DEVELOPING AN ARCHITECTURAL FRAMEWORK FOR FACILITATING TRANSFORMATION TOWARDS DATA-DRIVEN ORGANIZATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

KEREM KAYABAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

February 2022

Approval of the thesis:

DEVELOPING AN ARCHITECTURAL FRAMEWORK FOR FACILITATING TRANSFORMATION TOWARDS DATA-DRIVEN ORGANIZATIONS

Submitted by Kerem Kayabay in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics, METU** ________________

Prof. Dr. Sevgi Özkan Yıldırım
Head of Department, **Information Systems, METU** ________________

Assoc. Prof. Dr. P. Erhan Eren
Supervisor, **Information Systems, METU** ________________

Assoc. Prof. Dr. Ebru Gökalp
Co-Supervisor, **Computer Engineering Dept., Hacettepe University** ________________

Examining Committee Members:

Asst. Prof. Dr. Murat Perit Çakır
Cognitive Science, METU ________________

Assoc. Prof. Dr. P. Erhan Eren
Information Systems, METU ________________

Prof. Dr. Türksel Kaya Bensghir
Department of Business, Ankara Hacı Bayram Veli University ________________

Assoc. Prof. Dr. Oumout Chouseinoglou
Information Systems and Technologies, Bilkent University ________________

Assoc. Prof. Dr. Altan Koçyiğit
Information Systems, METU ________________

Date: 11.02.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Kerem Kayabay

Signature : __________________

ABSTRACT

DEVELOPING AN ARCHITECTURAL FRAMEWORK FOR FACILITATING TRANSFORMATION TOWARDS DATA-DRIVEN ORGANIZATIONS

Kayabay, Kerem

Ph.D., Department of Information Systems

Supervisor: Assoc. Prof. Dr. P. Erhan Eren

Co-Supervisor: Assoc. Prof. Dr. Ebru Gökalp

February 2022, 143 pages

Thesis Abstract

Paradigm shifts such as digital transformation and Industry 4.0 produce complex data, also called big data. Businesses increasingly focus on exploiting big data for competitive advantage, leveraging data science. However, many industries cannot effectively leverage data science since no comprehensive approach allows strategic planning for organization-wide data science projects and data assets. After recognizing the industry`s need, this thesis explores the Data Science Roadmapping Framework`s (DSR) development to help businesses align their business strategy with data-related, technological, and organizational resources. First, it utilizes a systematic approach to identify factors related to data science usage in organizations and challenges that data-driven transforming organizations face. In the proposed DSR framework, the resulting knowledge is synthesized with well-established technology roadmapping (TRM) literature, customizing TRM according to context, architecture, and process. Lastly, this study adopts the action research design to validate and refine the proposed framework in multiple iterations. The results indicate that the framework can help businesses initiate data science roadmapping initiatives, taking a step towards becoming data-driven. The DSR initiative also facilitates communication among business functions and generates consensus between stakeholders, including data owners, domain experts, and IT experts. While contemporary studies in the literature illustrate prebuilt roadmaps to help businesses get data-driven, this study focuses on the process of roadmapping to generate a tailored roadmap, providing the benefits above.

Keywords: technology roadmapping, data science, digital transformation, data-driven organization, architectural framework

ÖZ

VERİ GÜDÜMLÜ KURULUŞLARA DÖNÜŞÜMÜ KOLAYLAŞTIRMAK İÇİN MİMARİ BİR ÇERÇEVENİN GELİŞTİRİLMESİ

Kayabay, Kerem

Doktora, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Doç. Dr. P. Erhan Eren

Tez Eş Danışmanı: Doç. Dr. Ebru Gökalp

Şubat 2022, 143 sayfa

Tez özeti

Dijital dönüşüm ve Endüstri 4.0 gibi paradigma değişiklikleri, büyük veri olarak da adlandırılan karmaşık veriler üretmektedir. İşletmeler ise, veri biliminden yararlanarak rekabet avantajı için büyük veriden yararlanmaya giderek daha fazla odaklanmaktadırlar. Bununla birlikte, kuruluş çapında veri bilimi projeleri ve veri varlıkları için stratejik planlamaya izin veren kapsamlı bir yaklaşım bulunmadığı için, birçok endüstri veri biliminden etkili bir şekilde yararlanamamaktadır. Endüstrinin ihtiyacını anladıktan sonra, bu tez, işletmelerin iş stratejilerini veriyle ilgili, teknolojik ve organizasyonla ilgili kaynaklarıyla hizalamasına yardımcı olmak için Veri Bilimi Yol Haritası Oluşturma (DSR) çerçevesinin gelişimini araştırmaktadır. İlk olarak, kuruluşlarda veri bilimi kullanımıyla ilgili faktörleri ve veri güdümlü dönüşüm geçiren kuruluşların karşılaştıkları zorlukları belirlemek için sistematik bir yaklaşım kullanılmıştır. Elde edilen bilgiler, teknoloji yol haritası oluşturma (TRM) literatürü ile sentezlenmiş ve TRM yaklaşımı bağlama, mimariye ve sürece göre özelleştirilerek DSR çerçevesi geliştirilmiştir. Son olarak, bu çalışma, ortaya çıkan çerçeveyi ardışık yinelemelerde doğrulamak ve iyileştirmek için eylem araştırması tasarımını benimsemiştir. Sonuçlar, önerilen çerçevenin işletmelerin veri bilimi yol haritası oluşturma girişimlerini başlatmasına ve veri güdümlü olma yolunda bir adım atmasına yardımcı olabileceğini göstermektedir. Önerilen yaklaşım ayrıca iş fonksiyonları arasındaki iletişimi kolaylaştırmaktadır ve veri sahipleri, alan uzmanları ve bilişim uzmanları dahil olmak üzere paydaşlar arasında fikir birliği oluşturmaktadır. Literatürdeki çağdaş çalışmalar, işletmelerin veri güdümlü olmasına yardımcı olmak için önceden oluşturulmuş yol haritalarını gösterirken, bu çalışma, yukarıdaki faydaları sağlamak için işletmeye özel yol haritası oluşturma sürecine odaklanmaktadır.

Anahtar Sözcükler: teknoloji yol haritası geliştirme, veri bilimi, dijital dönüşüm, veri güdümlü organizasyon, mimari çerçeve

To my family

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Assoc. Prof. Dr. P. Erhan Eren for his support, guidance, and patience throughout my research. Together, we brainstormed and tested many ideas, developed state-of-the-art courses, and provisioned on-premise resources for research and education. He provided knowledge, experience, and many insights throughout our lengthy and enthusiastic discussions. I would like to thank my co-advisor, Assoc. Prof. Dr. Ebru Gökalp for sharing her experience, knowledge, and patience. She guided three out of four breaking points of the study, and I believe many of this study's accomplishments would not be possible without her endless encouragement and support. I would like to thank Assoc. Prof. Dr. Altan Koçyiğit for his help, suggestions, comments, and sharing his priceless experience and knowledge from the very beginning of my Ph.D. study. I am also grateful to Assist. Prof. Dr. Murat Perit Çakır for his valuable feedback and continuous support.

I would like to thank all my colleagues and friends from METU Informatics Institute, starting with Dr. Mert Onuralp Gökalp. I want to thank all my dear friends from there: Selin Gökalp, Umut Şener, Yücelen Bahadır Yandık, Ece Işık-Görkem Polat, Özgün Ozan Kılıç, Kerem Nazlıel, Alp Bayar, Arif Ozan Kızıldağ, Serhat Murat Alagöz, Tayfun Eylen, Utku Civelek, Ahmet Kuzubaşlı, Emrah Akkoyun, and Efe Can Yılmaz. I would like to thank Dr. Aykut Eken, who guided me to METU Informatics Institute in the first place. I am grateful to all academic and administrative staff, including Sibel Gülnar, Nilüfer Şeker, Hakan Güler, Ayşe Nur Özdere Yüksel, Sibel Ergin, Yaşar Sayın, Çetin İnci, Ali Kantar, and Eda Arbas. Hakan Güler: If you weren't there at the right moment, I would not apply to this programme.

