unpredictability of the actions of competitors and customers" (Lawrence and Lorsch, 1967; Thompson, 1967; Burns and Stalker, 1961). The following section includes the findings regarding the different aspects of dynamism: *change & innovation in the industry, uncertainty of the actions of competitors, uncertainty of the actions of customers, and uncertainty of the actions of users.*

8.3.2.1 Dynamism – change and innovation in the industry & BDAC

Change and innovation in the industry is an aspect of dynamism. Table 8.5 below presents the definition and proof quotes, and links them to the extant literature.

Table 8.5 Dynamism – change and innovation in the industry

Environmental uncertainty	
Construct	Dynamism – change & innovation in the industry
Definition	The rate of change and innovation in the industry as well as the uncertainty or unpredictability of the actions of competitors and customers (Lawrence and Lorsch, 1967; Thompson, 1967; Burns and Stalker, 1961).
Proof quotes	*"... the development of big data, data solutions are growing very fast, and developing very fast. The risk of a disruptive innovation like self-driving cars is a very real risk for us. So, despite our risk-averse nature, we have to be open to these kinds of innovations and the use of big data to stay competitive in this very fast-changing world." (programme coordinator, Case 1)*
Link to the extant literature	Managing technology development projects (Cooper, 2006).
Differences from the literature	The finding aligns with the current studies and addresses technology development as part of innovation in the industry, enabling companies to gain competitive advantages.

Change and innovation in the industry emerged in all four cases and are embodied in the development of new technology, specifically technology such as big data, cloud and navigation. From the point of view of the public transport provider, new technology such as autonomous driving will become a threat that may claim market share from public transport (see quote in the table above). Merely focusing on its current business will not help the company grow.

When asked about the project's origins, developers from different companies mentioned the development of technology. In Case 4, change and innovation such as cloud technology facilitate big data analytics capabilities and increase the data volume: *"Now, using the cloud, the capacity can be unlimited or theoretically expanded all the time, so the computing power is stronger, the scope of data collection is also wider, and the amount of data is larger."* (developer B, Case 4). The technology behind big data sets the foundation for developing big data analytics capabilities: *"Where does the foundation of big data come from? It comes from the capabilities of big data. Big data helps you analyse and do a series of calculations"* (data scientist, Case 2). The development of new technologies such as increased computing power and the cloud set the foundation for developing big data analytics capabilities in product development projects. Without corresponding computing power, it is impossible to use big data as a technology because dealing with such amounts of data requires big computing power. As computing power improves, more data can be dealt with, and along with this process, big data analytics capabilities are developed.

The foundation of developing big data analytics capabilities (BDAC) is the technical knowledge behind it: *"To put it bluntly, regarding the data sources used in this project, the bottom layer consists of these basic capabilities, our location service capabilities, our navigation planning capabilities, route algorithms, and the accuracy of road traffic information collection. These are all the underlying data that need to be tested. When the accuracy of our underlying data has reached a certain percentage after years of testing, then we will have the confidence to apply big data to the following scenarios"* (data analyst, Case 3). In Case 3, big data-enabled innovation was ready to be implemented only when the navigation technology and transportation data collection were prepared to support it. So, technologies for data analysis foster BDAC in digital product development. Proposition 5 reflects this phenomenon.

8

> **Proposition 5: Dynamism (change and innovation in the industry), along with technical knowledge, are facilitating the development of big data analytics capabilities.**

Change and innovation in the industry as one aspect of dynamism is prominent across all four cases. The research contributes to the literature by proposing that it is dynamism (change in innovation in the industry) together with technical knowledge that makes a difference in developing BDAC.

8.3.2.2 Dynamism – uncertainty of the actions of competitors & BDAC

Uncertainty of the actions of competitors is one aspect of dynamism (Lawrence & Lorsch, 1967; Thompson, 1967; Burns & Stalker, 1961). Table 8.6 below includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences.

Table 8.6 Overview of dynamism – uncertainty of the actions of competitors

Environmental uncertainty	
Construct	Dynamism – uncertainty of the actions of competitors
Definition	The rate of change and innovation in the industry as well as the uncertainty or unpredictability of the actions of competitors and customers (Lawrence and Lorsch, 1967; Thompson, 1967; Burns and Stalker, 1961).
Proof quotes	"It has originated from several aspects. When the project was established, we were inspired by the trends or the voices from the industry, from domestic and foreign manufacturers." (developer B, Case 4)
Link to the extant literature	Competitor actions (Jauch & Kraft, 1986).
Differences from the literature	The study contributes to the literature by finding the positive role of dynamism (uncertainty of the actions of competitors) on developing BDAC.

When facing an uncertain environment, the reactions of competitors are unpredictable. Market trends are formed when most of the competitors are doing something similar. What is observed across the cases is that companies follow the market trends to use new technologies such as big data. If the leading players in the market start using the latest technology, others will follow in their footsteps, as quoted in the table above.

8

When the majority of the competitors in the market adopt new technologies, many will follow. Although it is hard to predict what action competitors will take, these actions influence the company's decision to set up big data-based innovation projects. While operating these projects, companies develop their big data analytics capabilities. Thus, it is proposed:

> **Proposition 6: Dynamism (uncertainty of the actions of competitors) stimulates the development of big data analytics capabilities.**

This phenomenon is observed across all four cases. In Case 1, involving a railway transport company, unpredictable competitors are introducing autonomously driving cars, which stimulates the company to innovate with new technology like big data to improve their service to stay competitive. In Case 2, the trend of big data-enabled personalised news services has stimulated the case company to do the same. In Case 3, in the oil industry, where products on the market are highly similar, competitors' actions such as marketing campaigns can have a big influence on the market; therefore, to gain better market knowledge, big data-enabled products are needed, for example, to obtain information on traffic levels around competing petrol stations. In Case 4, the market trend was to use big data for cybersecurity and many competitors were exploring the possibilities to do so, and the case company decided to do so as well.

Previous research found that dynamism has a moderating role. Specifically, dynamism positively moderates dynamic capabilities' effect on incremental innovation (Mikalef et al., 2019). However, the current studies did not test the effect of dynamism on BDAC. The research found that dynamism has a positive role that stimulates the development of BDAC.

8.3.2.3 Dynamism – uncertainty of the actions of customers & BDAC

Uncertainty about the actions of customers can influence the development of BDAC, sometimes hindering it. In Case 2, the scenario was that the clients could not continue to do business with the case company during Covid (see quote in Chapter 5). The project had to be stopped – and thus the development of BDAC was hugely influenced by a lack of financial income and the uncertainty of the actions of customers. It is hard to predict what actions customers will take in a fast-changing uncertain world.

Uncertainty about the actions of customers only emerged in Case 2, and not in the other cases. A possible explanation for that is that the project stopped due to lack of financial income caused by the negative consequences of uncertainty; in Cases 1, 3 and 4, there were no such negative consequences, and the projects went well. People pay more attention to negative consequences and thus mention them in interviews.

8.3.2.4 Dynamism – uncertainty of the actions of users & BDAC

Users can be different from customers. Customers pay for products, while users use products. Most of the time, paying customers also use the product themselves; however, sometimes customers pay for the product for someone else to use it. The actions of users could affect data variety and volume. In Case 3, the input data for the product come mainly from two sources, both depending on the actions of users.

First, the data come from users who are close to the petrol station using the app for navigation. Second, the data also come from users who are using other apps with location functions near the petrol stations, because the location function is also realised through the company's software development kits. However, the usage of the app or relevant apps is unstable because people might not use any app to navigate to the petrol station. The actions of the users are not easy to predict, which influences the accessibility of these user data sources and the amounts of data acquired. The quote from Chapter 6 illustrates this.

> "Before you arrive at the petrol station, or when you are at the petrol station, you have used your mobile phone, and there is this app on this mobile phone. It calls our service or installs our brand's SDK, or it uses our map software itself, and then we can collect the user's data. But if he doesn't use a mobile phone or doesn't install our app and SDK, we won't be able to get his data. The biggest challenge now is that this is not a stable probability or ratio." (solution architect, Case 3)

This factor did not show up in Cases 1, 2 and 4, but that does not mean it has no effects on data variety and volume in these cases. The possible reason it stands out in Case 3 is that the users are not necessarily using the apps at petrol stations, which influences data variety and volume. In the other cases, there is a much higher chance that the users are in fact using the case company's app – on public transport when checking in and checking out or reading the news. In Case 4, the users are also the buyers of the cybersecurity system, and thus it would not make sense for them not to use the product.

8.3.3 Hostility & BDAC

Hostility is another aspect of environmental uncertainty. Table 8.7 below presents the definition and proof quotes, and links it to the extant literature. Hostility refers to the availability of resources and the degree of competition in the external environment. Hostility can impact the usage of big data in innovation projects – specifically, product differentiation and the degree of competition in the market.

The degree of competition in the external market is embodied in the research data. In Case 1, for example, respondents considered big data as one of the primary resources to help the company stay competitive. For public transport operators, alternative transport solutions are upcoming threats in the fast-changing environment. As the programme coordinator mentioned: *"I think the main motivation for this project – and also other projects – is that big data, and data in general, represent one of the main sources for staying in the competition, actually ..."* (programme coordinator, Case 1).

Table 8.7 Overview of hostility

Environmental uncertainty	
Construct	Hostility
Definition	"Hostility involves both the availability of resources and the degree of competition in the external environment. Researchers have measured it in terms of the threats posed by labour and material scarcity, price and product-quality competition, and product differentiation." (Newkirk, Lederer, p. 483)
Proof quotes	*"Why do I have to work hard when I can 'lie flat' and still make a profit? When the market goes down, the competition is tough. When the competition is tough, they look for tools. And that's what we are offering. The pandemic has caused a decline in overall sales in the petrol retail business, but it actually positively impacts our business." (domain expert, Case 3)*
Link to the extant literature	Porter's five forces (Mullins & Walker, 2012)
Differences from the literature	There are four forces that can increase rivalry among existing competitors. Among them, this research provides empirical evidence for the bargaining power of buyers and the threats posed by substitute products.

Another related quote from Case 3: *"I think the impact of the coronavirus on our product is positive. It has caused a sharp decline in travel. Big companies in the oil industry have also deeply felt the crisis. Because of the crisis, they were willing to embrace new technologies, data, and products to strengthen themselves and be more competitive. The crisis let them see the danger and the downhill trend. I think this is vital. Otherwise, even if we have a good product, but the customers are living comfortably and ignore the product, it will not work"* (solution architect, Case 3). In a competitive external environment, companies look for technologies to help them survive and compete with others. In Case 3, during the Covid lockdown, transportation dropped dramatically. Thus, the bargaining power of buyers strengthened, and the rivalry was already intense in the petrol industry, where there was little product differentiation (Mullins & Walker, 2012). Petrol retailers also struggled in the increasingly competitive market and needed new tools to help them compete. The struggle of petrol retailers became an opportunity for Company C; they initiated the SaaS project to help petrol retailers understand their sales and operations in more depth and width.

The two quotes from Case 1 and Case 3 above reveal the link between the high level of hostility and developing BDAC. The crisis ushered in intense competition for petrol retailers, and then the retailers looked for new technologies like big data to help them compete. Hostility influences the company's decision to use big data and develop BDAC. Thus, the following proposition:

> **Proposition 7: Hostility stimulates the development of big data analytics capabilities.**

Hostility emerged in Cases 1, 2 and 3, but not in Case 4. A possible explanation for this is that in Case 3, the product is petrol or oil, which are highly similar products that can be easily replaced by competing products available on the market. This phenomenon perfectly reflects the product differentiation aspect of hostility and only applies to Case 3; thus hostility stands out from the case. In the other cases (2, 3 and 4) the product differentiation is not as small as in Case 3. In Case 1, although the seat finder product is unique in the region where the company operates, its main business is transportation, and alternative modes of transportation are very competitive in adopting new technology, which puts the company under pressure to adopt new technology to stay in competition. In Case 2, the competition from the external environment was quite strong, and competition is the main aspect of hostility. Before the product started applying big data technology, the previous product was losing its place in the market. At that time, competitors were already using big data, and thus the company decided to harness big data to catch up. In Case 4, the phenomenon was more about the uncertainty of the actions of competitors rather than hostility. Consider the definition of both; the former emphasises uncertainty, while the latter emphasises competition. In Case 4, the possibility of using big data for cybersecurity was just a proven and feasible concept that no other company had applied yet. Thus, the market is more uncertain – competitors might use it, or they might not.

The current literature had not tested the effect of hostility on BDAC (Mikalef et al., 2019). The research contributes to the literature by showing that hostility could stimulate the development of BDAC.

8.3.4 Corporate social responsibility & BDAC

Corporate social responsibility (CSR) emerged from the research findings but only in Case 1. The company is a public transport provider, which means it has responsibility to the public society. Due to its CSR, the company prioritises reliability in its products and services. The company can be very cautious about implementing new technologies such as big data in the development of a product or service. This shows that social responsibility can sometimes act as a barrier to the development of BDAC.

This phenomenon is only observed clearly in Case 1. A possible explanation is that in Case 1, the company has the strongest social responsibilities compared to others. The company's products have a direct and significant impact on public society. For instance, if its product, the seat finder, provides misleading information to users, it can cause chaos and complaints from a large number of commuters. Yet, in other cases, like the news app and the SaaS product for petrol stations, any issues would generally have a less severe impact on public society. So, CSR can have less of an impact on the decision to adopt big data technology. As for Case 4, the cybersecurity system product is inherently digital and data-driven, so the company's CSR does not hinder the implementation of big data. Thus, CSR did not hinder the development of BDAC in Cases 2, 3, and 4.

8.3.5 Summary – external influencing factors of BDAC

To summarise, this research identified several external factors that could influence the development of BDAC. First, in a B2B scenario, customers who are open to new technologies, or are experts in the field, or have a budget for the new technology, would be willing to purchase big data-enabled products or services. This would facilitate the development of BDAC. Second, *uncertainty about the actions of competitors* could push companies to use big data, thus facilitating the development of BDAC. On the other hand, *uncertainty of the actions of customers* could hinder or even stop the development of BDAC. *Uncertainty of the actions of users could hinder the development of BDAC by hindering data variety.* Third, *hostility stimulates the* company's decision to use big data and develop BDAC. Fourth, *corporate social responsibility* could make companies hesitate about whether to start using big data, thereby delaying or hindering the development of BDAC.

8.4 Influencing factors of data variety

In this research project, the findings are also clustered around data variety. Data variety is an element of BDAC. This research has identified three factors that influence data variety, namely, *communication skills, privacy & regulations,* and *coordinating*. Communication skills belong to the category of relational knowledge, data variety is under the category of tangible resources, and coordinating is one of the dynamic capabilities. Tangible resources and relational knowledge are both constructs of BDAC. This section explains these relationships in detail. Figure 8.3 below illustrates these factors and relationships.

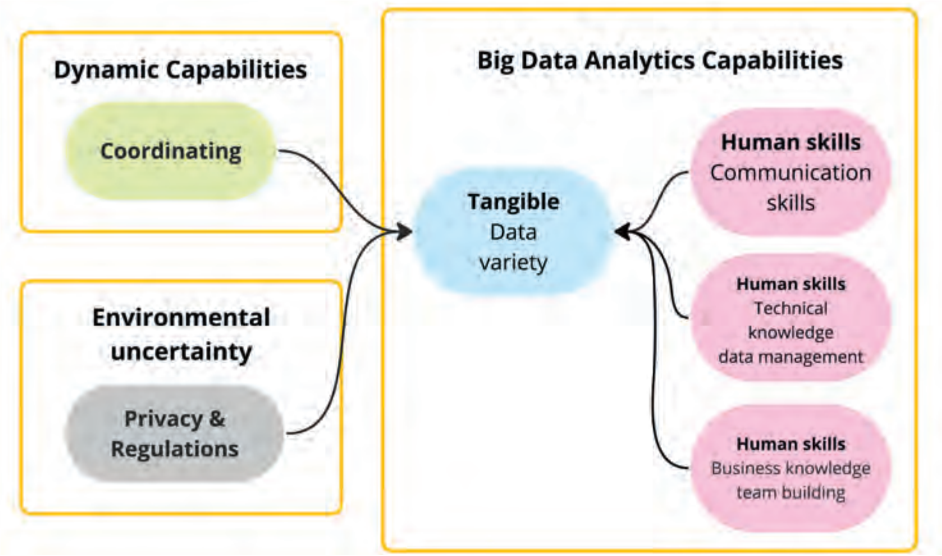


Figure 8.3 The relationships regarding the influencing factors of data variety.

8.4.1 Human skills and knowledge – communication skills & data variety

When mentioning prioritising tasks in teams, communication skills emerge as a factor. Table 8.8 below includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences.

Table 8.8 Overview of communication skills

Human skills – communication skills	
Construct	Communication skills
Definition	The ability to convey or share ideas and feelings effectively.
Proof quotes	"Lots of talks. Oh, it's, err, networking and talking, showing everybody that what we are building will deliver value to the customers and our organisation. And from that, you try to get everybody to add it to the systems. But it's not centrally managed." (product owner, Case 1)
Link to the extant literature	Communication skills (Mikalef, Boura, Lekakos & Krogstie, 2019).
Differences from the literature	The study enhances the existing literature by offering empirical evidence that illustrates the influence of communication skills on specific BDAC components.