Throughout my Ph.D. study, besides METU Informatics Institute, I had the chance to work at the University of Cambridge Institute for Manufacturing (IfM) and TÜBİTAK ULAKBİM, TRUBA High-Performace Computing Center. I would like to thank Dr. Mohamed Zaki and Professor Andy Neely for accepting me as a visiting researcher to IfM and the Cambridge Service Alliance. I would like to thank the excellent and devoted team at the TRUBA HPC Center, Onur Temizsoylu, Feyza Eryol, Sefa Arslan, Dr. Hakan Bayındır, Dr. Selçuk Bilmiş, Dr. Burcu Ortakaya, Dr. Sevil Sarıkurt, Dr. Merve Demirtaş, Dr. Özlem Sarı, and Dr. Sezen Bostan. Thank you for trusting me during my Ph.D. study.

I would like to express my gratitude to my wife, Gökçe Kayabay, for her invaluable support and patience. Thank you for loving me and believing in me. Lucky to have you in my life. I would like to thank my family, who always loved, supported, and trusted me:

My mother and father, Neslihan-Taner Kayabay, and my grandparents, Nebahat (Olcay)-Sami Koçtaş. I also want to thank my grandparents, Müjgan-İbrahim Kayabay and Kızhanım (Safiye) Sadık for loving me. Rest in peace. I shall mention Olcay Koçtaş twice, from whom I learned the importance of kindness, leadership, generosity, and modesty. I hope I am strong enough to carry on these values for the rest of my life and act accordingly. Special thanks to Asuman-Nevzat and Can Öztangut, who always supported me, loved me, and shared their experiences in the multiple critical junctures of my life. I want to thank Gülay-Habib and Pınar Doğar and also Aysun Tonka for helping Gökçe and me with love and amazing food. I shall not forget Leyla Şengezer and Hüseyin Haluk Şengezer, who always loved me and prayed for me in my most challenging moments. I express my gratitude to Birşen-Şeref Öztangut, Nermin Öztangut, and Batuhan Uzel, for always trusting me. Rest in peace, Birşen. Last but not least, I want to thank Emel-Bedri Erim for helping me throughout my decisions. I am privileged to have all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AHP	Analytic Hierarchy Process
BI&A	Business Intelligence and Analytics
BPMN	Business Process Model and Notation
CEO	Chief Executive Officer
COO	Chief Operating Officer
CRISP-DM	Cross-Industry Standard Process for Data Mining
DELTA	Data-Enterprise-Leadership-Targets-Analytics
DIPPS	Discover, Innovate, Predict, Perform, and Sustain
DOTS	Data-Organization-Technology-Strategy
DSR	Data Science Roadmapping
HR	Human Resources
HPC	High-Performance Computing
HPDA	High-Performance Data Analytics
IoT	Internet of Things
IIoT	Industrial Internet of Things
IT	Information Technology
KDD	Knowledge Discovery in Databases
KPI	Key Performance Indicator
OSEMN	Obtain, Scrub, Explore, Model, and iNterpret
RQ	Research Question
SEMMA	Sample, Explore, Modify, Model, and Assess
SLR	Systematic Literature Review
TDSP	Team Data Science Process
TRM	Technology Roadmapping

CHAPTER 1

INTRODUCTION

1.1. Research Background

Paradigm shifts are reshaping the competition across all industries. Digital transformation has changed the competition rules, creating a world where geographical advantages do not matter, protective regulations are questionable, and proprietary technologies are easy to copy (Davenport et al., 2017). With the rise of digital transformation technologies, most of the data available to any business today is unstructured and generated in large quantities without planning (Rogers, 2016). Such technologies, including the Internet of Things (IoT) and social media, have provided companies with complex data. Similarly, Industry 4.0 is a paradigm shift expected to change the production models. In this concept, heterogeneous IoT devices produce unstandardized and unstructured high-speed data whose value diminishes over time (Gokalp et al., 2016). The ability to collect arbitrary data from many divergent sources led to the transition to big data (Mayer-Schönberger & Cukier, 2014).

Numerous distributed models and architectures, notably supported by web-based companies, have emerged to deal with the different characteristics of big data that make storage and processing challenging (Gökalp et al., 2019). In return, businesses have leveraged data science using distributed technologies, seeking a competitive advantage in the market. Data science spans many fields, including mathematics, statistics, computer science, and provides the fundamental principles for extracting knowledge from data (Provost & Fawcett, 2013). Academia and industry have created process models such as Knowledge Discovery in Databases (KDD) (Fayyad et al., 1996) and Cross-Industry Standard Process for Data Mining (CRISP-DM) (Wirth & Hipp, 2000) to help organizations create business value with data science. While the process models steer individual data science projects, organizations carry out multiple data science projects aligned with their strategies to produce cross-functional value from big data.

1.2. Problem Statement

Industries that are well-suited for big data have become overachievers with data (e.g., e-commerce and online advertising). However, others have dragged behind, becoming either data-disadvantaged or underachievers with data (e.g., manufacturing, healthcare, B2B firms) (Davenport, 2014). Many businesses in the latter industries are aware of the potential, yet they face challenges while employing data-driven approaches to create business value. As a result, many organizations that pilot experiments with data analysis systems cannot deploy these systems in production (Benbya et al., 2020). For this purpose, businesses need a comprehensive approach to strategically plan for organization-wide data science endeavors and data assets. Such planning should align data science initiatives with business strategies and market opportunities. Data-related, technological, and organizational resources should also support organization-wide data science activities.

The literature acknowledges that reaching analytics goals depends on strategy, skills, culture, and leadership, but studies mainly offer prebuilt roadmaps to businesses striving to become data-driven (Dutta & Bose, 2015; Mousannif et al., 2016; O'Donovan et al., 2016; Rane & Mishra, 2018). A roadmap captures the strategic landscape and links an organization's commercial and technological functions (Phaal et al., 2004a). Nevertheless, the value lies in developing the roadmap rather than the roadmap itself (Willyard & McClees, 1987). The roadmapping (i.e., roadmap development) process facilitates communication and builds a consensus between different business functions. It is a flexible and scalable approach that requires customization to suit a particular purpose and organizational context (Phaal et al., 2013). This customization requirement makes it harder for organizations to launch roadmapping initiatives. Numerous studies have customized Technology Roadmapping (TRM) to help organizations with digital transformation (Al-Ali & Phaal, 2019), hardware development (Pearson et al., 2020), patent development (Jeong & Yoon, 2015), and smart cities (J. H. Lee et al., 2013). Although data is a fundamental asset for businesses to stay competitive, the literature lacks capturing data as an orthogonal dimension in roadmap architectures. Han and Geum (2020) recently offered data-integrated roadmaps that introduced a data layer for planning smart service systems. However, the introduced data layers did not cover the data lifecycle, and data-related processes were overlooked in the respective technology layers (Han & Geum, 2020). Indeed, adding a data layer to roadmapping architecture requires capturing and connecting organizational, technological, and strategic perspectives. Moreover, incorporating forecasting in these angles, considering the data lifecycle, catalyzes data-driven organizational transformation. The roadmapping process also needs to support the social change towards a data-driven culture of data literacy, openness, trust, experimentation, and continuous learning (Anderson, 2015).

1.3. Research Aim and Objectives

This research aims to bridge multiple research gaps by developing a Data Science Roadmapping (DSR) framework to facilitate the data-driven organizational transformation in data-disadvantaged or underachiever organizations. To strategically plan for organization-wide data science projects and data assets and overcome a set of challenges, we need to bridge several research gaps. While numerous studies focus on the challenges that data-driven transforming organizations face, mainly from the big data perspective (Section 2.4), the studies do not comprehensively identify the challenges from the data science perspective. The prebuilt roadmaps in the literature do not provide a roadmap development (roadmapping) process that enables communication and consensus between business functions. On the other hand, existing roadmapping frameworks do not connect strategic, technological, and organizational perspectives and also incorporate forecasting into these angles considering the data lifecycle. This study has the following objectives to bridge the gaps in the literature:

1. To systematically identify and categorize the challenges that data-driven transforming organizations face from the data science perspective.

2. To develop a roadmapping framework that synthesizes TRM, data science, big data, and data-driven organization literature.

3. To apply the framework to validate the framework's applicability and usefulness.

1.4. Significance of the Study

The significant contributions of this study are: (1) We identify the challenges organizations face amassing over 2,100 pieces of literature and categorizing them in the Data-Organization-Technology-Strategy (DOTS) research framework. (2) We tailor-make the widely adopted TRM according to context, architecture, and process synthesizing data science, big data, and data-driven organization literature. The resulting hybrid roadmap development methodology allows organizations to connect business strategies with data-related, technological, and organizational resources. Becoming data-driven is a long and challenging journey, requiring significant structural and social change. DSR recognizes this situation and incorporates workshop-based approaches to generate consensus among stakeholders and data-driven agile approaches to back up practitioner knowledge with quantitative evidence to overcome challenges. (3) We present detailed and extensible process models supported by specifications and templates to create an agile methodology for planning a roadmap with a segmented data layer iteratively. They show the roles, tasks, and data flow, making it easier for organizations to adopt the framework.

(4) We have refined and validated the framework by applying DSR with a research group and an oil and gas company to create data science roadmaps. Both groups identified data science trends and strategic vision for planning a data-integrated roadmap in the novel applications. Then, they planned the data, technology, and organization layers according to the identified strategies to bridge the gaps, incorporating forecasting into these angles. The demonstrations can further guide practitioners in developing their data science roadmaps. (5) Business Process Model and Notation (BPMN) (White, 2004) has been used to graphically represent the micro-level roadmapping processes during the development and application of the framework. The modeling tools have enabled the tracking of all revisions and comments in a single place, leading to systematic method development. We believe that visual process models can improve comprehension, thereby easing the execution and customization of roadmapping and serving as tools similar to the checklist-based templates for scoping roadmapping initiatives (Kerr & Phaal, 2019).

1.5. Research Strategy and Organization of the Thesis

The need for a comprehensive approach for planning organization-wide data science projects and data assets was recognized while collaborating with the industry on data science projects. A literature review helped identify the knowledge gaps in the literature. Next, a systematic literature review (SLR) helped (1) understand the factors related to the usage of data science in organizations and (2) determine the challenges that data-driven transforming organizations face. The research advanced based on the hypothesis that the widely adopted, flexible, and customizable TRM framework can also help organizations deal with data science challenges. Accordingly, TRM was customized according to context, process, and architecture. Lastly, the action research approach was adopted to refine and validate the framework in consecutive applications.

Accordingly, the thesis organization is as follows: Chapter Two presents a literature review providing the research background for data science, data-driven organizations, related studies, and technology roadmapping. Then it presents the identified knowledge gaps. Chapter Three first identifies research questions and objectives regarding knowledge gaps. It further elaborates on the research approach, depicting the study's timeline. The systematic development of the DOTS research framework presents the data science factors and challenges that answer the "what" questions in Chapter Four. In return, Chapter Five explores the development of the DSR framework in response to the "how" question. Chapter Six describes the three applications of the DSR framework that refine the framework and validate its applicability and usefulness. Lastly, Chapter Seven summarizes the overall findings, achievements, limitations, and future work.

CHAPTER 2

LITERATURE REVIEW

2.1. From BI&A to Big Data and Data Science

The IS domain generates historical data, including user records, medical images, and weblogs. Business intelligence and analytics (BI&A) technologies allow decision-makers to analyze their historical data conveniently by utilizing analytical dashboards and ad-hoc queries without dealing with the underlying technical complexities (H. Chen et al., 2012). Recently, the advancements in IoT, smart devices, and social media have enabled data collection from disparate sources. The ability to collect voluminous and continuous data from various data sources led to the transition from traditional to big data (Mayer-Schönberger & Cukier, 2014). Businesses seek competitive advantage by extracting knowledge from big data, but factors such as heterogeneous nature, volume, flow speed, and uncertainty complicate knowledge extraction.

Many distributed models and architectures have emerged to handle the characteristics of the exponentially increasing amount of data. Companies that mainly operate on the Internet, such as Google, Twitter, and LinkedIn, largely support these platforms. They develop and utilize distributed platforms to increase their ability to store, process, and manage data from different heterogeneous sources (Gökalp et al., 2019). Furthermore, to expedite their development processes, these platforms have been open-sourced. Hence, open-source platforms have become the standard big data processing platforms (Davenport & Dyché, 2013).

The availability of scalable big data technologies has given businesses opportunities to use big data to gain a competitive advantage. Businesses have started leveraging data science, which provides the fundamental principles that guide extracting knowledge from data (Provost & Fawcett, 2013). It spans many fields, including mathematics, statistics, and computer science. The concept recently attracted attention by both industry and academia with the availability of big data and distributed big data technologies. Related concepts such as machine learning, artificial intelligence (AI), and deep learning also gained their deserved attention.

2.2. Process Models for Data Science Projects

To establish standards, both academia and the industry have created process models to carry out data science and related data analytics and data mining projects. Among these process models, the most successful methodology emerged from academia (Moreira et al., 2018) is the KDD (Fayyad et al., 1996). The KDD process proposes a sequence of nine steps with possible backtracks. The process starts with learning the application domain and ends with using discovered knowledge in practice (Moreira et al., 2018). On the other hand, the CRISP-DM (Wirth & Hipp, 2000) is a highly adopted process model developed and used by the industry (Moreira et al., 2018). It proposes six sequential phases. The outcome of each phase determines the next step. Hence, like KDD, backtracks and iterations are common. The six phases are (1) business understanding, (2) data understanding, (3) data preparation, (4) modeling, (5) evaluation, and (6) deployment. Loops can occur between business understanding and data understanding, modeling and data preparation, and evaluation and business understanding.

There are OSEMN (Mason & Wiggins, 2010), SEMMA (Azevedo & Santos, 2008), and the Team Data Science Process (TDSP) (Microsoft, 2020), among other process models. Understanding these models and adopting a suitable one is vital for building the right technological and organizational capabilities. Nevertheless, these process models essentially guide the execution of individual data science projects. To maximize the benefits, an organization must carry out multiple projects aligned with its strategies and harness the outcomes of its operations. Doing so requires an overarching data-driven decision-making focus throughout the organization. Therefore, successful adoption and diffusion of data-driven approaches also require organizational transformation (Mikalef, van de Wetering, et al., 2018). The following section distinguishes between traditional and data-driven organizations and provides background on data-driven organizational transformation.

2.3. Data-Driven Organization and Data-Driven Organizational Transformation

An organization is a stable, formal, and social structure that transforms resources into products and services. The information systems perspective also considers an organization a hierarchical structure with unique culture, rules, procedures, and routines (Laudon & Laudon, 2019). Informal arrangements occur among superiors, colleagues, and subordinates, and people develop attachments to their relationships. On the other hand, data-driven organizations base business decisions on insights extracted from data rather than intuitions, superiority, and arrangements. Hence, descriptive analytics, diagnostics, predictive analytics, and prescriptive analytics are the primary tools for deriving actionable insights from many unstructured raw data and realizing its business value

(Sivarajah et al., 2017). Moreover, being data-driven is not about leveraging data to understand the past with BI&A. It is a matter of understanding the organization's future. It requires embracing recent data-driven approaches to reveal and tackle business problems even before they exist (McAfee & Brynjolfsson, 2012).

There are four types of organizational change enabled by information technology: (1) automation, (2) rationalization, (3) redesign, and (4) paradigm shifts (Laudon & Laudon, 2019). Paradigm shifts involve the highest complexity and risk but yield the highest return. Today, digital transformation is the trending paradigm shift as it reshapes many industries disrupting businesses. This phenomenon affects entire business models, segments, and functions. Companies unable to keep pace in this new world will face difficulties in the marketplace. One primary domain of digital transformation is data (Rogers, 2016). Traditionally, companies produced data with careful planning, mainly for evaluating, forecasting, and decision-making. With the rise of big data, most of the data available to any business today is unstructured and generated in large quantities without planning. Utilizing this data as a strategic asset is how a business differentiates itself in the market.

Creating cross-functional and strategic business value from big data can trigger organizational transformation. Mikalef et al. (2018) point out that adopting and diffusing big data analytics necessitate organizational transformation. Accordingly, their research question contains the term "big data-driven organizational transformation." Baesens et al. (2016) indicate that business transformation can occur due to insightful adoption and innovative applications of data science. Lastly, Wang et al. (2018) introduce a big data analytics-enabled transformation model and explain how big data analytics capabilities transform organizational practices.