Case 1 involved a collaborative effort between different groups to collect the data (quote in the table above). While one team collects new data sources, another team modifies the data model to reflect the updated data sources. In this scenario, different teams need to work together as one cross-functional team. The key is to maintain good communication and coordination in a timely manner to ensure a seamless fit between data collection and analysis work. Proposition 8 summarises this phenomenon.

> **Proposition 8: Communication skills facilitate collecting different data sources from different teams, thus benefiting data variety.**

Communication skills are commonly identified across the four cases. In Cases 1 and 2, communication skills are key to data collection, because data are provided by different parties, and the collection of data involves good communication between the parties. In Case 3, communication skills are important in understanding customers' needs and preferences so that the product can be adjusted accordingly. In Case 4, communication skills are associated with technical infrastructure management.

Current literature considers communication skills as an element of BDAC, but it is unclear regarding the effect of communication skills on BDAC. The current study contributes by offering empirical evidence to illustrate the influence of communication skills on specific BDAC components. The findings clarify the mechanisms through which communication skills affect these components. Specifically, communication skills are key to data collection, and they are also vital for understanding the customer's needs.

8.4.2 Environmental uncertainty – privacy and regulations & data variety

Parent (1983) defined privacy as the absence of undocumented personal knowledge about a person (Parent, 1983). This study aligns with Parent's definition and extends it to the lack of undocumented knowledge about a person or an organisation. Table 8.9 below includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences.

In a B2B innovation project (Case 4) that does not involve individual customer data, enterprise customers' data privacy emerged as a factor associated with data variety, which is a construct of big data analytics capabilities. In Case 4, for example, although respondents did not consider privacy a big issue for data usage, the client's data privacy may hinder using specific data sources to further train models for updates (see quote in the table above). For various reasons, data generated within the company are considered confidential and cannot be shared outside of the company. *"It would be great if we could combine the data from the petrol station, but they won't give us real data because this raises privacy issues. So, they won't share user information with us. They won't tell me who arrived today and what their mobile* *phone number is to compare with our data for validation. It is also a violation of the privacy law. But we can still carry out some validations using summary numbers"* (data analyst, Case 3). The B2B innovation project (Case 3) also revealed that the client's data privacy may limit the availability of data sources.

Table 8.9 Overview of privacy & regulation

Environmental uncertainty	
Construct	Privacy & regulations
Definition	Privacy: the lack of undocumented knowledge about a person or an organisation. Privacy regulations: the relations protecting privacy.
Proof quotes	*"If they [clients] don't want to give us their data for further training, that is fine. But we encourage our clients to bring us their data to update our algorithms. Especially in large enterprises, which are strict about this, some data cannot leave their intranet." (developer B, Case 4)*
Link to the extant literature	Data privacy in marketing (Martin & Murphy, 2017)
Differences from the literature	The existing literature focuses on user privacy. The study also includes business privacy. It provides empirical support to show the influence of both user and business privacy on using big data in innovation. Specifically, privacy can limit the data sources companies are using.

Practitioners who are adopting big data once had very blurry ideas about privacy. In recent years, privacy has been taken more seriously in the country of the case company. Privacy laws and regulations have caught up with the development of big data use in innovation. Companies that fail to follow these regulations are facing serious consequences. As a result of various privacy-violation cases, the line between legal and illegal became much clearer in practice. This has caused practitioners using big data to pay great attention to privacy aspects. As one of the participants mentioned: *"When I first started to make these big data products, we actually had very vague ideas about privacy and [data] security at that time. Sometimes these ideas had no basis in reality. We thought we could just produce the products and make money. But after these two or three years, we have experienced a lot during product development, including compliance and security inspections, various laws and regulations issued in our country, and several negative examples of competitors being caught. The red line is becoming clearer. We have high and clear red lines in mind when working on this product – what we should not touch, and what we should be careful with, what is compliant and legal. In fact, everyone has a standing point and thoughts, right? They also want to do a good job with their product and business, but they would not and should not violate the law" (solution architect, Case 3).*

Apart from business clients' data privacy, individual user data regulation also plays a role. In B2C projects, personal data are used as input to train models to build products. In this case, algorithm accuracy validation relies on personal data, and not being permitted to use these data would make validation less precise. Currently, the regulation in Asian countries is not as strict as in Europe. In Case 2, specific data sources are disclosed, such as the user's personal information and user-generated data. User-generated data emerged as a result of user actions (Saura, Ribeiro-Soriano & Palacios-Marqués, 2021); one example is click-through from apps. The company can use such data sources for product development without legal issues. *"For example, the data uploaded by users, including their gender and hobbies, are generated by the users when using our app. We also leverage users' behaviour data on other apps or websites because we have access to them. There is cross-app tracking. Now it's gradually, umm ... I don't know if this would be the case in Europe. People [users] in our country don't pay as much attention to privacy as those abroad"* (data scientist, Case 2).

In a B2C project, the regulation regarding privacy issues influenced the data source that can be used. Some data sources may be useful for the project but could not be used due to regulations. "*And we have some technology that has a privacy aspect – we have cameras on board, and we'd like to use a counting system based upon the camera images, but that has a very big legal impact because of GDPR, so I'm not sure if we will ever use it*" (product owner, Case 1). Thus, privacy regulation limits the data sources (variety) that can be used in digital product development projects. Proposition 9 summarises this phenomenon.

Proposition 9: Privacy & regulation restrict the development of big data analytics capabilities, specifically data variety.

Privacy is a common factor that emerged from all four cases. In Case 1, it clearly limits which data sources can be used in the innovation project, while in Cases 2, 3 and 4 it is less strictly managed, but regulations are getting stricter thus will have an influence on the data sources available for the projects in the future.

Current literature considers privacy and rule structures as intangible resources under the concept of BDAC (Mikalef, Pappas, Krogstie & Giannakos, 2018). The research contributes to the literature regarding big data analytics capabilities by suggesting privacy and regulation as a new construct under environmental uncertainty. As privacy and regulations often first come from outside of the organisations, the company then initiates privacy and rules for themselves. Privacy regulation rules can also be changing; for example, in Case 3, practitioners were aware that privacy rules might become stricter in the future.

The current literature mainly discusses user privacy (Saura, Ribeiro-Soriano & Palacios-Marqués, 2021). The results instead focus on business privacy. Based on a survey study with practitioners, privacy is the most frequently mentioned risk associated with adopting big data technologies (Raguseo, 2018). According to a content analysis study, security, privacy, and ethical issue concerns are the most important environmental factors influencing the organisational adoption of big data (Sun, Cegielski, Jia, & Hall, 2018). The research supports these findings as it shows that privacy and regulations can be considered a barrier for companies. The study further reveals that privacy regulations limit but do not exclude all data sources that can be used.

8.4.3 Dynamic capabilities & data variety

Dynamic capabilities refer to an organisation's ability to address rapidly changing environments by integrating, building, and reconfiguring internal and external competencies (Mikalef, Boura, Lekakos & Krogstie, 2019). The first-order constructs of dynamic capabilities include (1) sensing, (2) coordinating, (3) learning, (4) integrating and (5) reconfiguring routines. These constructs are adapted from past empirical studies (Mikalef and Pateli, 2017; Pavlou and El Sawy, 2011; Protogerou, Caloghirou and Lioukas, 2011). In this research, coordinating is the element of dynamic capabilities that emerged from interview data analysis.

8.4.3.1 Dynamic capabilities – coordinating & data variety

Coordinating is one of the dynamic capabilities. It is the ability to align the activities with stakeholders and manage tasks and resources. Table 8.10 below presents the definition, proof quotes about coordinating, links the findings to the extant literature and discusses the differences from the literature. The interview data suggested that coordinating capability is associated with data variety and volume.

From Case 1, it was discovered that coordinating capability is needed for efficiently gathering and processing data because there was a third-party supplier involved in data collection: "Company X is in charge of the infrastructure, which is where we got a lot of data from" (product owner, Case 1). Accurate or near-real-time data are essential to ensure optimal product quality. The company needs to collaborate with external stakeholders efficiently to fulfil the demand for real-time prediction, as exemplified in this quote: "*That ... really has to do with the speed, the time the data is collected until the moment that data reaches our data platform ... you have to improve the speed of processing in this whole chain of data collection and things for real-time availability. And it's just not there at the moment*" (programme coordinator, Case 1). Coordinating with the third-party supplier is essential in Case 1 because the supplier provided a large amount of crucial data that serve as the input for the prediction model. Efficient alignment with the supplier can improve the data collection process, thereby facilitating data variety.

Table 8.10 Overview of dynamic capabilities

Dynamic capabilities	
Definition: An organisation's abilities to address rapidly changing environments by integrating, building, and reconfiguring internal and external competencies (Mikalef, Boura, Lekakos & Krogstie, 2019).	
Construct	**Coordinating**
Definition	The abilities to integrate, build and reconfigure internal and external competencies to address rapidly changing environments (Mikalef, Boura, Lekakos & Krogstie, 2019).
Proof quotes	"Did he really go to refill his tank with petrol? Then how long does it take to refuel? We have to compare the data – we need customer data feedback in order for us to understand the situation and then decide whether my model is trained correctly or not. We continuously perform scene classification and machine learning verification." (solution architect, Case 3)
Link to the extant literature	Managing and partnering with external stakeholders in management (Harrison & John, 1996)
Differences from the literature	In literature regarding managing and partnering with external stakeholders, the development of managing external stakeholders is considered to blur the line between internal and external stakeholders. The findings echo the literature and also propose that customers and suppliers can serve as data source providers of the kind that have traditionally come from inside companies.

In Case 2, the data source is the content created by individual new media and traditional media institutions, which also requires coordinating. In Case 3, coordinating with external stakeholders was also needed for data collection, and the data were used for model validation (see quote in Table 4). The model predicts the number of customers visiting petrol stations at a particular time. The number of customers visiting the petrol station is needed to validate the model; however, the clients own this data. Thus, coordinating with the clients is vital for the data collection to validate the prediction model.

In both cases, external stakeholders can provide a large number of data sources to be used as the input for the prediction model or to validate the model. Thus, it is proposed:

> **Proposition 10: Coordinating is positively associated with big data analytics capability, specifically data variety.**

Coordinating as a factor is prominent in Cases 1, 2 and 3 because data sources are owned by external stakeholders; thus, coordinating is key for data collection. When it comes to Case 4, the company also needed data sources owned by clients. The possible reason why coordinating was not prominent is that clients in the cybersecurity industry are more willing to share data and feel safe doing so, and large-scale data sharing is part of the deal when purchasing cybersecurity services. For this reason, coordinating did not stand out as a factor in Case 4, but that does not mean coordinating with clients is not necessary for data source collection.

8.4.4 Summary – influencing factors of data variety

This research identified several factors that could influence data variety. *Communication skills* and *coordinating capability* could facilitate collecting different data sources from different stakeholders, thus benefiting data variety. *Privacy & regulations* could make certain data sources inaccessible for use, thus hindering data variety.

8.5 Influencing factors of innovation performance

In this research, two aspects of innovation performance were identified: product performance and project performance. Product performance is influenced by three factors: data variety, data reliability, and business analytics. Project performance, on the other hand, is influenced by one factor, which is data velocity (as seen in Case 1). Figure 8.4 below highlights these factors and relationships. Factors and relationships that only emerged from one case are marked with * in the figure. Innovation performance change is the consequence of using big data. The following paragraphs first discuss the measurements of innovation performance identified from this research; this is followed by a discussion about the four factors and how they influence innovation performance.

8.5.1 Innovation performance (the consequence of developing BDAC)

This research identified two consequences of big data usage: product performance and project performance. Most of the findings concern product performance, and the main measurement that stood out from the findings is prediction accuracy. Table 8.11 below presents an overview of product performance and prediction accuracy, including their definitions and proof quotes from this research, then links them with extant literature and discusses the differences. There are many ways to measure and evaluate product performance. This research revealed several of them that are used in practice: the ones that focus more on the characteristics and function of the product – accuracy and usefulness (this research has determined that in terms of prediction accuracy, closer to real-time prediction is more accurate); the ones that focus more on the customers – customer feedback; and the ones that focus on the market – sales volume, market penetration, market development (Ansoff, 1957). The quotes below illustrate these measurements. For project performance, time to market is the measurement identified from this research and corporate social responsibility (CSR) could have an impact on it; this is only shown in one case and is thus not included in the overview table (see 8.2 CSR).

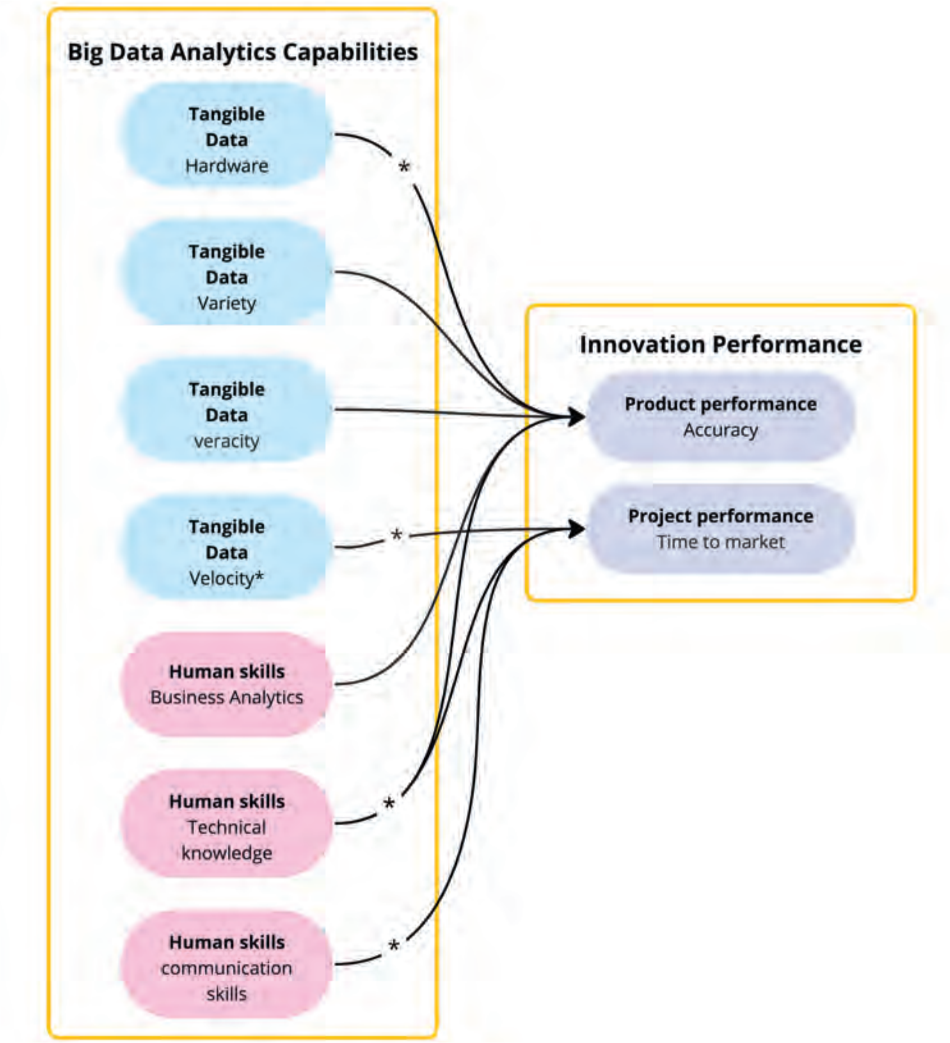


Figure 8.4 The relationships regarding the influencing factors of innovation performance. *Presents factors and relationships identified from only one case.

8

Table 8.11 Overview of the consequences of big data usage

	Consequences of big data usage	
Construct	**Product performance**	**Prediction accuracy**
Definition	The product's response to external actions in its working environment. It is realised through the performance of its constituent components (Zeng & Gu, 1999).	The measurement of correctly predicted data points out of all data points.
Proof quotes	Accuracy – *"There are these indicators: accuracy rate, false-positive rate. These are the two core indicators."* (developer A, Case 4) Customer feedback – *"We are also working on a complete digital feedback loop where we offer travellers possibilities to give their online feedback."* (programme coordinator, Case 1) Usefulness, sales volume, market penetration, market development, innovativeness – *"There are three levels. First, your product has been established now. Is it useful? Does it sell? Second, how widely is your product used and sold? Is it just a single application or can it be duplicated to extend across the whole chain in the industry? Third, the degree of innovation of your product and the value it brings to customers."* (solution architect, Case 3).	*"I'd like to improve the service by using more short-time data and make it provide more real-time predictions. So that each morning, you can run the model – it's a more accurate prediction when you make it on shorter notice."* (programme coordinator, Case 1)
Link to the extant literature	Evaluate Performance Measures (Keegan, Eiler & Jones, 1989); (van Veen-Dirks, 2010).	Big data accuracy assessment tool (Mylavarapu, Thomas & Viswanathan, 2019).
Differences from the literature	The study revealed various factors associated with digital product performance from practice.	The study provides empirical evidence supporting what accuracy means in describing or predicting reality.