2.4. Challenges for Data-Driven Transforming Organizations

Numerous studies focus on the challenges that data-driven transforming organizations face, primarily from the big data perspective. Davenport et al. (2010) introduce the Data-Enterprise-Leadership-Targets-Analytics (DELTA) model in Analytics at Work. In this concept, D stands for data, E for enterprise, L for leadership, T for targets, and A for analysts. Academic studies such as (Lismont et al., 2017) and (Wahdain et al., 2019) use the DELTA model to describe the analytics success factors. Katal et al. (2013) lay out various challenges and issues while adopting big data technologies. These challenges are (1) privacy and security, (2) data access and sharing of information, (3) storage and processing issues, (4) analytical challenges, (5) skill requirements, and (6) technical challenges. Fosso Wamba et al. (2015) present the issues related to big data-enabled business value based on a systematic review and a longitudinal case study. The issues are

(1) data policies, (2) technology and techniques, (3) organizational change and talent, (4) access to data, and (5) industry structure.

Kim and Park (2016) use the Analytic Hierarchy Process (AHP) method to investigate factors for big data usage in healthcare in Korea. They propose a research framework where the significant domains are data, organization, technology, and support. Sivarajah et al. (2017) investigate big data challenges confronted by organizations. The study presents a conceptual classification of data, process, and management challenges. Surbakti et al. (2019) use a systematic research method to examine big data usage factors and provide a research framework. They report the findings under (1) organizational aspects, (2) systems, tools, and technologies, and (3) people aspects themes. Lastly, Svensson and Taghavianfar (2020) reveal challenges organizations face in implementing a data-driven business in practice by conducting semi-structured interviews. They report the challenges in four categories: data, organization, decision-making, and management.

2.5. Prebuilt Roadmaps in Literature

Academia has produced prebuilt roadmaps to help organizations overcome data science challenges and fulfill data analytics goals. For example, Dutta and Bose (2015) develop a framework that provides organizations with a roadmap for conceptualizing, planning, and implementing big data projects. The framework comprises three phases: strategic groundwork, data analytics, and implementation. They conclude that, besides other factors, a big data project's success is mainly determined by a well-planned roadmap, active involvement of top management, and data-driven decision-making culture.

Mousannif et al. (2016) suggest a roadmap consisting of three phases: global strategy elaboration, implementation, and post-implementation. The implementation and post-implementation phases cover the lifecycle of a big data project, while the global strategy elaboration phase seeks answers to the following questions:

- Why a big data project?
- What data to use?
- Where to store and process data?
- How to protect data?

O'Donovan et al. (2016) offer a three-phase methodology to develop industrial data analytics capabilities. These phases are: (1) classifying teams and roles, (2) implementing the Information Technology (IT) architecture, and (3) applying the analytics process. Rane and Mishra (2018) follow a staged approach to introducing a roadmap for business analytics implementation in organizations. The stages are Discover, Innovate, Predict, Perform, and Sustain (DIPPS). Through case studies, they found that deploying DIPPS increased the probability of achieving analytics goals.

The academic literature acknowledges that reaching analytics goals depends on strategy, skills, partnership, culture, and leadership. However, as these studies focus on offering prebuilt roadmaps, they disregard the iterative nature of roadmaps and the benefits of carrying out roadmapping activities. To the best of our knowledge, there is no roadmapping methodology that aims to align different perspectives, such as strategy, data, technology, and organization, while fostering communication and consensus throughout the organization. Organizations benefit the most from the process of building a roadmap rather than the roadmap itself (Phaal & Muller, 2009). They realize full benefits while devising their roadmap and customizing the roadmapping process according to their context. The following section elaborates on the well-established TRM framework (Phaal et al., 2004a).

2.6. Technology Roadmapping

The technology roadmap is a time-based chart comprising the market, product, and technology layers (Phaal et al., 2004a). The middle layer of the roadmap, typically the products or services, links the firm's commercial and technological functions. As explained in the earliest published study in this domain (Willyard & McClees, 1987), Motorola used the product technology roadmap to provide managers with a comprehensive evaluation of technologies and a long-range view of future product needs. Willyard and McClees (1987) find that the value lies in developing the roadmap rather than the finished document as with any managerial tool. The roadmapping process enables the communication between business functions to eliminate misalignments, create a shared understanding, and build consensus. This is a flexible approach. Roadmaps take other forms to help with various situations, including business reconfiguration, process development, research network development, and sector foresight (Phaal et al., 2004b). Roadmapping is a scalable approach. It applies to firms, supplier networks, and entire industries, taking multiple forms (Phaal et al., 2013).

There are several methods for developing a roadmap: expert-based, computer-based, and hybrid (Pora et al., 2020). Expert-based approaches rely on qualitative analysis, such as brainstorming and scenario analysis. For example, workshop-based fast-start methods enable rapid initiation of roadmapping, bringing together diverse participants to focus on immediate issues and quick wins (Phaal et al., 2013). The original T-Plan (Phaal et al., 2004a) brings together medium-sized groups in four half-day workshops focusing on product-technology roadmapping. In comparison, the S-Plan (Phaal et al., 2007) assembles large groups in a one-or two-day workshop to explore and prioritize strategic issues. Computer-based methods are built upon quantitative data analysis that utilizes large textual databases and other novel data sources (Pora et al., 2020). These approaches are useful when it is challenging to articulate practitioner knowledge in roadmapping workshops (Park et al., 2020). The best of both worlds is possible with a hybrid approach that can leverage quantitative evidence to support practitioner knowledge in roadmapping workshops (Park et al., 2020). Nevertheless, a universal roadmapping approach that fits all contexts does not exist (Pearson et al., 2020), so adaptations are indispensable.

Despite its versatility, a one-size-fits-all approach is not appropriate for roadmapping. A successful application requires adaptation of the roadmap structure and roadmapping process to suit an organization's particular purpose and context (Pearson et al., 2020; Phaal et al., 2013). Numerous studies in the literature acknowledge the importance of customization, and they propose customization approaches to guide organizations to start roadmapping initiatives. Kerr et al. (2019) customized the S-Plan reference process to a specific organizational context by introducing pre-workshop and post-workshop steps. Al-Ali and Phaal (2019) introduced an agile roadmapping framework customized for digital transformation by integrating prototyping into the process. Souza et al. (2020) developed an agile roadmapping approach for digital start-ups, replacing workshops with small meetings with the help of digital software. Pearson et al. (2020) customized the methodology to guide agile and lean innovation and manage uncertainty and risk at a fusion start-up. The roadmap development process involved successive iterations to the roadmap every six months following the production of the first-pass roadmap aligned with the build-measure-learn cycles (Ries, 2011).

Data is a fundamental asset for businesses to stay competitive; however, the literature lacks studies that consider data as a distinct perspective in TRM. Han and Geum (2020) introduce roadmaps that add a data layer as a functional link for planning smart service systems and illustrate data-integrated roadmap development based on ex-post analysis of case studies. However, their roadmap development process does not identify the required technologies according to data-related processes. Additionally, the roadmap architectures, particularly the data layers, do not cover the data lifecycle.

2.7. Literature Summary and Knowledge Gaps

The shift from traditional data to big unplanned data has created business opportunities. Businesses have leveraged data science to extract knowledge from data, and they have adopted well-established process models while doing so. While adopting a suitable process model helps build the right capabilities, diffusion of data-driven approaches requires organizational transformation. The literature investigates challenges that data-driven transforming organizations face, mainly from the big data usage perspective. On the other hand, data science spans multiple fields and leverages scalable big data technologies to tackle business problems.

Academia offers prebuilt roadmaps to help businesses in their data-driven transformation journeys. In a nutshell, a roadmap is a time-based visual tool that provides managers comprehensive evaluation of technologies. As with any managerial tool, the businesses achieve real value while developing the roadmap rather than the roadmap itself. The process is called roadmapping, and it eliminates information asymmetries and develops a shared understanding. Roadmapping needs customization to suit a particular purpose and context. Despite data being a fundamental asset for businesses to stay competitive, the roadmapping literature lacks studies to plan organization-wide data assets and data science projects. Consequently, this study identifies the following knowledge gaps in the literature:

1. The literature lacks a study that comprehensively identifies challenges that data-driven transforming organizations face from the data science perspective.