8

8.5.1.1 Tangible resources – data variety & product performance

Data comprise a type of tangible resource that is indispensable for developing big data analytics capabilities. Data set the foundation for the project by offering crucial insights. Data variety refers to different data formats and sources. Different data sources can offer different insights that complement each other. In Case 3, the company provides a SaaS product that can predict the number of cars arriving at a petrol station to fill their tanks based on traffic data. However, to give price suggestions, more data are needed from the petrol stations. For example, the number of customers who went to the station to fill up their tanks is based on sales records. These records could help the SaaS product validate the prediction of the number of customers and sales based on traffic, while traffic prediction provides closer to real-time information than sales records. The prediction is improved, leading to better product performance. Data variety has a positive influence on product performance and this phenomenon is observed in all four cases.

Proposition 11: Data variety brings various insights that could positively influence product performance.

8.5.1.2 Tangible resources – data reliability & product performance

Data reliability refers to how well the data reflects reality. Table 8.12 below includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences. For example, in Case 3 (quote in the table below), the reliability of data could influence the quality of the prediction model. The model is the main part of the product, thus leading to the following proposition:

Proposition 12: Data reliability is positively associated with product performance.

Table 8.12 Overview of data variety and reliability

	Tangible resources	
Construct	**Data variety**	**Data reliability**
Definition	The different data formats and sources (Kaisler, Armour, Espinosa, and Money, 2013).	How reliable are the data in terms of reflecting the reality?
Proof quotes	*"I only have this part of the data. I don't know what the price is at each petrol station. How many cars arrived for petrol filling? […] All of these, I don't know […] We provide a piece of data that our customer has not used before. We added the data, but the problem is that I don't have the data that our clients have. In fact, our data complement theirs."* (domain expert, Case 3)	*"Can the 'white samples' and 'black samples' that we are taking now represent the current trend? Some of the previous attack methods keep on changing."* (developer A, Case 3)
Link to the extant literature	Big data sources and methods for social and economic analyses (Blazquez & Domenech, 2018); Big data analytics in marketing strategy (Ducange, Pecori & Mezzina, 2017). Big data variety significantly impacts innovation performance, while volume does not (Ghasemaghaei & Calic, 2020). Big data variety has significant impact on data-driven insights (Ghasemaghaei & Calic, 2019).	Data quality model (Merino, Caballero, Rivas, Serrano & Piattini, 2016). Concept of reliability from quality control, specifically, absolute reliability – the level of resemblance of data items to reality (Agmon & Ahituv, 1987).
Differences from the literature	The study provides empirical examples illustrating how exactly different data sources can provide different insights that complement each other.	The study provides empirical evidence of absolute reliability in data and emphasises the importance of absolute reliability.

8

8.5.1.3 Tangible resources – data velocity & project performance

Data velocity shows an influence on project performance, specifically influencing time to market; this can only be clearly observed in Case 1. The quote from Chapter 4 regarding Case 1 illustrates this (see Chapter 4 – data velocity & project performance). Data collection speed is part of data velocity (Gandomi & Haider, 2015; George et al., 2016). Data collection is an early step in using big data for innovation and the subsequent steps include analysis for digital product prototyping. The speed of data collection could then influence how quickly the product would go through the innovation process and finally reach the customers. Thus, velocity influences time to market.

Velocity did not stand out in the other three cases; a possible reason for this is that in Cases 2, 3, and 4, the speed at which data are generated, collected, stored, and analysed was not delayed or did not have a noticeable influence on time to market. Thus, participants did not mention it. However, that does not mean that velocity does not matter in terms of shortening time to market.

8.5.1.4 Tangible resources – data volume

This research identified two out of the three commonly known characteristics of big data, which are variety and velocity, but not volume. Volume, according to the definition concerns the magnitude of data. Data come from a variety of data sources. More variety means more volume as well. The research did not consider volume as a factor because the influence is on the variety side; more sources of data matter. However, obtaining a greater volume of data from the same source does not seem to have an influence. Thus, in this research, variety and velocity emerged rather than volume.

8

8.5.1.5 Human skills and knowledge – business analytics & product performance

There are many different definitions of business analytics proposed by different authors. Examples include the Management Study Guide (2018), Pratt (2017), and Gartner IT glossary (2018). This research adopts a comprehensive definition proposed by a business analytics definition review paper. Business analytics is the systematic thinking process of using qualitative and quantitative computational tools and methods to analyse data and gain insights to inform and support decision-making (Power, Heavin, McDermott & Daly, 2018). Table 8.13 includes its definition and proof quotes from this research, then links them with extant literature and discusses the differences.

Table 8.13 Overview of human skills – business analytics

Human skills – business analytics	
Construct	**Business analytics**
Definition	Business analytics is the systematic thinking process of using qualitative and quantitative computational tools and methods to analyse data and gain insights to inform and support decision-making (Power, Heavin, McDermott & Daly, 2018).
Proof quotes	Real-time analytics – *"And then, there is the recommendation algorithm. In fact, we can deal with the data closer to real-time. Support this more real-time recommendation algorithm ... For example, you processed the data once an hour, and now you can turn it into a minute ... The effect is a substantial improvement. For example, in the past, the algorithm generated a personalised user page that might be calculated one day in advance. Now it might be that every time you swipe it or open the app, you get a new one. The user experience will definitely be better than before."* (data scientist, Case 2).
Link to the extant literature	Business analytics is an elements of human skills (Mikalef et al., 2019).
Differences from the literature	The research contributes to the literature by illustrating that different resources often function together to enable business analytics.

One example of business analytics that emerged from the research is real-time analytics. Real-time analytics applies logic and mathematics to data instantaneously to provide insights for decision-making (Gartner, n.d.). As an aspect of business analytics, real-time analytics makes it possible to improve product performance. In Case 2, for example, closer to real-time analytics ensure that the product is more effectively updated with current news. Considering the changing content preferences of users, real-time analytics is ideal for users who are always interested in breaking news and content to stay informed. Thus, real-time analytics improves product performance. The quote in the table above illustrates this.

Several human skills and tangible resources underlie real-time analytics. Table 8.14 below lists a few types of resources based on the categorisation as examples. Together, these tangible resources and human skills are associated with product performance.

Table 8.14 Example skills and resources

Resource type	Characteristics	Definitions
Tangible resources (infrastructure)	Edge computing	A distributed information technology (IT) architecture (Brush, 2022).
Tangible resources	Data warehouse appliances	A combination of hardware and software products designed specifically for analytical processing (Brush, 2022).
Human skills – technical knowledge	Programming	A technical process in which programmers create instructions for computers to follow.

Proposition 13: Business analytics, specifically real-time analysis, could improve product performance.

Business analytics is identified across all four cases. Cases 1 and 2 demonstrate that business analytics is associated with product performance. Case 3 shows that coordinating facilitates data collection, which then improves the quality of business analytics. Case 3 also illustrated that business analytics together with data enabled the development of BDAC. In Case 4, technical knowledge enables better business analytics and further influences product performance. The research contributes to the literature by illustrating that often different resources function together to enable business analytics.

8.5.2 Summary – influencing factors of innovation performance

This research revealed that *data variety* could bring various insights and triangulations that could improve *product accuracy*; thus, it facilitates *product performance*. *Data reliability* could benefit prediction accuracy; thus, it also facilitates product performance. *Business analytics*, specifically *real-time analysis*, keeps products up-to-date, thereby benefiting product performance. *Data velocity*, specifically *the speed of data collection*, could then influence how quickly the product goes through the innovation process and finally reaches the customers. Thus, data velocity influences project performance – time to market.

8.6 Summary – cross-case analysis

This chapter presents the findings from cross-case analysis. The findings are based on four product development projects from four companies in four industries.

To give an overview, the results show that factors from both inside and outside of the company influence the development of BDAC, which then influences innovation performance.

Figure 8.5 below presents an overview of all the factors identified from this research. External factors such as dynamic capabilities and environmental uncertainty could facilitate or hinder the development of BDAC. The development of BDAC can then have a positive influence on innovation performance. The following Chapter 9 discusses these main findings, its theoretical and practical implications, along with limitations and future perspectives.

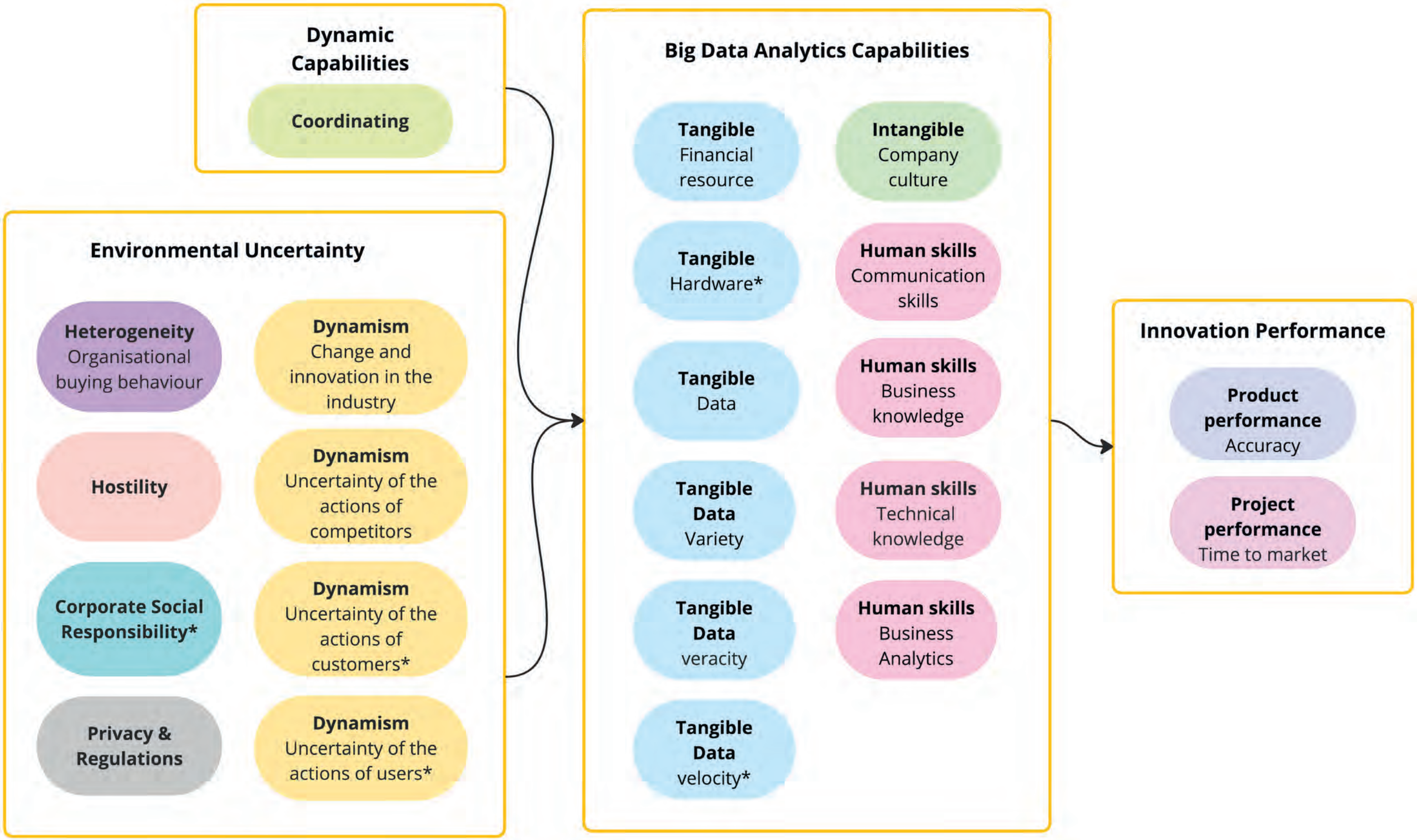


Figure 8.5 Overview of factors and main relationship found in this research.
*Presents factors and relationships identified from only one case.
Different colours are used to distinguish between resource and capabilities on the second layer constructs of BDAC, dynamic capabilities, environmental uncertainty, and innovation performance.

CHAPTER 9

Discussion

This chapter presents the discussion of the dissertation. It first briefly recaps the aim and the main findings of this research project. Then, it highlights the most intriguing findings. Furthermore, it discusses how the result of this research contributes to literature and theories in the field of innovation management. The chapter concludes with the research contributions to the practical world, the limitations of this research, and future research directions.

9.1 Main findings

This research project aims to explore the antecedents and consequences of developing big data analytics capabilities (BDAC) in digital product innovation, how these antecedents influence BDAC, and how the constructs of BDAC influence the outcomes. It adopted the theoretical lens of the resource-based view (RBV) and dynamic capabilities (DC) for the research investigation. The research is guided by the following main research question, followed by four sub-questions:

RQ: How do companies develop big data analytics capabilities in digital product innovation?

> Sub-RQ1: What are the antecedents of developing big data analytics capabilities in digital product innovation at the project level?
>
> Sub-RQ2: How do the identified antecedents influence the development of big data analytics capabilities in digital product innovation at the project level?
>
> Sub-RQ3: What are the consequences of developing big data analytics capabilities in digital product innovation at the project level?
>
> Sub-RQ4: How do big data analytics capabilities influence its consequences in digital product innovation at the project level?

9

Figure 9.1 Conceptual model illustrating the research questions of this study.

Figure 9.1 above visualises the research questions of this study. This study conducted a cross-case analysis to answer these questions, and used interviews as the data collection method. The main findings are provided after iterative analysis.

Figure 9.2 below illustrates the main findings in response to the research questions in the form of a theoretical framework. BDAC is at the centre of the theoretical framework, the antecedents are on the left of BDAC, and the consequences are on the right. There are both internal and external antecedents influencing the development of BDAC. Within the context of BDAC, its constructs play a significant role in contributing to its overall development. From the external side, coordinating, an element of dynamic capabilities, and several environmental uncertainty elements impact BDAC. Regarding the consequences, BDAC influences product performance and project performance.

The propositions below explain the relationships between the antecedents and consequences in the theoretical framework in detail. The propositions are grouped by the three dependent variables: BDAC, data variety, and product performance.

9.1.1 Internal influencing factors of BDAC:

Proposition 1: Basic resources such as financial resources set the foundations and could facilitate the development of big data analytics capabilities.

Proposition 2: Company culture facilitates the development of big data analytics capabilities.

Proposition 3: Both technical and business knowledge are needed in initiating big data analytics capabilities development.

9.1.2 External influencing factors of BDAC:

Proposition 4: Heterogeneity (organisational buying behaviour) is positively associated with developing big data analytics capabilities, because customers who are open to new technologies, are experts in the field, or have a budget for the new technology are willing to purchase big data-enabled products or services.

Proposition 5: Dynamism (change and innovation in the industry) and technical knowledge are facilitating the development of big data analytics capabilities.

Proposition 6: Dynamism (uncertainty of the actions of competitors) stimulates the development of big data analytics capabilities.

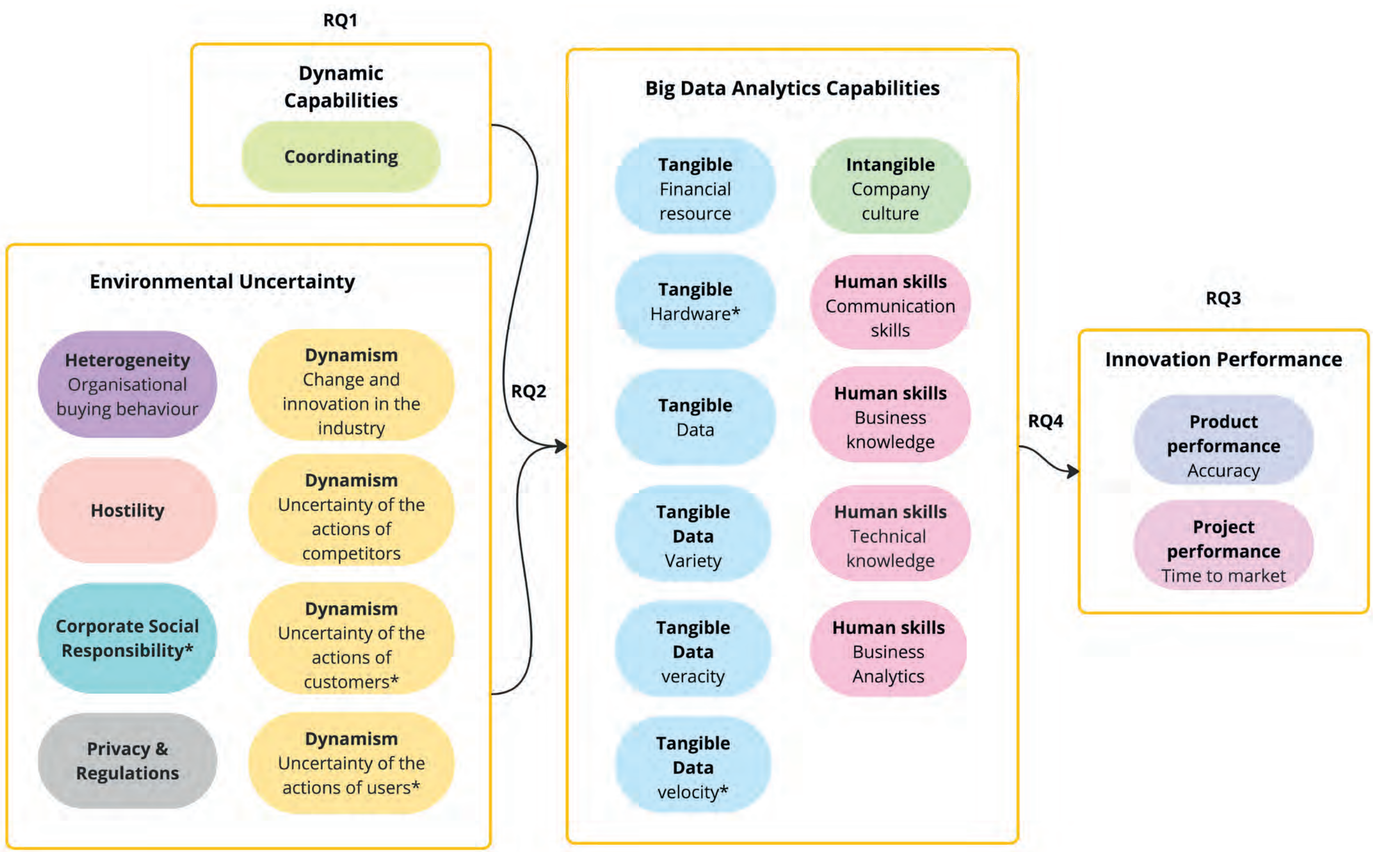


Figure 9.2 Result theoretical framework.
*Presents factors and relationships identified from only one case.
Different colours are used to distinguish between resources and capabilities on the second layer constructs of BDAC, dynamic capabilities, environmental uncertainty, and innovation performance.