2. No roadmapping framework enables organizations to plan for organization-wide data assets and data science projects.

3. There is no application of roadmapping that considers data in a separate layer linking different functions.

CHAPTER 3

RESEARCH APPROACH

3.1. Research Questions and Objectives

This study aims to facilitate data-driven organizational transformations, acknowledging the need from data-disadvantaged and underachiever type organizations. Three research questions identifying the scope of this thesis are:

1. What factors relate to the usage of data science in organizations?

2. What challenges do organizations face on their journey to become data-driven considering the data science perspective?

3. How should an organization build a holistic roadmap that blends strategical, data-related, technological, and organizational perspectives to facilitate its data-driven organizational transformation?

Accordingly, this study has the following objectives considering the research questions together with the knowledge gaps in the literature:

1. To systematically identify and categorize the challenges that data-driven transforming organizations face from the data science perspective.

2. To develop a roadmapping framework to help organizations overcome data science challenges and become data-driven.

3. To apply the framework in organizational settings to validate the framework's applicability and usefulness.

3.2. Research Approach

This study began by recognizing the need for a comprehensive approach for planning organization-wide data science projects and data assets while collaborating with industry experts on data science projects. An SLR was conducted to determine the data science challenges organizations face while using data-driven approaches to create business value and confirm the need for a comprehensive approach. The research progressed based on the hypothesis that the widely adopted, flexible, and customizable TRM framework (Phaal et al., 2004a) could help organizations deal with data science challenges since numerous studies successfully applied the framework in various organizational contexts for different purposes.

TRM is a flexible approach; however, a successful application requires tailoring to a particular situation (Phaal et al., 2013). The roadmap architecture and the roadmapping process must provide a framework for structuring knowledge. Some studies guide and facilitate customization activity (S. Lee & Park, 2005; Phaal et al., 2004b). However, this requirement makes it challenging for organizations to start such initiatives because they lack expertise in roadmapping. We follow the guidelines recommended by Phaal et al. (2004a) to tailor-make the roadmap architecture and the standard T-Plan according to context, architecture, and process, leveraging CRISP-DM and DOTS research framework.

The resulting framework provides a complete methodology comprising the macro-and micro-level process models developed iteratively. Several experts who carried out interdisciplinary work in both industry and academia provided feedback during the development process. These experts work in data-disadvantaged or underachiever type organizations and match the roles in Table 23. They are aware of the need for organizational transformation. They have data science sector experience, and some also have software engineering experience. During the development process, the primary researcher's interaction with experts included guiding modifications, keeping logs, and facilitating workshops during the application. After the methodology was sufficiently mature, we adopted an action research approach to refine the framework and validate DSR's applicability and usefulness. Action research design is suitable when the research question is about understanding the process of change or improvement to learn from it (Coughlan & Coghlan, 2002). The underlying beliefs in action research designs are (Easterby-Smith et al., 2015): (1) The best way to learn about an organization is by attempting to change it, and (2) the people implementing these changes should become part of the research process itself. Accordingly, action research is appropriate for developing practical tools such as roadmapping that require working together on "live" management problems and challenges (Kerr et al., 2019).

We applied the DSR methodology with a cross-disciplinary research group with sector experience and an oil and gas company to plan their future data science initiatives and their respective resources. Figure 1 depicts the research approach adopted while developing and iterating on DSR using a timeline. All meetings were recorded for evaluation purposes. We also kept journals capturing information and experiences. In both iterations, we asked the participants questions Kerr and Phaal (2019) suggested and collected unstructured qualitative data about their roadmapping experience. We report the results and discuss the implications for research and practice in the respective sections.

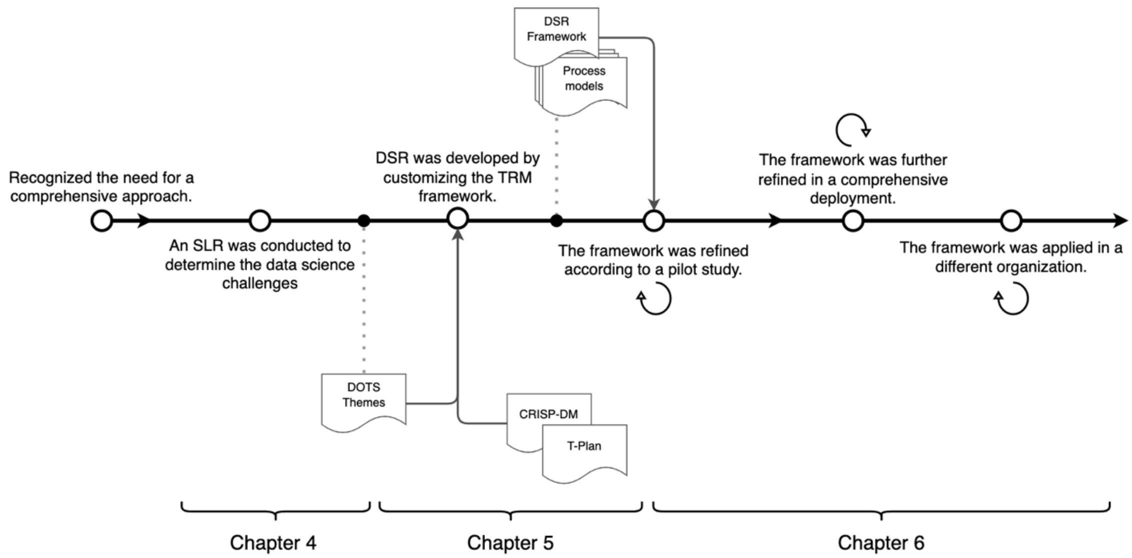

Figure 1: Research approach adopted in this thesis

CHAPTER 4

DOTS RESEARCH FRAMEWORK

4.1. The Need for a Systematic Approach

As paradigm shifts, including digital transformation and Industry 4.0, are reshaping the competition, businesses try to create cross-functional and strategic business value from unplanned data (Rogers, 2016). However, they face numerous obstacles, leveraging data-driven approaches. Numerous studies investigate these challenges considering the big data perspective (Section 2.4). On the other hand, the data science perspective is critical since data science leverages big data technologies to tackle business problems while spanning multiple fields. In an attempt to systematically identify all the challenges and provide directions for data-driven transforming organizations, this study applies systematic literature review method to answer the following research questions:

- RQ1: What are the factors identified in the literature related to the usage of data science in organizations?
- RQ2: In a broader sense, what are the challenges that organizations face on their journey to become data-driven?

4.2. SLR Methodology

4.2.1. Planning the Review

To account for all relevant prior work and present results as valuable insights to researchers and practitioners, the systematic review method recommended by Tranfield et al. (2003) is used. This section elaborates on the planning, conducting, and reporting steps of the systematic review.

This study begins with an extensive search for prior reviews on the subject topic. After the research gap is identified, in an iterative process, the review protocol is prepared, clarifying (1) the specific research questions, (2) the sample of studies, (3) the search strategy, and (4) the criteria for inclusion & exclusion of studies. The systematic search comprises Scopus, Web of Knowledge, and IEEE Xplore electronic databases. Following search strings and rules are used during this process:

1. In title, abstract, or keywords, at least one occurrence of the following keywords is demanded: (business* OR organi?ation* OR enterprise)

2. In title, abstract, or keywords, at least one occurrence of the following keywords is asked for: ("data science" OR analytics) The keyword analytics spans other keywords like "data analytics", "big data analytics", "business analytics", "business intelligence and analytics".

3. In all of the text, at least one occurrence of the following keyword is asked for: ("case stud*"). The focus is real case scenarios to increase the applicability of the study's findings.

4. Since this is a cross-disciplinary study, the subject area is limited to business, engineering, computer science, social science, and decision science.

5. The document types are limited to journal articles, conference papers, and book chapters to put a threshold on quality. Conference proceedings are excluded only for the IEEE database.

In the first pass, the inclusion and exclusion of studies are done based on the title. In the second pass, abstracts and keywords are examined. Among the criteria defined to minimize researcher bias, there are:

1. Any study that explores factors related to the usage of data science in an organization is included.

2. Any study that explores the challenges confronted by businesses trying to become data-driven organizations is included.

3. All studies that focus on business intelligence & analytics, big data, and data analytics are included.

4. Studies related to smart cities are excluded.

An extended set of criteria is applied to the remaining sample considering abstract and keywords in the second pass. The full text is also considered if a decision cannot be made based on abstract and keywords. There are additional criteria: the article must be in English, and full text must be available.

4.2.2. *Conducting the Review*

The search is done in the first quarter of 2019. At the end of stage one, 2,153 papers are found. At stage two, 557 articles are found relevant, applying the first set of criteria. At stage three, 288 articles remained, applying the second set of criteria. The remaining articles are extended through backward citations and articles that are assessed as crucial to this research. The diagram of the selection process is delineated in Figure 2.

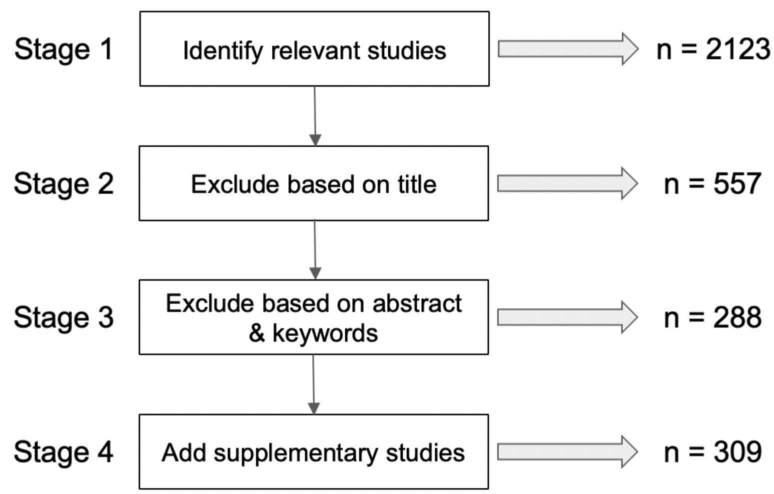

Figure 2: The study selection process [based on (Mikalef, Pappas, et al., 2018; Surbakti et al., 2019)]

4.2.3. *Reporting and Dissemination*

Synthesizing findings, factors (see RQ1), and challenges (see RQ2) are extracted from studies. In a bottom-up approach, similar factors and challenges are aggregated together. Broader categories started to appear, but it took several iterations and meetings to reach a consensus among researchers for these categories. Eventually, these categories are fit into four prominent themes, which are named "Data," "Organization," "Technology," and "Strategy." Accordingly, DOTS research framework is composed, titled after the initials of the four themes.

4.3. DOTS Research Framework

Upon aggregating factors influencing the usage of data science in organizations, four themes emerge. In the DOTS research framework, four pillars of a data-driven transformation are "Data," "Organization," "Technology," and "Strategy." Figure 3 delineates the primary dimensions of each theme. The following subsections summarize each theme's key findings, accompanied by secondary dimensions and these dimensions' references.

Figure 3: Primary dimensions of DOTS themes

4.3.1. Data Theme

The data theme accounts for the data science lifecycle and data governance. The latter attempts to overcome data quality, sensitivity, security, reliability, accessibility, availability, and usability challenges (Ardagna et al., 2016b). It emphasizes master data management (L. K. W. Fernando & Haddela, 2017), well-defined standards and guidelines (Ardagna et al., 2016b), and metadata management (L. K. W. Fernando & Haddela, 2017). Some practical challenges in different phases of a data science process (for the CRISP-DM) are as follows:

- **Business understanding**: Organizations should identify valuable use cases while identifying the data objects that business functions should store and prioritize

(Pape, 2016). Since data science involves a degree of uncertainty, exploratory analysis of available data at this stage may lead to counterintuitive findings (Alexander & Lyytinen, 2017).

- **Data understanding**: Accessing data may be a challenge if it requires other organizations' involvement. Data might be unavailable or proprietary (Golightly et al., 2018).

- **Data preparation:** Every data source in a heterogeneous system may require a dedicated data collector (Derguech et al., 2014), and the level of detail (i.e., data granularity) can vary among data sources (Gopalkrishnan et al., 2012). Fragmented data across vendors, business functions, and services raise integration challenges (Malaka & Brown, 2015).

- **Modeling**: The characteristics of data and modeling may necessitate the use of specific approaches. For example, a large-scale business analytics problem may force modelers to employ scalable and fault-tolerant mechanisms. Tool selection is a problem since many tools and application frameworks implement data analytics architecture components in a distributed system (Gökalp et al., 2019). Additional challenges include experimental tracking and code reusability (Google, 2020).

- **Evaluation**: It is challenging to examine the models' generalizability and validate their performance under different conditions since we create a model based on the available data. Furthermore, unlike software systems, systems with machine learning components are usually non-deterministic and may not reproduce the same results, making these systems challenging to test and verify (Ozkaya, 2020; Sculley et al., 2015). Moreover, only the proper representation of the data products, such as insights or recommendations to domain experts, increases the enterprise's analytics adoption (Daradkeh, 2019).

- **Deployment and maintenance**: Models in production are usually ephemeral due to the context-dependent nature of data. Delivering production models requires automated training and validation steps performed manually by data scientists. A continuous delivery mechanism that leverages best practices, such as continuous integration and testing, is necessary. However, applying DevOps principles to artificial intelligence systems does not guarantee returns since AI systems differ from software systems (Google, 2020).

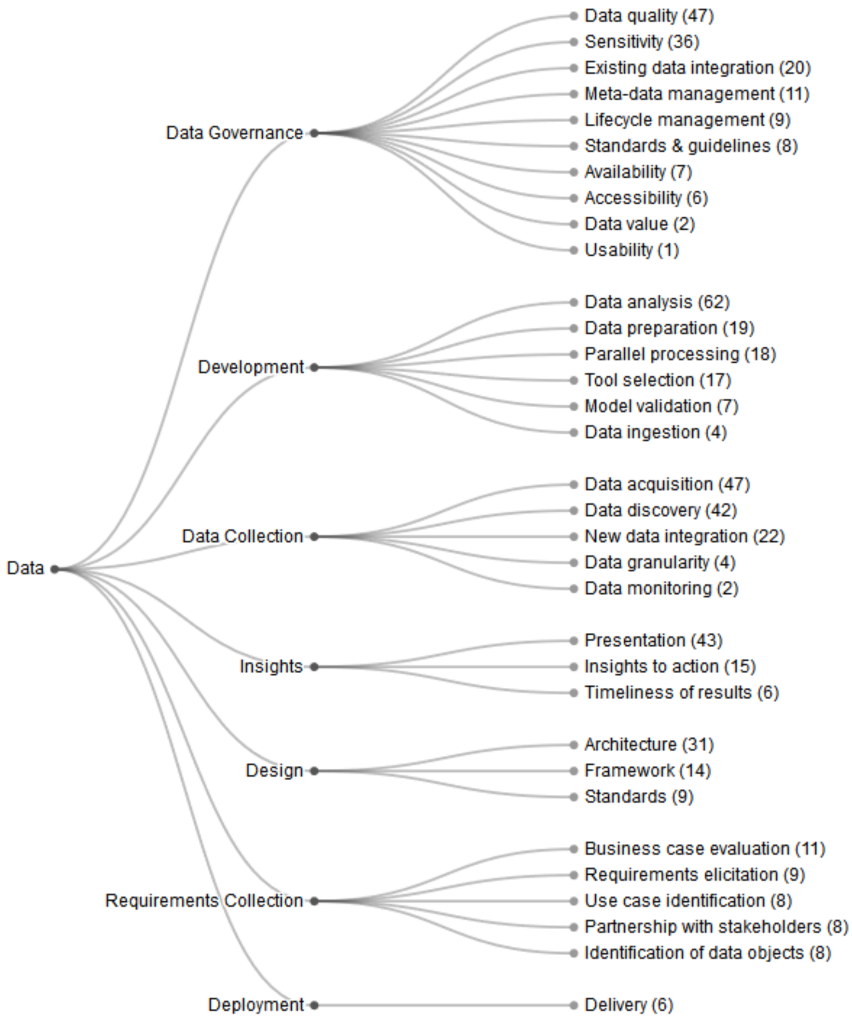

Figure 4: Secondary dimensions of data theme and number of articles that reference each secondary dimension