Proposition 7: Hostility stimulates the development of big data analytics capabilities.

9.1.3 Influencing factors of data variety:
Proposition 8: Communication skills facilitate collecting different data sources from different teams, thus benefiting data variety.

Proposition 9: Privacy & regulation restricts the development of big data analytics capabilities, specifically data variety.

Proposition 10: Coordinating is positively associated with big data analytics capabilities, specifically data variety.

9.1.4 Influencing factors of product performance:
Proposition 11: Data variety brings various insights that could positively influence product performance.

Proposition 12: Data reliability is positively associated with product performance.

Proposition 13: Business analytics, specifically real-time analysis, could improve product performance.

9.2 Theoretical implications

This section explains the different aspects of the theoretical implications of this study. Starting from the middle of Figure 9.2 above, this study explained how the constructs of BDAC facilitate its development, identified the antecedents and consequences of data variety, and revealed the interactions between BDAC elements. Then, moving on to the main categories presented in Figure 9.2, this study proposed new relationships between dynamic capabilities and BDAC, and new relationships between environmental uncertainty and BDAC, and unpacked the impact of BDAC constructs on innovation performance. Finally, this study made a methodological contribution to the field of developing BDAC in product innovation. The following paragraphs are organised and sub-headed by the aspects listed above.

9.2.1 Explained how the constructs of BDAC facilitate its development
BDAC is a capability that is made up of many constructs. Previous studies classified these BDAC constructs or elements into the categories of tangible resources, intangible resources, and human skills (Mikalef et al., 2019). Although these studies classified the constructs of BDAC, it remained unknown how these constructs form

or contribute to the development of BDAC (Mikalef et al., 2019). This study adds to this stream of literature by explaining how these constructs contribute to the development of BDAC.

The findings of this study reveal that several constructs contribute directly to the development of BDAC. These constructs include financial resources, hardware, data, company culture, business, and technical knowledge. Based on the classification (Mikalef et al., 2019), they belong to three different categories: tangible resources, intangible resources, and human skills & knowledge. Figure 9.3 presents the constructs with the categories they belong to. The arrows illustrate the relationships between each construct and BDAC, along with an arrow illustrating the relationship between two of the constructs (business knowledge and technical knowledge). The following paragraphs discuss these relationships in detail, starting from tangible resources and proceeding to intangible resources and, finally, human skills & knowledge.

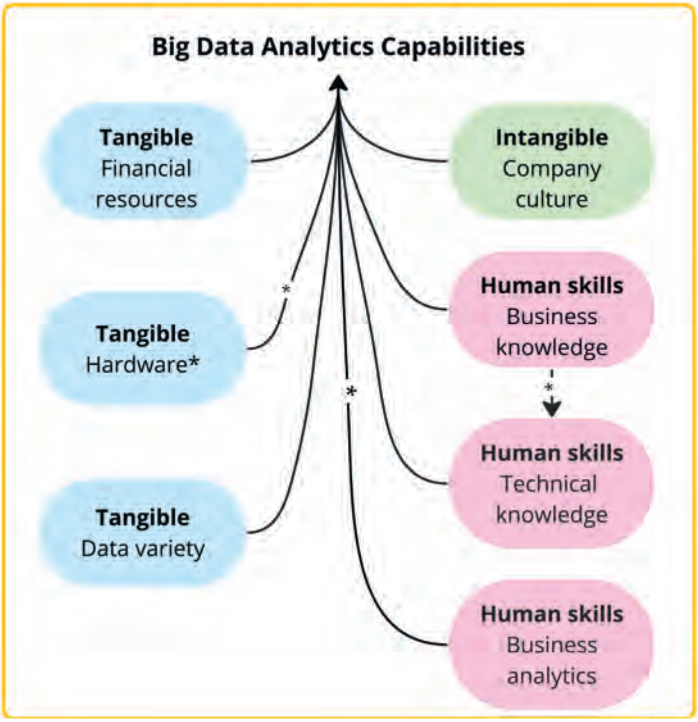


Figure 9.3 The relationships regarding the internal influencing factors of BDAC.
*Presents factors and relationships identified from only one case.
Different colours are used to distinguish between resources and capabilities on the second layer constructs of BDAC.

Financial resources are a type of tangible resource. The study shows that financial resources are the foundation for developing BDAC. Without these resources, a big data-enabled product development project cannot move forward, hindering BDAC development. In Case 1, the project could only start its operation once the management team approved the funding. In Case 2, the project had to stop due to a lack of financial resources.

Another tangible resource identified from this study is hardware. Hardware is also essential, and its advancement could benefit the development of BDAC. For example, in Case 2, the evolution of semiconductor chips means faster data processing and analysis, thus facilitating BDAC.

As a tangible resource, data is another foundation for developing BDAC. The factor only stood out in some cases because, as an essential resource, it is easily neglected.

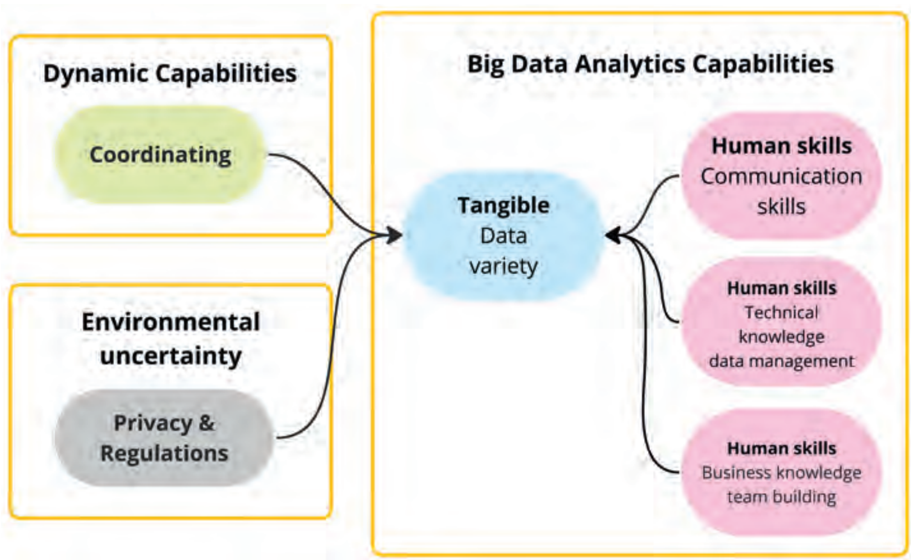
Based on the findings of this study, it was observed that, once tangible resources have been allocated, intangible resources also contribute to the development of BDAC. Company culture is considered an intangible resource. Data-driven culture increases the perceived value of data. Top-management support gets the project initiated. An open culture makes data sharing and usage free from data silos. Companies with these cultures are building supportive environments for BDAC development.

Human skills & knowledge also contribute to the development of BDAC. Technical knowledge and business knowledge belong to the human skills & knowledge category. This study found that initiating a big data-enabled product innovation project typically requires a combination of relevant business and technical knowledge within the company to kickstart the project.

Tangible resources, intangible resources, and human skills are the elements that together contributed to the development of BDAC. When a company is developing its BDAC, in essence, it is realised through the development of these elements of BDAC. Together, the development of these elements pushes the development of BDAC. This research adds to the current literature by revealing how these elements push BDAC development.

9.2.2 The antecedents and consequences of data variety

Regarding the antecedents of data variety, previous studies have focused on companies' orientation towards innovation and revealed the impact of exploration and exploitation orientations on data variety (Johnson et al., 2017). This study took a different angle and identified additional antecedents of data variety from dynamic capabilities, from environmental uncertainty and from within BDAC. The antecedents are coordinating, privacy & regulations, and communication skills. Figure 9.4 below illustrates the antecedents of data variety. It can be seen that the external antecedents coordinating and privacy & regulations influence BDAC by influencing data variety, and communication skills also influence it internally. The possible reason why there are more antecedents influencing data variety than other factors is that data variety could yield rich and diverse insights that further enhance the product. A change in data variety makes a difference, while a change in data volume does not because greater amounts of data from the same sources cannot yield more insights. In addition, data variety has more antecedents, which might be because the influence of antecedents on data variety is more noticeable than the influence of antecedents on other elements of BDAC. For example, the influence of environmental factors on company culture tends to be subtler.



9

Figure 9.4 The relationships regarding the influencing factors of data variety. Different colours are used to distinguish between resources and capabilities on the second layer constructs of BDAC, dynamic capabilities, and environmental uncertainty.

Regarding consequences, previous studies have identified that data variety has a significant impact on data-driven insights (Ghasemaghaei & Calic, 2019) and innovation efficacy & efficiency (Ghasemaghaei & Calic, 2020). Efficiency focuses on the time and money spent on projects, while efficacy focuses on the extension of the product range and market. However, it was not clear how data variety exerted this influence and whether data variety impacts other aspects of innovation outcomes. This study fills that gap by indicating that data variety could also influence product performance – product prediction accuracy and explained how data variety impacts data-driven insights and product performance. This study reveals that when data-driven insights corroborate each other, this has a positive impact on innovation performance, specifically, product prediction accuracy. For example, in Case 3, the SaaS product predicts petrol station traffic using navigation data. To refine the predictions, the company leverages sales record data on customer numbers, improving the SaaS product's performance by validating navigation data-based predictions against sales data. Combining these two data sources enhances traffic prediction accuracy. Thus, this finding contributes to the literature by adding explanations as to how data variety yields data-driven insights and, by validating through different data sources, how these insights further impact specific product performance – product prediction accuracy.

9.2.3 Revealing interactions: BDAC element dynamics

Considering there are many elements or constructs of BDAC, there are also relationships within BDAC; the constructs could facilitate each other. This research revealed that communication skills facilitate collecting different data sources from different teams, thus benefiting data variety (Proposition 8). This is illustrated in Figure 9.4. On the other hand, innovation process management could negatively affect data management if the data management strategy is considered from a short-term perspective. These findings highlight that while there are antecedents influencing BDAC from the outside, there are also dynamics within BDAC.

Previous studies have focused on testing the relationship between the BDAC concept as a whole and its antecedents and consequences (see Chapter 2, Table 5). For example, one study tested alliance management capability as an antecedent of BDAC (Dubey et al., 2021) and two investigated innovation capabilities as consequences (Mikalef et al., 2019; Mikalef et al., 2020). While previous studies examined BDAC as a holistic concept, this research contributes to the literature by providing empirical evidence to reveal the relationship within BDAC, specifically between BDAC constructs. Therefore, this research highlights the presence of the internal dynamic nature of BDAC.

9.2.4 New relationships between dynamic capabilities and BDAC

Moving on to the external antecedents of BDAC, Figure 9.4 illustrates the research findings regarding the relationship between dynamic capabilities and BDAC. This study revealed different relationships between dynamic capabilities and BDAC compared with other current empirical studies. Current empirical studies have reported that dynamic capabilities mediate the effect of BDAC on innovation capabilities (Mikalef et al., 2019). This research result suggests that coordinating can serve a different role than a mediator. As the element of dynamic capabilities, coordinating facilitates BDAC element data variety. By coordinating with third parties, the company can gain access to new data sources, and these data sources have the potential to improve the product. For example, in Case 3, the model can predict the number of visitors to the petrol station based on navigation app data. However, sales transaction data from the petrol station is needed to validate predictions and increase accuracy. Coordinating effectively with petrol stations is the key to getting access to the data.

9.2.5 Revealing new relationships between environmental uncertainty and BDAC

This research also identified relationships between environmental uncertainty and BDAC that differ from those found in other current studies. For example, how does hostility influence BDAC (Proposition 13)? Figure 9.5 below illustrates the relationships between elements of environmental uncertainty and BDAC. Existing literature has tested the moderating role of hostility on the effect of BDAC on innovation capabilities without significant results (Mikalef et al., 2019). This research contributes to the literature by adding a new finding about hostility. While hostility might not moderate the effect of BDAC on innovation capabilities, according to Mikalef et al. (2019), it could still influence BDAC. As observed across cases, companies start to use new technologies when there is high competition in the external environment. For example, petrol stations face relatively high competition because they all sell a largely similar product and, furthermore, demand dropped dramatically during Covid due to the limited need for travel or commuting. Petrol suppliers felt the pressure, and thus initiated a project to use big data to help them understand the market, gaining information such as a heatmap illustrating relative business in specific areas and how many potential customers visited petrol station A compared with other petrol stations.

Regarding the current literature about the role of environmental uncertainties, previous studies suggest that BDAC can influence innovation capability through the mediator of dynamic capabilities, and environmental uncertainties moderate the effects (Mikalef et al., 2019). Figure 9.6 presents the theoretical framework from Mikalef's study. This research result suggests new relationships whereby the constructs of environmental uncertainty can impact a firm's BDAC. The constructs include dynamism, heterogeneity, hostility, and corporate social responsibility. This means that while the environmental uncertainties could serve as a moderator, they can also directly affect BDAC.

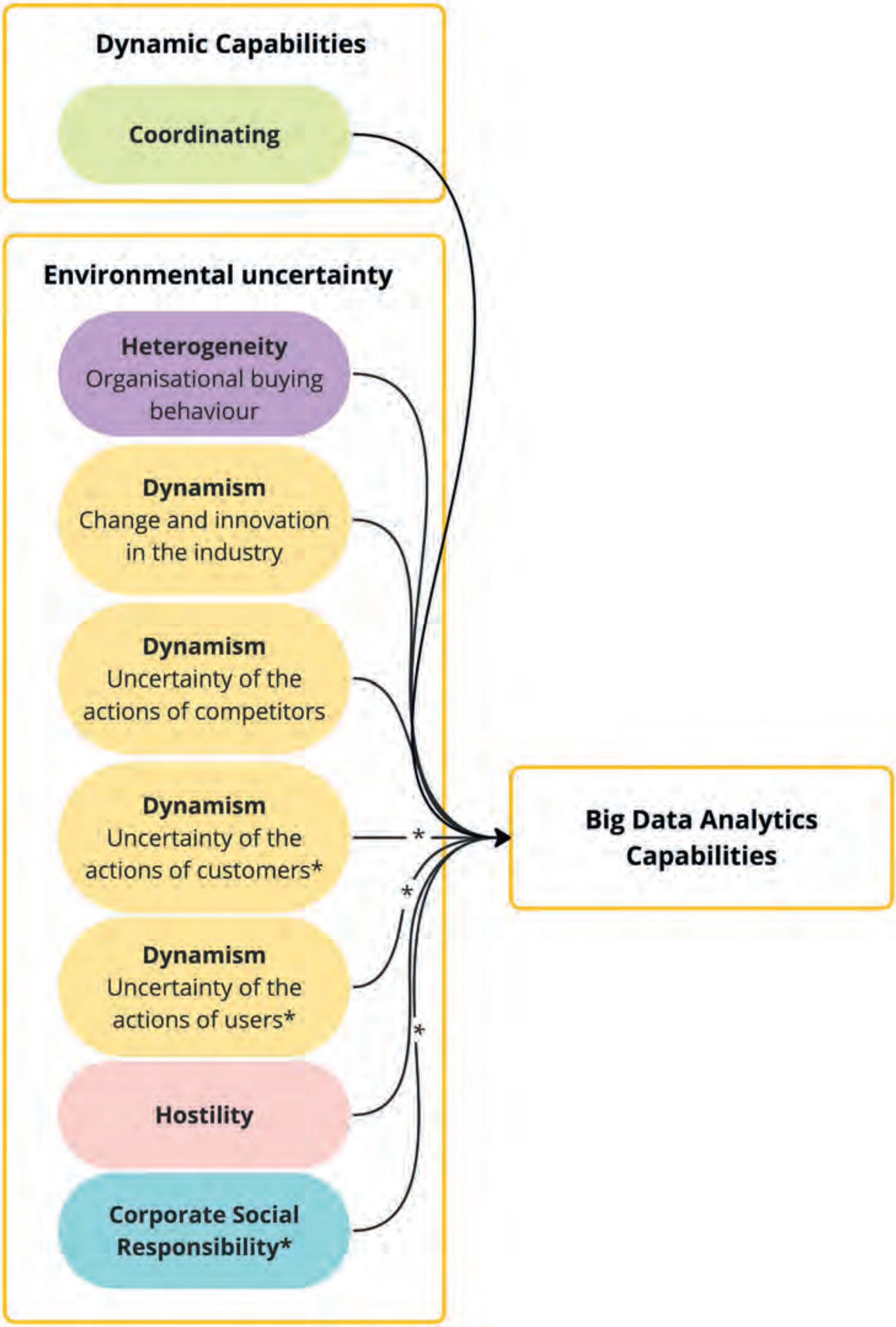


Figure 9.5 The relationships regarding the external influencing factors of BDAC.
*Presents factors and relationships identified from only one case.
Different colours are used to distinguish between resources and capabilities on the second layer constructs of environmental uncertainty.