Table 1: The definitions and references for the primary dimension: Data governance

Secondary Dimension	**Definition and references**

Data quality	Data quality has several dimensions, including accuracy (error-free data), timeliness (up-to-date data), consistency (same format), and completeness (Hazen et al., 2014). Poor data quality decreases the reliability of results and increases time spent in preprocessing. (Mikalef, Pappas, et al., 2018), (Liu & Shi, 2015), (Batra, 2017), (Rao et al., 2018), (Kowalczyk & Buxmann, 2015a), (S. Verma et al., 2018), (Amankwah-Amoah & Adomako, 2019), (Ardagna et al., 2016b), (F. Fernando & Engel, 2018), (Koronios et al., 2014), (Muhammad Habib ur et al., 2016), (Rikhardsson & Yigitbasioglu, 2018), (Malaka & Brown, 2015), (Tucker et al., 2017), (Miller, 2018), (V. Grover et al., 2018), (Hee Yeong Kim et al., 2018a), (Kasim et al., 2012), (Pérez-González et al., 2019), (Seddon et al., 2017), (L. K. W. Fernando & Haddela, 2017), (M.-K. Kim & Park, 2016), (Dwivedi & Kulkarni, 2008), (Mounir et al., 2018), (Segooa & Kalema, 2018), (Lam et al., 2017), (Vidgen et al., 2017), (Khatri, 2016), (Kademeteme et al., 2017), (Lamba & Singh, 2018), (Ji-fan Ren et al., 2017), (Kowalczyk & Buxmann, 2015a), (Rubab et al., 2019), (Shanmuganathan, 2019), (Seddon & Constantinidis, 2012), (Hazen et al., 2017), (Lai et al., 2018), (Riungu-Kalliosaari et al., 2017), (Zahid et al., 2018), (Janssen et al., 2017), (Baars & Ereth, 2016), (Ramesh & Ramakrishna, 2018), (Lismont et al., 2017), (Baesens et al., 2016), (Li et al., 2019)
Sensitivity	Security, privacy, and sensitivity challenges are highlighted when cloud computing is used to perform parallel processing on Big Data (Chaoui & Makdoun, 2017). Additional challenges include data loss, data breach, and data theft (Ardagna et al., 2016b). (Banda & Ngassam, 2017), (Iqbal et al., 2018), (Liu & Shi, 2015), (Chong & Shi, 2015), (Y. Wang, Kung, & Byrd, 2018), (S. Verma, 2017), (Kimble & Milolidakis, 2015), (Koronios et al., 2014), (Al-Hakimi, 2017), (Gopalkrishnan et al., 2012), (Maleki et al., 2016), (Malaka & Brown, 2015), (Tucker et al., 2017), (Miller, 2018), (V. Grover et al., 2018), (Sivarajah et al., 2017), (Roderick et al., 2017), (Hee Yeong Kim et al., 2018a), (Carter & Sholler, 2015), (Kasim et al., 2012), (Günther et al., 2017), (Bihl et al., 2016), (Coleman et al., 2016), (M.-K. Kim & Park, 2016), (Lamba & Singh, 2018), (Vo et al., 2017), (Schüll & Maslan, 2018), (Fromm et al., 2012), (Angelopoulos et al., 2016), (Goben & Raszewski, 2015), (Zahid et al., 2018), (Bygstad et al., 2019), (Li et al., 2019), (Atyeh et al., 2017)
Existing data integration	Since big data is collected from heterogeneous sources, the integration of data stored in different business divisions' storage is a challenge. Data virtualization is an approach to provide encapsulated views for data stored in a heterogeneous set of data stores (Janković et al., 2018). (J. Y. Lee et al., 2017), (Batra, 2017), (Rao et al., 2018), (Phillips-Wren & Hoskisson, 2015), (Derguech et al., 2014), (Jose et al., 2017), (V. Grover et al., 2018), (Shanks & Sharma, 2011), (Hee Yeong Kim et al., 2018a), (Phillips-Wren & Hoskisson, 2014), (Coleman et al., 2016), (M.-K. Kim & Park, 2016), (Mounir et

	al., 2018), (Khatri, 2016), (Roy et al., 2014), (Aldea et al., 2018), (Bygstad et al., 2019), (Baars & Ereth, 2016), (H. M. Chen, Schutz, et al., 2016)
Meta-data management	Proper meta-data should be implemented to support data preservation and re-use (Kasim et al., 2012). (J. Y. Lee et al., 2017), (Spruit & Sacu, 2015a), (Coleman et al., 2016), (L. K. W. Fernando & Haddela, 2017), (Khatri, 2016), (Vo et al., 2017), (Chung & Chung, 2013), (Beheshti et al., 2018), (Baars & Ereth, 2016), (Ramesh & Ramakrishna, 2018)
Data lifecycle management	Data lifecycle management is the process of managing data throughout its lifecycle from collection to ingestion, processing, archiving, and disposal (Y. Wang, Kung, & Byrd, 2018). (Maleki et al., 2016), (Hee Yeong Kim et al., 2018a), (Günther et al., 2017), (L. K. W. Fernando & Haddela, 2017), (Khatri, 2016), (Goben & Raszewski, 2015), (Baars & Ereth, 2016), (Li et al., 2019)
Standards and guidelines	There are legal issues with data analytics, such as accounting for intellectual properties. Well-defined and internationally recognized standards and guidelines can help with legal and regulatory compliance (Ardagna et al., 2016b). (S. Verma, 2017), (Kimble & Milolidakis, 2015), (Maleki et al., 2016), (Hee Yeong Kim et al., 2018a), (Günther et al., 2017), (Spruit & Sacu, 2015a), (Li et al., 2019)
Availability	The extent of readily available high-quality data increases data analytics benefits (Seddon et al., 2017). (Al-Hakimi, 2017), (Hee Yeong Kim et al., 2018a), (Carter & Sholler, 2015), (Lam et al., 2017), (Rubab et al., 2019), (Zahid et al., 2018)
Accessibility	Quality of data access affects traceability and analytics quality (Kowalczyk & Buxmann, 2015a). On the other hand, while proprietary data sets owned by organizations limit what questions can be asked, privileged access to data is a strong incentive to work for an organization (Carter & Sholler, 2015). (Iqbal et al., 2018), (Chong & Shi, 2015), (Gudfinnsson & Strand, 2018), (Lavalle et al., 2011)
Data value	As data is being created through social media and the Internet of Things, companies need to evaluate their data assets and quantify their value (Enders, 2018). (Baesens et al., 2016)

Usability	Data usability may increase with a more compact description of the dataset (Ardagna et al., 2016b).