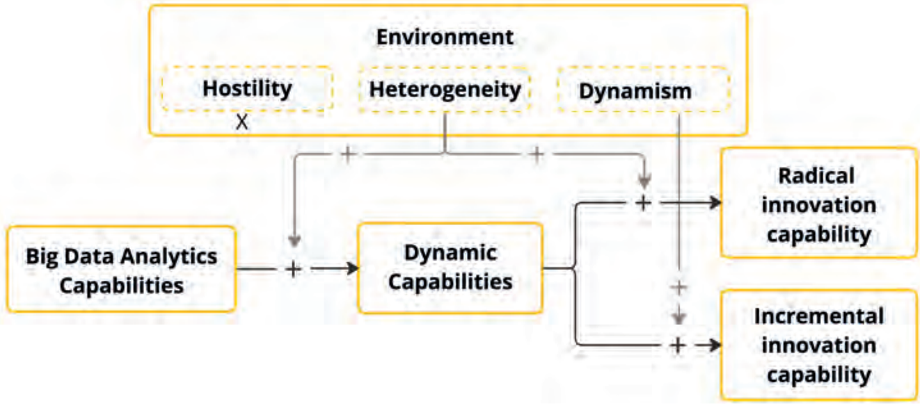


Figure 9.6 Theoretical framework from Mikalef, Boura, Lekakos & Krogstie (2019)

9.2.6 Unpacking the Impact of BDAC constructs on innovation performance

The previous paragraphs discussed the implications regarding the antecedents of BDAC from within and outside of BDAC. Now, this study moves on to the impact of BDAC on innovation performance. As discussed above in the section on how elements of BDAC interact with each other, these many elements of BDAC together push its development. Thus, the impact of BDAC on innovation performance could mean the impact of elements of BDAC on innovation performance. Innovation performance is embodied in product performance (accuracy) and project performance (time to market). This research identified BDAC elements that could impact both.

Starting from the antecedents of product performance, diversity in data sources could bring various new insights that may benefit product performance. Data reliability is also vital for product performance. For example, in order to predict the number of visitors to a petrol station, navigation data, including navigation route and road traffic information, are input for the algorithm to enable the functionality of the product. The reliability of these data sources would impact the accuracy of predictions of the number of visitors to the petrol station (Case 3). Business analytics are associated with improving product performance. Take real-time analysis, for example. For news apps, real-time browsing history analysis means staying closer to users' changing preferences and enabling the better prediction of content that users might be interested in (Case 2). Thus, real-time analysis improves product performance.

Regarding project performance, velocity has a positive impact on time to market. Velocity measures how fast the data are generated, collected, stored, and analysed (Gandomi & Haider, 2015; George et al., 2016). A participant in Case 1 pointed out the importance of increasing data collection and data storage by building an efficient data platform to improve time-to-market.

In the literature, previous studies have tested the relationship between concepts regarding big data and its innovation consequences (see Chapter 2, Tables 5 & 6). For example, studies have focused on the impact of the 3Vs (volume, variety, and velocity) on new product revenue (Johnson et al., 2017) or on innovation efficiency and efficacy (Ghasemaghaei & Calic, 2020). Another example is from Mikalef et al. (2020); their study investigated the relationship between BDAC as a holistic concept and its impact on radical and incremental innovation capability. To contribute to the field, this research brings the next level of detail by exploring the influence of BDAC constructs on innovation performance measurements. By focusing on the constructs of BDAC rather than the concept of BDAC as a whole, this research reveals the mechanism behind BDAC's impact on innovation performance. It revealed that increased data velocity could shorten the time to market. It also uncovered how data variety, reliability, and business analytics could influence product performance (propositions 2, 3, 6). In addition, product reliability is a measurement that has not been discussed in previous empirical studies on this topic.

9.2.7 Methodology contribution

In addition to its findings, this study also contributes to the literature from a methodology perspective. It is one of the first studies to use interviews as a data collection method to build multiple case studies to explore the phenomenon of developing BDAC in digital product innovation projects. Previous empirical studies have adopted quantitative large-scale surveys to examine the antecedents and consequences of using big data for innovation (e.g. Mikalef et al., 2020; Dubey et al., 2021). Quantitative studies are beneficial in testing relationships; however, they cannot explain how the relationship or impact happens, which is the strength of case studies (Creswell & Creswell, 2018; Yin, 2013). For this reason, this study adopts the case study method to answer the questions about how the antecedents influence BDAC and how BDAC then influences the consequences, thereby seeking to explain and understand the phenomenon.

9.2.8 Theoretical contribution in general

In general, the findings from this study shed light on how antecedents influence BDAC and how it then influences innovation performance to make the field more comprehensive. When using RBV as a theoretical lens to understand the mechanism behind improving innovation performance, it is necessary to look at the antecedents of BDAC constructs, explore the relationship between BDAC constructs, and their impact on innovation performance. This study adds to the current literature by explaining how the constructs of BDAC facilitate its development, uncovering the dynamics within BDAC and revealing the importance of data variety. It also contributes to the current literature by uncovering new relationships between dynamic capabilities and BDAC, between environmental uncertainty and BDAC, expanding the category of environmental uncertainty and unpacking the impact of BDAC constructs on innovation performance. In addition, it is one of the first studies that conducts interview data analysis to build multiple case studies for exploring the phenomenon.

9.3 Practical implications

This research output is a framework that demonstrates the mechanisms of enabling innovation through big data analytics capabilities. The research findings illustrate how BDAC constructs, dynamic capabilities, and environmental uncertainty influence BDAC and its constructs, ultimately influencing innovation performance. This study draws on experiences and lessons learnt across four cases from different industries and geolocations, providing broader knowledge than a study focusing on a single company. Through this framework, practitioners such as managers, designers, and data scientists can see how their work can contribute to and influence innovation performance and how to improve it when managing digital product development projects. The following practical implications are organised by themes, starting from how to increase data variety to how to build internal and external environments to facilitate BDAC and, finally, what practitioners should focus on to improve innovation performance.

Starting from data variety, because different data sources could bring valuable insights to big data-enabled innovation projects, it is important to build an environment that fosters data variety. The findings from this study could help practitioners such as product owners, programme coordinators and data scientists alike to realise that coordinating capability and communication skills could facilitate data collection from different stakeholders. New data sources added to big data innovation projects yield new insights that could complement existing ones, which could improve product performance in the end. On the other hand, considering this significant role of data variety in the theoretical model, companies are attracted by the advantages that it can bring, motivating them to explore all available options for acquiring more and better data sources. However, this pursuit must be balanced with privacy concerns and the implementation of privacy regulations and policies to safeguard the rights of relevant stakeholders.

Coordinating and communication skills are important for accessing data sources, which benefits BDAC development. Other internal factors besides data variety are also important, including financial resources, hardware, data, company culture, business, and technical knowledge. Sufficient funding, effective hardware, and data that could bring valuable insights, along with business and technical knowledge, are essential resources and skills that need to be ready at the start of the project. Building and maintaining an open and data-driven culture and gaining support from top management would facilitate the big data-enabled innovation project. Data scientists and other data enthusiasts could act as advocators and influence others to make the company more data-driven and open to data sharing and usage. Project managers could win top management support by proving the viability of the concept of the big data-enabled product and showing its reliability and value.

With respect to the external environment in which companies operate, many environmental uncertainty aspects influence the development of BDAC. Thus, practitioners should be aware of their impact and act upon it. To build good relationships with potential clients, they create products that match customers' needs and the company's capabilities. For instance, the company in Case 3 initiated its SaaS product in order to cater to the need of petrol stations to know the market status, and to this end harnessed its navigation data and system. Companies must also be aware of the new technologies available in the industry. For example, Chat GPT was enabled by generative pre-trained transformer models; shortly after its launch, many companies started building digital products based on it. Keeping an eye on the market and competitors' innovation actions is also important. Companies must understand that customers' and users' actions are not stable and could thus influence data collection. Increasing data variety might help, as different data sources could triangulate each other. When facing hostile market competition, companies should seek help from new technologies like big data and develop big data analytics capabilities in order to enable growth and the innovation of new products.

On the other hand, companies must comply with corporate social responsibility and carry out strict testing of prototypes to ensure reliability. Understanding the potential influence of privacy regulations and corporate social responsibility can remind decision-makers in big data innovation projects to be aware of them and keep themselves up to date or look ahead to anticipate possible future regulations. Regulations may be brought up to speed later than technology but will catch up eventually.

After considering the aspects that influence the development of BDAC in innovation projects, it is time to consider how to improve innovation performance. These findings could bring insight for practitioners involved in big data innovation projects, especially product managers, designers, data scientists and analysts, enabling them to understand how certain aspects of big data analytics capabilities could influence the results of product performance. To increase product accuracy, practitioners should get access to more data sources while complying with data policy and regulations and use closer-to-real-time data and analysis in their prediction models. To improve time-to-market, practitioners should shorten the time required to set up database and analysis systems so that the product development project can initiate and operate more quickly.

From a practical point of view, this study helps product managers, designers, data scientists, and people involved in big data product innovation projects to see the project more clearly. The findings of this study share a picture unveiling the mechanism behind the project regarding how the resources, skills, and capabilities are integrated and influence each other in forming the big data-enabled digital product. To improve product performance, big data analytics capabilities need to be developed well. To that end, companies need basic resources such as funding, a data-driven and open culture, and a variety of reliable data. On the other hand, business knowledge, technical knowledge, communication skills, and business analytics are also necessary. Product owners and managers alike could make good use of these resources they have access to and keep themselves informed about the actions of competitors, any new technologies that can be applied in the industry, and changes in customers' buying behaviour changes when initiating or adjusting innovation schemes.

9.4 Limitations and future research

This final part of the dissertation focuses on the limitations of this research project regarding the research design, data collection and analysis, followed by the actions taken to minimise the effect they might have on the research. Finally, future research directions are suggested.

First, this exploratory multiple case study aims to identify antecedents and consequences of big data analytics capabilities. Case studies are great for answering how and why questions (Yin, 2013), which is suitable for this research in exploring how factors influence each other. Multiple case studies provide more situations for investigation than a single case, enabling the researcher to explore more factors that could be generalised to cases in similar situations. Aiming at reaching saturation, the study could always benefit from more cases. Furthermore, the resulting framework highlights only the relationships between antecedents and consequences that are prominent in the interview data. Other factors might also have an impact but need to be evident in the interview data. However, this is always the case, considering the nature of the type of research. It is challenging, if not impossible, to exploit the data and extract all potential factors. Therefore, further exploration of the framework through other studies with different case samples is desirable to identify more factors and relationships. This study focused on four cases involving the production of digital products. Future studies could explore the factors that influence tangible products, perhaps in the smart wearable industry, where data are used as input for analysis algorithms in order to test whether the factors can be generalised with tangible products as well. Future studies could test this framework with large-scale quantitative empirical studies across different industries for corroboration and confirmation. This study suggests that structural equation model analysis should be conducted to test the antecedents and determine whether BDAC fully mediates the effects of the antecedents on the consequences of innovation performance.

Second, although the cases were carefully selected to aim at a balance, the findings are based on only four projects. Each case was characterised by a specific combination of social and economic factors and decision processes, and it is difficult to isolate their impact from these four cases. Most of the findings from the four cases can be generalised to other industries and companies , considering if using big data as input for algorithms to do predictive analysis to enable key functions of a digital product. Some findings might be more suitable for some cases, but not generalisable for other industries and companies. Thus, further quantitative studies are needed to test the associations of the antecedents and consequences identified in this study and find out which ones are generalisable and which ones are not. Quantitative studies that use surveys as the primary data collection method could reach more participants and help test whether the findings from this research project are significant.

The third point concerns the interview participants' potential biases. For this research, to understand the phenomenon, it was a good idea to interview people who were involved in the project, because other people do not know much about it. This unavoidably might have brought up biases in assessing the innovation performance. For example, the interviewees may have overestimated the performance of the product. However, the result of the assessment is not the focus of the output of this research, because the aim of this research is to find the relevant factors that might influence the result. So, the performance was not intentionally measured, but through the conversation it is possible to see the link between certain factors and their effect on the performance.

In addition, it could be beneficial to interview more participants, such as designers, to understand the user experience side of the project and what designers think about the performance of the product. However, the designers might not provide information regarding big data-related factors. It is also beneficial to do user interviews to understand what users think about the usability of the product, but our research focus is on big data and its effect on the product, so the input from users can provide some side input. Users cannot help in providing information that helps to identify the influencing factors of product performance, whereas people who participated in the development project can provide much information to help with this analysis.

Fourth, innovation with big data could pose many potential ethical issues, including privacy protection, equality and non-discrimination, autonomy, controlling one's own identity, transparency, solidarity, contextual integrity, property and copyright (Christen et al., 2019). However, this research was aiming at understanding the mechanism of influencing innovation performance in terms of project and product performance. It did not focus on ethical considerations, but this could be a very important area for future research.

Finally, although this study aimed to identify as many factors and relationships as possible, some factors and aspects remain to be explored. Table 9.1 below presents the summary of the main findings along with some directions for future research. Future studies could identify more antecedents and consequences and explore relationships between them. The following paragraphs explain these future directions in detail:

Regarding the relationships within BDAC, this research project has identified many tangible resources and human skills compared to intangible resources. Intangible resources are more subtle than the other two types of resources and skills. For example, organisational learning as an intangible resource is not easily observed. It might only stand out from the interview data if intentionally measured. Future studies could explore the impact of organisational learning on BDAC constructs. How does creating, retaining and transferring knowledge influence the development of BDAC?

This project revealed the relationships between several BDAC constructs and their impact on innovation performance. Constructs such as data variety, velocity and reliability could influence innovation performance, while volume does not. In addition, business analytics also influence product performance. However, it is not clear whether business knowledge has an impact on product or project performance. Business knowledge refers to recognising the value of big data and where to focus efforts to gain insights (Akter & Wamba, 2016). It also includes modelling and analysis, such as statistical analysis and predictive modelling. It would be logical to assume that improving analysis and models would boost product performance; however, this was not shown in this study, perhaps because it was not tested intentionally. This can be remedied by comparing the differences after optimising the analysis or model. On the other hand, it is also unclear whether the improvement of business technology would shorten or prolong the time to market. However, as optimisation takes time, an optimised analysis method could also shorten the data process.

This study identified that dynamic capabilities could influence BDAC and that, specifically, coordinating influences data variety. However, this is the only element of dynamic capability for which this study shows an impact. Future research could explore other aspects of dynamic capabilities, including sensing, learning, and integrating. These elements could have an impact on other constructs of BDAC. The factors did not emerge from this study, perhaps due to the data collection method. Sensing, learning, and integrating are challenging to observe from interview data. Future studies could adopt action research methods in which the researchers get involved in the business to observe them better.

Regarding environmental uncertainties, all three elements of environmental uncertainty are identified as antecedents of BDAC: dynamism, heterogeneity, and hostility. What remains to be clarified is whether these aspects of environmental uncertainty impact specific BDAC elements. For example, high hostility stimulates companies to develop their BDAC. Future studies could explore which elements of BDAC are particularly stimulated – whether data variety, velocity, or other elements – and the role of hostility.

In addition, aspects of environmental uncertainties still need to be discovered, such as other aspects of hostility, including the availability of key resources. For instance, how do changes in the availability of key resources influence BDAC?

In this study, several factors and relationships were revealed in one case but not in others. A few examples from the four cases are provided here. Future studies could further explore the impact of these factors:

The result from this research indicates that hardware is essential as a foundation for BDAC development and further benefits product performance. In Case 2, faster semiconductor chips could start a series of chain reactions, increase computing power, and lead to faster data processing and analysis, thus benefiting BDAC development and eventually improving product performance thanks to faster operation. As hardware plays a fundamental role, it is sometimes easily neglected. To test its impact, future research could include hardware as an independent factor in quantitative empirical studies to test whether the level of BDAC differs.

In Case 4, data velocity influenced time to market because the speed of data collection and analysis could influence how quickly the product would go through the innovation process and finally reach the customers. This was only shown in Case 4, and not in the others, because it would be noticeable when there was a delay in the process. Future studies could test this factor by investigating delayed product innovation projects as case studies.

Business knowledge is identified as an influencing factor for technical knowledge in Case 2. Specifically, innovation process management influences data management. Case 2 involved the internet industry, where the pace of innovation is rapid, with several quick sprints. In this context, data management, such as the data format type, was inconsistent. The phenomenon only showed up in Case 2, which could be explained due to the fast pace of the internet industry. Future studies could test whether this would occur among other internet companies by choosing to focus on internet companies as samples in multiple case studies to see whether this is a commonly shared phenomenon.

Uncertainty of the actions of customers also emerged from only one case in these research findings. In this case, business customers stopped sponsoring the product due to a lack of financial income. Future studies could explore this phenomenon by investigating cases where customers' actions are unstable, for example, start-ups.

Finally, corporate social responsibility (CSR) was found in Case 1 but not in the others. CSR has been found to hinder the development of BDAC. The reason why it might only be present in Case 1 is that, as a public transport provider, the company has to uphold greater CSR than other companies. Future studies could test whether this is true for other companies with significant societal responsibilities. For example, these studies could observe multiple pharmaceutical companies by interviewing people in research and development and management or engage in observation or join the team to do action research to see whether CSR hinders their intention to use big data to develop new medicines.

This section identifies limitations and suggests future study directions. Future studies could explore this topic through other data collection methods, such as action research, in which researchers become part of the group they are studying. Future studies could also test the factors and relationships presented in these findings with large-scale quantitative studies in multiple industries. In this way, the research on using big data for product innovation would become more comprehensive.

Table 9.1 Main findings from this research and suggestions for future research directions

The main findings from this study	Suggestions for future directions
1. Revealing there are relationships within BDAC: • Communication skills & data variety • Business knowledge & technical knowledge • Financial resources & BDAC • Hardware & BDAC • Data & BDAC • Company culture & BDAC • Business knowledge & BDAC • Technical knowledge & BDAC	Explore and test more relationships between BDAC constructs in addition to what this research has found. For instance, do intangible resources such as organisational learning influence BDAC elements and, if so, how?
2. Exploring the influence of BDAC constructs on innovation performance constructs: • Data variety & product performance • Data reliability & product performance • Business analytics & product performance • Data velocity & project performance	Future research could explore the relationships between other constructs of BDAC and innovation results. For instance, does business knowledge influence product performance and, if so, how?
3. Revealing relationships between dynamic capabilities and BDAC that are different from current studies: • Coordinating & BDAC	Future research could test whether coordinating along with other constructs of dynamic capabilities such as sensing, learning, and integrating could have any impact on other constructs of BDAC.
4. Identified new relationships between environmental uncertainty and BDAC: • Heterogeneity & BDAC • Dynamism & BDAC • Hostility & BDAC • Corporate social responsibility & BDAC • Privacy & regulations • Corporate social responsibility	• Future research could test whether environmental uncertainty can influence specific constructs of BDAC. For instance, does hostility influence specific BDAC constructs like data variety and velocity? • Future research could also explore whether other aspects of environmental uncertainty could influence BDAC. For instance, other aspects of hostility, such as the availability of key resources. How do changes in the availability of key resources influence BDAC?
5. Antecedents of developing BDAC and innovation performance were identified, but they emerged only from one case in this research: • Hardware & BDAC • Data velocity & project performance • Business knowledge & technical knowledge • Dynamism – uncertainty of the actions of customers/users & BDAC • Corporate social responsibility & BDAC	Future research could explore these factors in other scenarios. Perhaps these factors are more prominent in other project settings.

9.5 Future perspective

Looking ahead, it is to be expected that big data will revolutionise digital product innovation while keeping morality and responsibility at its heart. Big data-enabled personalisation is here to stay. However, the focus will shift from using these products to capture attention to creating inspiring products, stimulating thinking, and improving human intelligence. Future digital products should also maintain a respectful distance from human life, allowing non-digital activities to have their space.

Product design should include mechanisms to avoid overuse so as to maintain real-world social connections, especially to protect vulnerable groups such as children.

At the moment, big data is widely used in tech-savvy industries. In the future, more big data-enabled products will continue to extend their applications in manufacturing, energy, and agriculture. There will also be more big data-enabled digital products in healthcare and education and products that offer solutions or mitigate social issues such as discrimination, political polarisation, and privacy protection. Moreover, big data-enabled products are set to support sustainability efforts, empowering industries to make more informed decisions and reduce their environmental footprint.

References

References

Abernathy, W. J., & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. *Research Policy*, *14*(1), 3-22.

Abrahamsson, P., Salo, O., Ronkainen, J., & Warsta, J. (2017). Agile software development methods: Review and analysis. *arXiv preprint arXiv:1709.08439*.

Agmon, N., & Ahituv, N. (1987). Assessing data reliability in an information system. *Journal of Management Information Systems*, *4*(2), 34-44.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics, 182*, 113-131.

Ahmed, P., & Shepherd, C. D. (2010). *Innovation Management: Context, Strategies, Systems and Processes*. (1 ed.) Pearson.

Amatriain, X. (2013, August). Big & personal: data and models behind Netflix recommendations. In *Proceedings of the 2nd international workshop on big data, streams and heterogeneous source mining: Algorithms, systems, programming models and applications* (pp. 1-6).

Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of Management, 40*(5), 1297–1333.

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, *15*(2), 94-101.

Ansoff, H. I. (1957). Strategies for diversification. *Harvard Business Review*, *35*(5), 113-124.

Appio, F. P., Frattini, F., Petruzzelli, A. M., & Neirotti, P. (2021). Digital transformation and innovation management: A synthesis of existing research and an agenda for future studies. *Journal of Product Innovation Management*, *38*(1), 4-20.

Argote, L., McEvily, B., & Reagans, R. (2003). Managing knowledge in organisations: An integrative framework and review of emerging themes. *Management Science*, *49*(4), 571-582.

Argote, L. (2012). *Organizational learning: Creating, retaining and transferring knowledge*. Springer Science & Business Media.

Art. 4 GDPR – Definitions. (n.d.). General Data Protection Regulation (GDPR). Retrieved February 10, 2022, from https://gdpr-info.eu/art-4-gdpr/

Bagno, R. B., Salerno, M. S., & da Silva, D. O. (2017). Models with graphical representation for innovation management: A literature review. *R&D Management*, *47*(4), 637-653.

Baker, W., Kiewell, D., & Winkler, G. (2014). Using big data to make better pricing decisions. *McKinsey analysis*.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99-120.

Bashir, N., Papamichail, K. N., & Malik, K. (2017). Use of social media applications for supporting new product development processes in multinational corporations. *Technological Forecasting and Social Change*, *120*, 176-183.

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019b). Big data adoption: State of the art and research challenges. *Information Processing & Management, 56*(6). https://doi.org/10.1016/j.ipm.2019.102095

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R. and Kern, J., 2001. *The agile manifesto.*

Bekker, A. (2022, July 19). 4 types of data analytics to improve decision-making. *ScienceSoft*. https://www.scnsoft.com/blog/4-types-of-data-analytics

Bharadwaj, A., El Sawy, O., Pavlou, P., Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quart. 37* (2), 471–482.

Bilbao-Osorio, B., & Rodríguez-Pose, A. (2004). From R&D to innovation and economic growth in the EU. *Growth and Change*, *35*(4), 434-455.

Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organisational research: Some implications for a theory of work performance. *Academy of Management Review*, *7*(4), 560-569.

Borovicka, T., Jirina Jr., M., Kordik, P., & Jirina, M. (2012). *Selecting Representative Data Sets* (Reviewed ed., Vol. 44). IntechOpen. https://doi.org/10.5772/50787

Breunig, K. J., Aas, T. H., & Hydle, K. M. (2014). Incentives and performance measures for open innovation practices. *Measuring Business Excellence*, *18*(1), 45-54.

Breznitz, D., & Ornston, D. (2018). The politics of partial success: fostering innovation in innovation policy in an era of heightened public scrutiny. *Socio-Economic Review*, *16*(4), 721-741.

Brown, T. (2008). Design thinking. *Harvard Business Review*, *86*(6), 84.

Bruneel, J., Ratinho, T., Clarysse, B., & Groen, A. (2012). The evolution of business incubators: Comparing demand and supply of business incubation services across different incubator generations. *Technovation*, *32*(2), 110-121.

Brunswicker, S., & Schecter, A. (2019). Coherence or flexibility? The paradox of change for developers' digital innovation trajectory on open platforms. *Research Policy*, *48*(8), 103771.

Brush, K. (2022). Real-time analytics – customer experience. https://www.techtarget.com/searchcustomerexperience/definition/real-time-analytics

Bughin, J., & van Zeebroeck, N. (2017). The best response to digital disruption. *MIT Sloan Management Review* 58 (4): 80–6.

Cakebread, C. (2017). You're not alone, no one reads terms of service agreements. *Business Insider*. Retrieved February 14, 2024, from https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11?international=true&r=US&IR=T

Cambridge dictionary. (2024). autonomy. In *Cambridge Dictionary*. Retrieved February 12, 2024, from https://dictionary.cambridge.org/dictionary/english/discrimination

Cambridge dictionary. (2024). discrimination. In *Cambridge Dictionary*. Retrieved February 12, 2024, from https://dictionary.cambridge.org/dictionary/english/discrimination

Cocchi, N., Dosi, C., & Vignoli, M. (2021). The hybrid model matrix: Enhancing stage-gate with design thinking, lean startup, and agile. *Research-Technology Management*, *64*(5), 18-30.

Collins. (2019). Technology. In Collins. Harper Collins Publishers Ltd. https://www.collinsdictionary.com/dictionary/english/technology#:~:text=Technology%20refers%20to%20methods%2C%20systems,Technology%20is%20changing%20fast.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, *30*, 44-57.

R

Cooper, R. G. (1986). *Winning at new products* (pp. 1-14). Reading, MA: Addison-Wesley.

Cooper, R. G., & Kleinschmidt, E. J. (1986). An investigation into the new product process: Steps, deficiencies, and impact. *Journal of Product Innovation Management*, *3*(2), 71-85.

Cooper, R. G. (2001). *Winning at new products.* Cambridge, MA: Perseus Publications.

Cooper, R. G. (2006). Managing technology development projects. *Research-Technology Management, 49*(6), 23–31. https://doi.org/10.1080/08956308.2006.11657405

Cooper, R. G., & Kleinschmidt, E. J. (2011). *New products: The key factors in success*. Marketing Classics Press.

Corbin, J. C., & Strauss, A. S. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). SAGE Publications, Inc.

Cote, C. (2021, October 19). 4 types of data analytics to improve decision-making. *Business Insights Blog*. https://online.hbs.edu/blog/post/types-of-data-analysis

Cozzolino, A., Verona, G., & Rothaermel, F. T. (2018). Unpacking the disruption process: New technology, business models, and incumbent adaptation. *Journal of Management Studies*, *55*(7), 1166-1202.

Creswell, J. W., & Creswell, D. J. (2018). *Research design: qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications, Inc.

Cukier, K. (2010). Data, data everywhere. *Economist*, *394*(8671), 3-5.

Deka, G. C. (2014). Big data predictive and prescriptive analytics. In *Handbook of research on cloud infrastructures for big data analytics* (pp. 370-391). IGI Global.

Deshpande, P. S., Sharma, S. C., & Peddoju, S. K. (2019). Predictive and prescriptive analytics in big-data era. In *Security and data storage aspect in cloud computing* (pp. 71-81). Springer, Singapore.

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*.

Diebold, F. X.(2012). On the origin(s) and development of the term "big data". Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, Tech. Rep. https://economics.sas.upenn.edu/sites/default/files/filevault/12-037.pdf

Dijcks, J. (2013). Oracle: big data for the enterprise. Oracle White Paper, Oracle Corporation, Redwood Shores, CA.

Dougherty, D., Borrelli, L., Munir, K., & O'Sullivan, A. (2000). Systems of organizational sensemaking for sustained product innovation. *Journal of Engineering and Technology Management*, *17*(3-4), 321-355.

Drechsler, K., Gregory, R., Wagner, H. T., & Tumbas, S. (2020). At the crossroads between digital innovation and digital transformation. *Communications of the Association for Information Systems*, *47*(1), 23.

Ducange, P., Pecori, R., & Mezzina, P. (2017). A glimpse on big data analytics in the framework of marketing strategies. *Soft Computing, 22*(1), 325–342. https://doi.org/10.1007/s00500-017-2536-4

Dykes, B. (2017, June 28). Big data: Forget volume and variety, focus on velocity. Forbes. https://www.forbes.com/sites/brentdykes/2017/06/28/big-data-forget-volume-and-variety-focus-on-velocity/?sh=2645248e6f7d

El Morr, C., & Ali-Hassan, H. (2019). Analytics building blocks. In *Analytics in healthcare* (pp. 15-30). Springer, Cham.

Etymonline. (n.d.). *data.* Etymonline. Retrieved September 10, 2021, from https://www.etymonline.com/search?q=data

European Union. (2022, January 7). Data protection and online privacy. Europa. Retrieved July 20, 2023, from https://europa.eu/youreurope/citizens/consumers/internet-telecoms/data-protection-online-privacy/index_en.htm

Fabian, C., & Fabricant, R. (2014). *The Ethics of Innovation (SSIR)*. Stanford Social Innovation Review. Retrieved February 5, 2024, from https://ssir.org/articles/entry/the_ethics_of_innovation#bio-footer

Fagerberg, J. (2009). Innovation: A guide to the literature. In *The Oxford handbook of innovation*. Oxford University Press.

Fagerberg, J., & Godinho, M. M. (2006). Innovation and catching-up. In *The Oxford handbook of innovation*. Oxford University Press.

Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: Impact on firm performance. *Management Decision, 57*(8), 1923–1936. https://doi.org/10.1108/md-07-2018-0825

Forbes Insights. (2015). *Betting on big data: How the right culture, strategy and investments can help you leapfrog the competition.* www.forbes.com/forbesinsights

Gandomi, A., & Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management,* 35(2), 137-144.

Gartner. (2019). Comments off on our top data and analytics predicts for 2019. Retrieved from https://blogs.gartner.com/andrew_white/2019/01/03/our-top-dataand-analytics-predicts-for-2019/

Gartner. (n.d.). Definition of real-time analytics – Gartner information technology glossary. https://www.gartner.com/en/information-technology/glossary/real-time-analytics#:~:text=Real%2Dtime%20analytics%20is%20the,the%20arrival%20of%20new%20data.

Gaubinger, K., Rabl, M., Swan, S., & Werani, T. (2015). Innovation and product management. *Innovation and product management: A holistic and practical approach to uncertainty reduction*, 83-113.

Giannopoulou, E., Gryszkiewicz, L., & Barlatier, P. J. (2014). Creativity for service innovation: A practice-based perspective. *Managing Service Quality: An International Journal*, *24*(1), 23-44.

Gibson, P. (2018, February 23). Types of data analysis. Chartio. https://chartio.com/learn/data-analytics/types-of-data-analysis/

Gimmon, E., & Levie, J. (2010). Founder's human capital, external investment, and the survival of new high-technology ventures. *Research Policy*, *39*(9), 1214-1226.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking qualitative rigor in inductive research. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Gobble, M. M. (2013). Big data: The next big thing in innovation. *Research-Technology Management*, *56*(1), 64-67.

Gordon, G. G. (1991). Industry determinants of organizational culture. *Academy of Management Review, 16*(2), 396–415. https://doi.org/10.5465/amr.1991.4278959

Grant, R. M. (1991). The resource-based theory of competitive advantage: Implications for strategy formulation. *California Management Review*, *33*(3), 114-135.

Größler, A., & Grübner, A. (2006). An empirical model of the relationships between manufacturing capabilities. *International Journal of Operations & Production Management*, *26*(5), 458-485.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information &*

Management, 53(8), 1049–1064. https://doi.org/10.1016/j.im.2016.07.004

Hadjimanolis, A. (2000). A resource-based view of innovativeness in small firms. *Technology Analysis & Strategic Management*, *12*(2), 263-281.

Harrison, J. S., & St. John, C. H. (1996). Managing and partnering with external stakeholders. *Academy of Management Perspectives*, *10*(2), 46-60.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. *Cambridge Service Alliance*, 1-29.

Hinings, B., Gegenhuber, T., & Greenwood, R. (2018). Digital innovation and transformation: An institutional perspective. *Information and Organization*, *28*(1), 52-61.

Hollinger, P. (2015, June 12). Smarter aircraft create a wealth of data but it remains underexploited. *Financial Times.* https://www.ft.com/content/3f956a92-0943-11e5-b643-00144feabdc0

Hopkins, M. (2005). Measurement of corporate social responsibility. *International Journal of Management and Decision Making*, *6*(3-4), 213-231.

Hornby, A. (2020). *Oxford advanced learners dictionary.*

Hultink, E.J., Griffin, A., Robben, H.S.J., Hart, S., 1998. In search of generic launch strategies for new products. *International Journal of Research in Marketing 15* (3), 69–285.

IDC. (2022, May). *Worldwide IDC global DataSphere forecast, 2022-2026: Enterprise organizations driving most of the data growth.* IDC: The Premier Global Market Intelligence Company. https://www.idc.com/getdoc.jsp?containerId=US49018922

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: A technological perspective and review. *Journal of Big Data*, *3*, 1-25.

Jauch, L. R., & Kraft, K. L. (1986). Strategic management of uncertainty. *Academy of Management Review*, *11*(4), 777-790.

Johnson, J. S., Friend, S. B., & Lee, H. S. (2017). Big data facilitation, utilisation, and monetisation: Exploring the 3Vs in a new product development process. *Journal of Product Innovation Management*, *34*(5), 640-658.

Kaplinsky, R., & Morris, M. (2000). *A handbook for value chain research* (Vol. 113). Brighton: University of Sussex, Institute of Development Studies.

Kappler, K., Schrape, J. F., Ulbricht, L., & Weyer, J. (2018). Societal implications of big data. *KI-Künstliche Intelligenz*, *32*, 55-60.

Keegan, D. P., Eiler, R. G., & Jones, C. R. (1989). Are your performance measures obsolete? *Strategic Finance, 70*(12), 45.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management, 34*(3), 387–394. https://doi.org/10.1016/j.ijinfomgt.2014.02.002

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management science*, *65*(7), 2966-2981.

Langerak, F., Hultink, E. J., & Robben, H. S. (2004). The impact of market orientation, product advantage, and launch proficiency on new product performance and organizational performance. *Journal of Product Innovation Management*, *21*(2), 79-94.

Lee, J., & Berente, N. (2012). Digital innovation and the division of innovative labor: Digital controls in the automotive industry. *Organization Science*, *23*(5), 1428-1447.

Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, *50*, 57-70.

Li, Y., & Chen, L. (2014). Big biological data: Challenges and opportunities. *Genomics, Proteomics & Bioinformatics*, *12*(5), 187.

Lippett, M. (2022, May 5). Privacy, intelligence, agency: Security in the smart home. *Forbes.* https://www.forbes.com/sites/forbestechcouncil/2022/05/05/privacy-intelligence-agency-security-in-the-smart-home/

Lusch, R. F., & Nambisan, S. (2015). Service innovation. *MIS Quarterly*, *39*(1), 155-176.

Maclnnis, D. J., Moorman, C., & Jaworski, B. J. (1991). Enhancing and measuring consumers' motivation, opportunity, and ability to process brand information from ads. *Journal of Marketing*, *55*(4), 32-53.

Marr, B. (2016). *Big data in practice: How 45 successful companies used big data analytics to deliver extraordinary results*. John Wiley & Sons.

Marr, B. (2021, August 12). Walmart: Big data analytics at the world's biggest retailer. Bernard Marr. https://bernardmarr.com/walmart-big-data-analytics-at-the-worlds-biggest-retailer/

Matt, C., Hess, T., & Benlian, A. (2015). Digital transformation strategies. *Business & Information Systems Engineering*, *57*(5), 339-343.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured text analytics to support new product development decisions. *Research-Technology Management*, *58*(2), 30-39.

Marr, B. (2015, April 27). Big data-as-a-service is next big thing. Forbes. https://www.forbes.com/sites/bernardmarr/2015/04/27/big-data-as-a-service-is-next-big-thing/?sh=2a38d21733d5

Marr, B. (2021, July 13). How much data do we create every day? The mind-blowing stats everyone should read. *Bernard Marr*. https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/

Marshall, A., Mueck, S., & Shockley, R. (2015). How leading organisations use big data and analytics to innovate. *Strategy & Leadership, 43*(5), 32-39.

Martin, K. D., & Murphy, P. E. (2017). The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, *45*(2), 135-155.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Mercier, K., Richards, B., & Shockley, R. (2013). Analytics: The real-world use of big data in retail. *Executive Report, IBM Global Business Services, Business Analytics and Optimization, Somers, NY*.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review 90*(10): 60–68.

McAvoy, J., & Butler, T. (2018). A critical realist method for applied business research. *Journal of Critical Realism*, *17*(2), 160-175.

Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A data quality in use model for big data. *Future Generation Computer Systems*, 63, 123–130. https://doi.org/10.1016/j.future.2015.11.024

Metcalfe, J. S. (1998) *Evolutionary economics and creative destruction*. London: Routledge

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management, 30*(2), 272–298. https://doi.org/10.1111/1467-8551.12343

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and E-Business Management, 16*(3), 547–578. https://doi.org/10.1007/s10257-017-0362-y

Miller, D., & Friesen, P. H. (1983). Strategy-making and environment: the third link. *Strategic Management Journal*, *4*(3), 221-235.

Morey, T., Forbath, T., & Schoop, A. (2015). Customer data: Designing for transparency and trust. *Harvard Business Review*, *93*(5), 96-105.

Mullins, J., & Walker, O. (2012). *Marketing management: A strategic decision-making approach* (8th ed.). McGraw Hill.

Muro, M., & Katz, B. (2011). The new “cluster moment”: How regional innovation clusters can foster the next economy. In *Entrepreneurship and global competitiveness in regional economies: Determinants and policy implications*. Emerald Group Publishing Limited.

Murphy, C. B. (2023). Tangible assets vs. intangible assets: What’s the difference? *Investopedia.* https://www.investopedia.com/ask/answers/012815/what-difference-between-tangible-and-intangible-assets.asp

Mylavarapu, G., Thomas, J. P., & Viswanathan, K. A. (2019, March). An automated big data accuracy assessment tool. In *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)* (pp. 193-197). IEEE.

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management: Reinventing innovation management research in a digital world. *MIS Quarterly*, *41*(1).

Newkirk, H. E., & Lederer, A. L. (2006). The effectiveness of strategic information systems planning under environmental uncertainty. *Information & Management*, *43*(4), 481-501.

Niebel, T., Rasel, F., & Viete, S. (2019). Big data–BIG gains? Understanding the link between big data analytics and innovation. *Economics of Innovation and New Technology*, *28*(3), 296-316.

O’Leary, D. E. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems*, *28*(2), 96-99.

Oliveira, T., Thomas, M., Baptista, G., & Campos, F. (2016). Mobile payment: Understanding the determinants of customer adoption and intention to recommend the technology. *Computers in Human Behavior, 61*, 404–414. https://doi.org/10.1016/j.chb.2016.03.030

Organisation for Economic Co-operation and Development, OECD. (2005). *The measurement of scientific and technological activities Oslo manual: Guidelines for collecting and interpreting innovation data* (3rd ed.). OECD Publishing.

OECD. (2015). *OECD innovation strategy 2015. An agenda for policy action.* Paris/Eurostat, Luxembourg: OECD Publishing.

Origin-destination (flow) data - Office for National Statistics. (2022, May 11). Office for National Statistics. Retrieved August 11, 2022, from https://www.ons.gov.uk/census/censustransformationprogramme/census2021outputs/2021dataproducts/origindestination

Packin, N. G. (2019). Social credit: much more than your traditional financial credit score data. *Forbes*. Retrieved February 13, 2024, from https://www.forbes.com/sites/nizangpackin/2019/12/13/social-credit-much-more-than-your-traditional-financial-credit-score-data/

Parent, W. A. (1983). Recent work on the concept of privacy. *American Philosophical Quarterly*, *20*(4), 341-355.

Pearson, T., & Wegener, R. (2013). Big data: the organisational challenge. *Bain Co*.

Pence, H. E. (2014). What is big data and why is it important? *Journal of Educational Technology Systems*, *43*(2), 159-171.

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big data and AI–A transformational shift for government: So, what next for research? *Public Policy and Administration*, *35*(1), 24-44.

Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers*, *20*(2), 209-222.

Power, D. J., Heavin, C., McDermott, J., & Daly, M. (2018). Defining business analytics: An empirical approach. *Journal of Business Analytics*, *1*(1), 40-53.

Project Management Institute. (2013). *A guide to the project management body of knowledge (PMBOK® Guide)* (Fifth Edition). Project Management Institute.

Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management*, *38*(1), 187-195.

Ransbotham, S., Kiron, D. & Prentice, P. K. (2016). Beyond the hype: The hard work behind analytics success. *MIT Sloan Management Review, 57*(3), 6.

Rathore, M. M., Shah, S. A., Shukla, D., Bentafat, E., & Bakiras, S. (2021). The role of AI, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. *IEEE Access*, *9*, 32030-32052.

Rice, M. (2023). *What is a data platform? 28 Examples of big data platforms to know.* Built In. https://builtin.com/big-data/big-data-platform

Rosenberg, N. (2006). Innovation and economic growth. In *Innovation and growth in tourism*, OECD Publishing, Paris, https://doi.org/10.1787/9789264025028-4-en.

Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE.

Saldanha, T. J., Mithas, S., & Krishnan, M. S. (2017). Leveraging customer involvement for fueling innovation: The role of relational and analytical information processing capabilities. *MIS Quarterly*, *41*(1).

SAS. (n.d.). Predictive analytics: What it is and why it matters. SAS. Retrieved November 14, 2022, from https://www.sas.com/en_us/insights/analytics/predictive-analytics.html

Savage, N. (2022). Breaking into the black box of artificial intelligence. *Nature*.

Schilling, M. A., & Hill, C. (1998). Managing the new product development process: Strategic imperatives. *Academy of Management Executive, 12*(3), 67–81.

Sandelowski, M. (1993). Theory unmasked: The uses and guises of theory in qualitative research. *Research in Nursing & Health*, *16*(3), 213-218.

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, *60*, 102331.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D. & Tufano, P. (2012). *Analytics: The real-world use of big data,* IBM Institute for Business Value, Said Business School, New York, NY.

Schumpeter, J.A. (1934). *The theory of economic development.* Harvard University Press, Cambridge, MA.

Sharma, A. K., Sharma, D. M., Purohit, N., Rout, S. K., & Sharma, S. A. (2022). Analytics techniques: Descriptive analytics, predictive analytics, and prescriptive analytics. In *Decision intelligence analytics and the implementation of strategic business management* (pp. 1-14). Springer, Cham.

Sharon, T. (2021). Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*, *23*(Suppl 1), 45-57.

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S., & Robinson, G. E. (2015). Big data: Astronomical or genomical? *PLOS Biology, 13*(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195

Sun, S., Cegielski, C. G., Jia, L., & Hall, D. J. (2018). Understanding the factors affecting the organisational adoption of big data. *Journal of Computer Information Systems*, *58*(3), 193-203.

Svahn, F., Mathiassen, L., & Lindgren, R. (2017). Embracing digital innovation in incumbent firms: How Volvo cars managed competing concerns. *MIS Q.*, *41*(1), 239-253.

Tabrizi, B., Lam, E., Girard, K., & Irvin, V. (2019). Digital transformation is not about technology. *Harvard Business Review*, *13*(March), 1-6.

Talend. (n.d.). Big data and agriculture: A complete guide. Talend – a Leader in Data Integration & Data Integrity. https://www.talend.com/resources/big-data-agriculture/

Teece, D. J. (2015). Intangible assets and a theory of heterogeneous firms. In A. Bounfour and T. Miyagawa (eds), *Intangibles market failure and innovation performance*, pp. 217–239. Cham: Springer.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509-533.

The Economist. (2014, May 22). Digital disruption on the farm. https://www.economist.com/business/2014/05/24/digital-disruption-on-the-farm

Thornhill, S. (2006). Knowledge, innovation and firm performance in high-and low-technology regimes. *Journal of Business Venturing*, *21*(5), 687-703.

Troilo, G., De Luca, L. M., & Guenzi, P. (2017). Linking data-rich environments with service innovation in incumbent firms: A conceptual framework and research propositions. *Journal of Product Innovation Management, 34*(5), 617-639.

Vance. (2013, October 2). Monsanto's billion-dollar bet brings big data to the farm. Bloomberg. Retrieved December 5, 2022, from https://www.bloomberg.com/news/articles/2013-10-02/monsanto-buys-climate-corporation-for-930-million-bringing-big-data-to-the-farm?leadSource=uverify%20wall

Verstegen, L., Houkes, W., & Reymen, I. (2019). Configuring collective digital-technology usage in dynamic and complex design practices. *Research Policy*, *48*(8), 103696.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118-144.

Varadarajan, R. (2018). Innovation, innovation strategy, and strategic innovation. In *Innovation and strategy* (pp. 143-166). Emerald Publishing Limited.

Waltman, L., Van Eck, N. J., & Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*(4), 629-635.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact:

Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234-246.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research,* 70, 356-365.

Wan-Jan, W. S. (2006). Defining corporate social responsibility. *Journal of Public Affairs: An International Journal*, *6*(3-4), 176-184.

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, *176*, 98-110.

Wang, Z., & Wang, N. (2012). Knowledge sharing, innovation and firm performance. *Expert Systems with Applications, 39*(10), 8899-8908.

Wang, L., Yang, M., Pathan, Z. H., Salam, S., Shahzad, K., & Zeng, J. (2018). Analysis of influencing factors of big data adoption in Chinese enterprises using DANP technique. *Sustainability*, *10*(11), 3956.

Webster Jr, F. E., & Wind, Y. (1972). A general model for understanding organizational buying behavior. *Journal of Marketing*, *36*(2), 12-19.

Wegener, R., & Sinha, V. (2013). *The value of big data: How analytics differentiates winners.* Bain and Company.

Westerman, G., Bonnet, D., & McAfee, A. (2014). The nine elements of digital transformation. *MIT Sloan Management Review*, *55*(3), 1-6.

White, M. (2012). Digital workplaces: Vision and reality. *Business Information Review, 29*(4), 205–214.

Wilberg, J., Schäfer, F., Kandlbinder, P., Hollauer, C., Omer, M., & Lindemann, U. (2017, December). Data analytics in product development: Implications from expert interviews. *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 818-822). IEEE.

Witell, L., Snyder, H., Gustafsson, A., Fombelle, P., & Kristensson, P. (2016). Defining service innovation: A review and synthesis. *Journal of Business Research*, *69*(8), 2863-2872.

Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive, 12*(2).

Wu, S. J., Melnyk, S. A., & Flynn, B. B. (2010). Operational capabilities: The secret ingredient. *Decision Sciences*, *41*(4), 721-754.

Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research, 69*(5), 1562–1566.

Yin, R. K. (2013, May 10). *Case study research: Design and methods (applied social research methods) (Fifth)*. SAGE Publications, Inc.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary—the new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research*, *21*(4), 724-735.

Zeng, Y., & Gu, P. (1999). A science-based approach to product design theory Part II: Formulation of design requirements and products. *Robotics and Computer-Integrated Manufacturing, 15*(4), 341-352.

APPENDIX

Abbreviated Interview Guide
Detailed conceptual models
Acknowledgements
About the Author

Abbreviated Interview Guide

Big data definition

- What is “big data” for you? What would you consider to be “big data”?

Big data innovation projects

- Could you think of a recent new service/product development project you were involved in that uses big data?

Project background

- How would you like to describe this project? What kind of product was involved?
- Were there similar products within your company and in the market?
- How different/or similar is it from the rest?

Project origin and stages

- What are the motivations/origins for this project? How did it start?
- How do you divide this innovation project in different stages?

Project participants

- Who were directly involved in this development project and what’s their role in it?
- How was the collaboration between these people?

Big data sources

- What sources of data are involved in this project?
- How can these data help you with the design and development of this service?

Big data analysis

- How do you combine all the data sources and analyse them?
- How do you interpret the data to bring insights?

Results

- What insights did you expect big data to bring?
- And what’s the result, what insight did big data bring in the innovation process?

The results

- How do you measure the impact or influence of using big data in this project?
- How happy you are with the result? Any improvement can be made?

&

Company characteristics

- What's the structure of departments in your company?
- How would you describe your company culture?

Environmental context

- How would you describe the market environment your company is in?

Challenges

- What challenges or barriers to using and creating value from big data do you see (managerial and technical)?

&

Detailed conceptual models

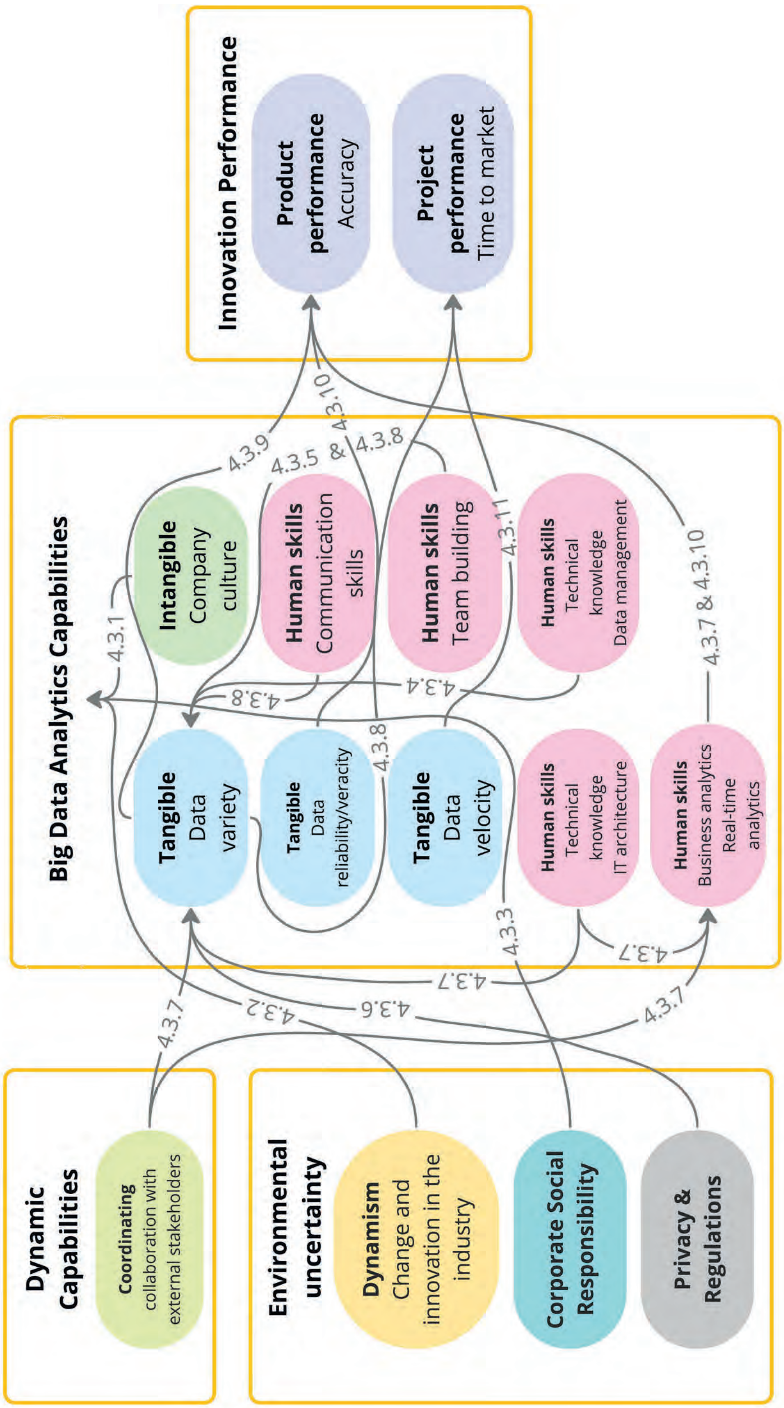


Figure A1. Detailed conceptual model of antecedents and consequences of big data usage in innovation in Case 1.

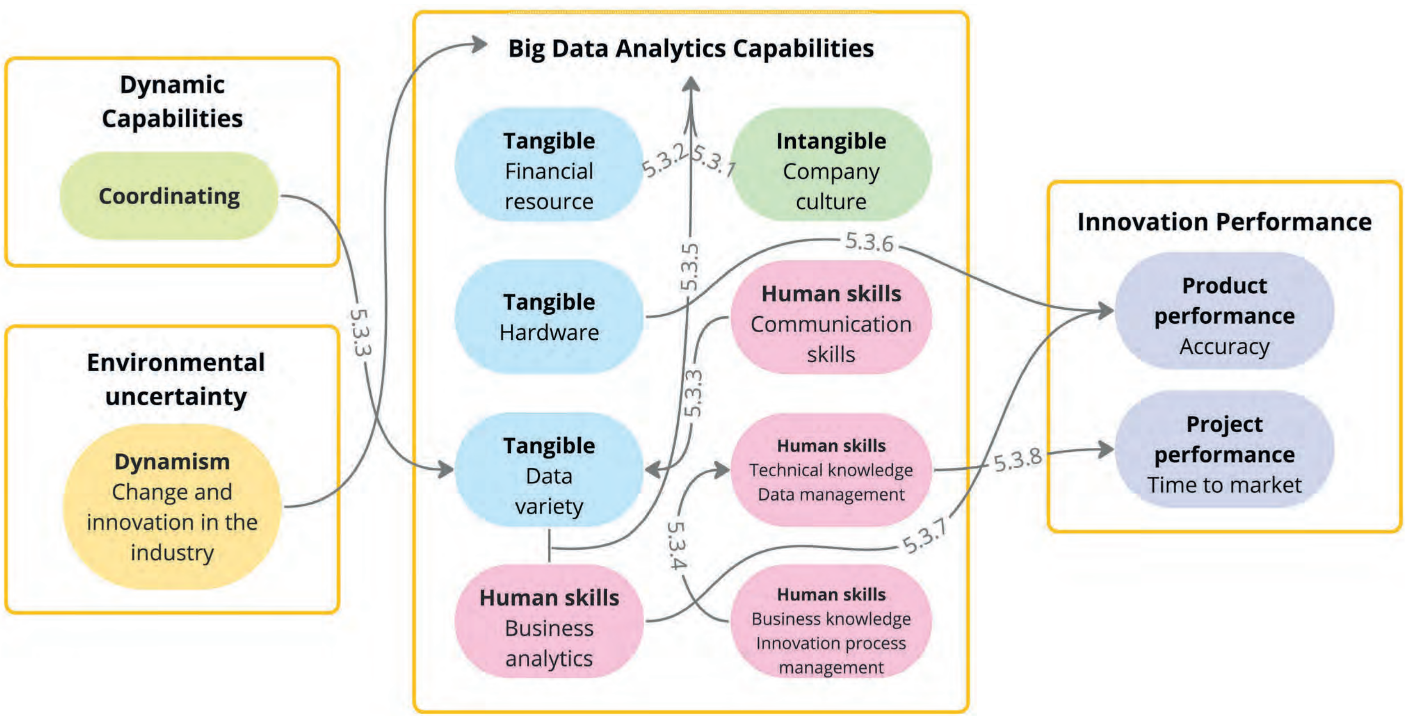


Figure A2. Detailed conceptual model of antecedents and consequences of big data usage in innovation in Case 2.

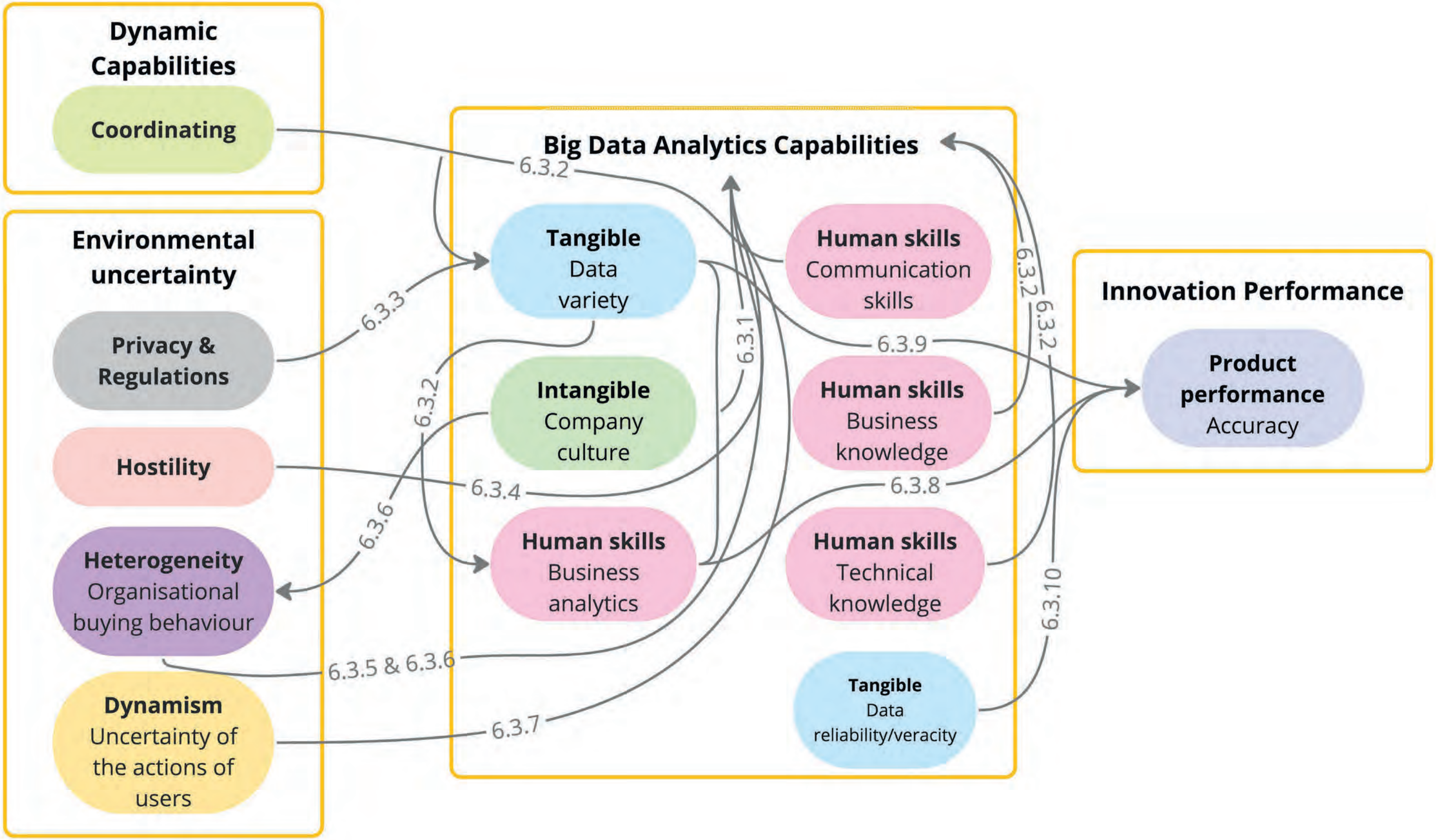


Figure A3. Detailed conceptual model of antecedents and consequences of big data usage in innovation in Case 3.

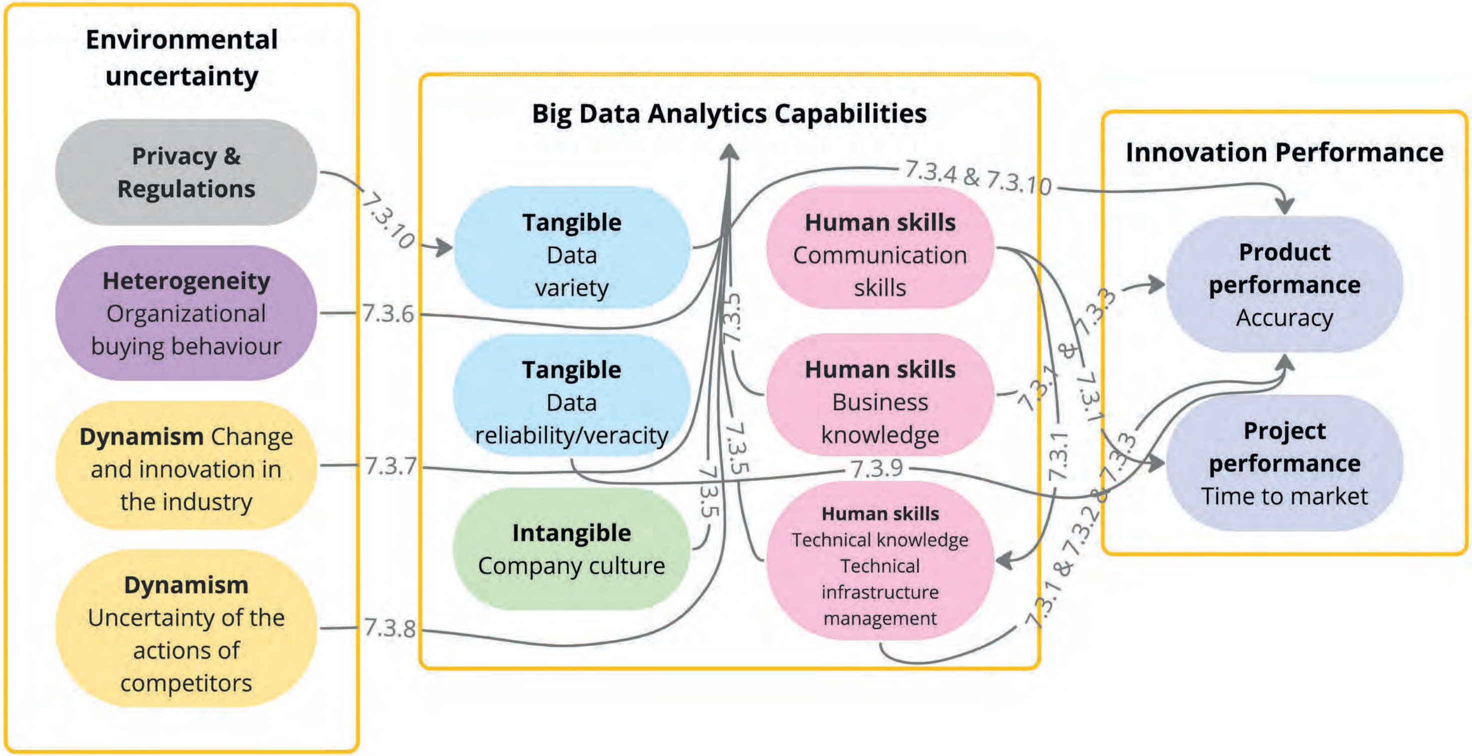


Figure A4. Detailed conceptual model of antecedents and consequences of big data usage in innovation in Case 4.

Acknowledgements

PhD projects usually take four or five years, which seems like a long time, but the journey runs very quickly. Despite the feeling that it went so quickly, it was also full of ups and downs. Reflecting on this journey, I feel thankful for what I've learned and for the people whom I've learned from.

I would like to say thank you to my promotor and co-promotor Erik and Pinar. You guided me directions to keep me on track, provided practical guidance on tackling tasks and made me think more critically by asking 'why' questions. I appreciated your professional feedback as well as your humour, making our progress meetings not only constructive but also fun! You have witnessed the highs and lows of my journey, and together, we made it to the end.

I would like to thank my doctoral committee members Dirk, Sicco, Jan Carel, Tucker, and Peter. Thank you for accepting the invitation to my defence, reading through this dissertation, and contributing to advancing the knowledge to the research topic.

This research project also received help from Guoqi, Euiyoung, Mignon, Milene, Sacha, Wilfred, Jacky and Kars. Thank you for asking questions, sharing insights, and helping me gaining access to the pilot interviews.

This study would not exist without the support from my contact persons in academic and industry for gaining access to the companies and connecting to the case study interview participants. Thank you for offering the opportunity.

I would like to say thank you to my participants for the pilot interviews and the data collection interviews. Thank you for answering all my questions and sharing your knowledge and experience working in big data enabled innovation projects.

To all the PhD fellows I've spent time with during my whole PhD journey, thank you for not only your inspiring talks but also your fun and warm company.

To whom I shared good news, jokes, and complaints with, had fun with, and reflected about the PhD journey with, Ahmee, Beatriz, Emilia, Garoa, Gerbera, Helen, Jiwon, Julieta, Jelle, Karl, Linda, Mahshid, Noa, Renske, Puck, Sagar, Xu Jun, Xueqing, thank you for making this PhD journey less painful, more encouraging, and fun! I was lucky to experience the PhD journey along with you.

I would like to say thank you to my lovely paranymphs Renske and Helen who are taking time off from work to support me, I am so happy to share this special day with you!

I also want to thank my external support, friends I met from conferences and platforms, Carlos, Douglas, and Susan. It's always great to talk to people from other sides of the world about research and get suggestions and opinions from different perspectives.

I would like to say thank you to my friends and family for helping me unwind and recharge, find balance away from work.

To my parents, thank you for making me who I am today. Thank you for believing in education and lifelong learning. It is also very encouraging and inspiring to see both of you pursuing your dreams and goals.

I save my final thanks to the man who is giving me the most support during the journey. To my boyfriend Yi Hengqian, you've seen me in my best and my worst, and you are always here with me. I am very lucky to have you in my life!

About the Author

Qiu Yunran (Grace) was born in Nanchang, Jiangxi province, and grew up in Beijing, China. She began her higher education journey with a curiosity about business operations and finance, starting with a grounded understanding of accounting as the foundation of observing business operations.

After completing the Bachelor degree, during her internship, she realised the importance of people, the capabilities, human resources and communication alongside financial matters, which motivated her to pursued a Master's degree in Human Resources Management in the United Kingdom, delving into person-organisation fit, capabities and company culture.

Transitioning into industry, she specialized in employee engagement and cross cultural communication. During her work, by observing the business in real world she realised the core of the business is innovation, the product or service as solution for the customers. And innovation goes hand in hand with the development of new technologies.

Motivated by the belief that innovation is critical to companies and to the economic growth and is driven by new technologies, she decided to learn more about technology driven innovations. This motivated her to start her Ph.D. journey at Delft University of Technology, focusing her research on understanding the mechanism behind using big data in digital product innovation.

Looking forward, Grace will continue her contribution to the field of innovation, leveraging her academic insights and practical experience to drive meaningful progress in product and service innovation.

&

&

ISBN 978-94-6384-562-5