Table 2: The definitions and references for the primary dimension: Development

Secondary Dimension	**Definition and references**
Data analysis	This activity collects programming models, algorithms, and techniques (e.g., machine learning, deep learning) to process, make sense of, and gain insights from data. (Mikalef, Pappas, et al., 2018), (J. Y. Lee et al., 2017), (Dittert et al., 2018), (Lu, 2018), (Iqbal et al., 2018), (Liu & Shi, 2015), (Holsapple et al., 2014), (Derguech et al., 2014), (Arora & Malik, 2015), (Hazen et al., 2018), (Wu et al., 2016), (Choi et al., 2018), (Vashisht & Gupta, 2016), (Chong & Shi, 2015), (Y. Wang, Kung, & Byrd, 2018), (S. Verma, 2017), (F. Fernando & Engel, 2018), (Assunção et al., 2015), (Sá et al., 2015), (J. Y. Lee et al., 2017), (Gopalkrishnan et al., 2012), (Maleki et al., 2016), (Huang et al., 2009), (Bedeley et al., 2018), (Y. Wang & Byrd, 2017), (H. Chen et al., 2012), (Mishra & Saini, 2016), (Alade, 2017), (Raffoni et al., 2018), (Nalchigar & Yu, 2018), (Malaka & Brown, 2015), (Sahu et al., 2017), (Nalchigar & Yu, 2017), (Sivarajah et al., 2017), (Roderick et al., 2017), (Kasim et al., 2012), (Bihl et al., 2016), (Sangupamba Mwilu et al., 2016), (Zafeiropoulos et al., 2018), (Pérez-González et al., 2019), (Coleman et al., 2016), (Hamister et al., 2018), (Khatri, 2016), (Dutta & Bose, 2015), (Roy et al., 2014), (Newman et al., 2016), (Vo et al., 2017), (Chung & Chung, 2013), (Blake & Gabb, 2014), (Rubab et al., 2019), (Milosevic et al., 2016), (Laha, 2016), (Mowrer et al., 2017), (Goben & Raszewski, 2015), (Shanmuganathan, 2019), (Flath & Stein, 2018), (Gončarovs & Grabis, 2017), (Janssen et al., 2017), (K. Kim & Lee, 2018), (Olszak, 2016), (Bi et al., 2016), (Duke & Ashraf, 2019)
Data preparation	Before analysis, a data preparation step is critical where data preprocessing techniques improve the quality of data. These techniques include but are not limited to noise reduction, outlier detection, and feature extraction (Muhammad Habib ur et al., 2016). (J. Y. Lee et al., 2017), (Iqbal et al., 2018), (Liu & Shi, 2015), (F. L. Wang, Rischmoller, Reed, et al., 2018), (Hazen et al., 2018), (Vashisht & Gupta, 2016), (S. Verma, 2017), (Assunção et al., 2015), (Nalchigar & Yu, 2018), (Nalchigar & Yu, 2017), (Sivarajah et al., 2017), (Sun et al., 2014), (Ordonez & Garcia-Garcia, 2016), (Goben & Raszewski, 2015), (Gončarovs & Grabis, 2017), (Janssen et al., 2017), (K. Kim & Lee, 2018), (Olszak, 2016)
Parallel processing	Parallel or distributed processing provides a high degree of flexibility by having the big dataset analyzed at the same time by multiple distributed processors (Choi et al., 2018).

	(Rahman & Aldhaban, 2015), (Wu et al., 2016), (Mishra & Saini, 2016), (Brichni & Guedria, 2018a), (Kasim et al., 2012), (Zafeiropoulos et al., 2018), (Pérez-González et al., 2019), (Gupta et al., 2013), (L. K. W. Fernando & Haddela, 2017), (S. Shah et al., 2018), (Pusala et al., 2016), (Rubab et al., 2019), (Milosevic et al., 2016), (Rekha & Parvathi, 2015), (Bi et al., 2016), (H. M. Chen, Kazman, et al., 2016), (Ren et al., 2014)
Tool selection	There is an abundance of tools and application frameworks for every data analytics architecture component (Gokalp et al., 2016). Selecting the right tools and programming models helps prevent technical debt (Sculley et al., 2014). (Chong & Shi, 2015), (P. Grover & Kar, 2017), (Alade, 2017), (O'Donovan et al., 2016), (Ricardo et al., 2008), (Sahu et al., 2017), (Bihl et al., 2016), (Zafeiropoulos et al., 2018), (Aho & Uden, 2014), (Gupta et al., 2013), (Coleman et al., 2016), (Schüritz et al., 2017), (Vidgen et al., 2017), (Rubab et al., 2019), (Atyeh et al., 2017)
Model validation	Data scientists need to provide reliable and relevant results (Riungu-Kalliosaari et al., 2017). After training a model, it is essential to evaluate it before putting the model into service (Gončarovs & Grabis, 2017). (H. M. Chen, Kazman, et al., 2016), (J. Y. Lee et al., 2017), (Spruit & Sacu, 2015a), (Tao & Gao, 2016), (Baesens et al., 2016)
Data ingestion	For real-time data moving at high velocity (e.g., large-scale continuous sources such as event logs (Beheshti et al., 2018)), a fault-tolerant ingestion mechanism is necessary to get data from sources to processing. (Zahid et al., 2018), (Milosevic et al., 2016), (O'Donovan et al., 2016)

Table 3: The definitions and references for the primary dimension: Data collection

Secondary Dimension	**Definition and references**
Data acquisition	Data acquisition is a complex problem since a massive amount of structured, semi-structured, and unstructured data is produced by heterogeneous sources (Chong & Shi, 2015). For every data source, the design of a dedicated data collector may be required (Derguech et al., 2014). (J. Y. Lee et al., 2017), (Lu, 2018), (Altarturi et al., 2018), (Liu & Shi, 2015), (Vashisht & Gupta, 2016), (Y. Wang, Kung, & Byrd, 2018), (S. Verma, 2017), (Sá et al., 2015), (Muhammad Habib ur et al., 2016), (Mishra & Saini, 2016), (Maturana & Asenjo, 2015), (Sahu et al., 2017), (Debortoli et al., 2010), (Sivarajah et al., 2017), (Roderick et al., 2017), (Kasim et al., 2012), (Pérez-González et al., 2019), (Coleman et al., 2016), (Sun et al., 2014), (M.-K. Kim & Park, 2016), (Dwivedi & Kulkarni, 2008), (Mounir et al., 2018), (Brennan et al., 2018), (Khatri, 2016), (Lamba & Singh, 2018), (Chung & Chung, 2013), (Rubab

	et al., 2019), (C. K. M. Lee et al., 2015), (Watson, 2018), (Janković et al., 2018), (Goben & Raszewski, 2015), (Castellanos et al., 2017), (Tanwar et al., 2015), (Zahid et al., 2018), (Zhao et al., 2018), (Baars & Ereth, 2016), (Lavalle et al., 2011), (Hartmann et al., 2016), (Olszak, 2016), (Gear et al., 1982), (H. M. Chen, Kazman, et al., 2016), (Atyeh et al., 2017)
Data discovery	Heterogeneous sources generate big data. Organizations can discover data passively by listening to social media and available tools or actively engaging customers and employees with technological platforms (Troisi et al., 2018) and crowdsourcing techniques. Accessing the data may be a challenge – it might require the manufacturer or supplier's involvement (Golightly et al., 2018). (Liu & Shi, 2015), (Kowalczyk & Buxmann, 2015a), (Phillips-Wren & Hoskisson, 2015), (Derguech et al., 2014), (Kretzer et al., 2014a), (P. Grover & Kar, 2017), (Y. Wang, Kung, & Byrd, 2018), (Sá et al., 2015), (Jose et al., 2017), (Muhammad Habib ur et al., 2016), (Gopalkrishnan et al., 2012), (Nanjappan et al., 2017), (Maleki et al., 2016), (Edge et al., 2018), (Huang et al., 2009), (Božič & Dimovski, 2019), (H. Chen et al., 2012), (Mishra & Saini, 2016), (Nino et al., 2015), (Tucker et al., 2017), (Hee Yeong Kim et al., 2018a), (Kasim et al., 2012), (Günther et al., 2017), (Spruit & Sacu, 2015a), (Mounir et al., 2018), (Khatri, 2016), (Dutta & Bose, 2015), (N. Verma & Voida, 2016), (Vo et al., 2017), (Schüll & Maslan, 2018), (Watson, 2018), (Mowrer et al., 2017), (Goben & Raszewski, 2015), (Popovič et al., 2018), (Flath & Stein, 2018), (Riungu-Kalliosaari et al., 2017), (Bygstad et al., 2019), (Janssen et al., 2017), (Hartmann et al., 2016), (Duke & Ashraf, 2019)
New data integration	Data is found in fragmented sources across vendors, business functions, tools, and services (Malaka & Brown, 2015). Integrating new data is vital before commencing further activities. (Lu, 2018), (Liu & Shi, 2015), (Kretzer et al., 2014a), (Vashisht & Gupta, 2016), (Chong & Shi, 2015), (Assunção et al., 2015), (Muhammad Habib ur et al., 2016), (Gopalkrishnan et al., 2012), (Nanjappan et al., 2017), (Maleki et al., 2016), (Božič & Dimovski, 2019), (Nino et al., 2015), (V. Grover et al., 2018), (Sivarajah et al., 2017), (Pérez-González et al., 2019), (Coleman et al., 2016), (Brennan et al., 2018), (Chung & Chung, 2013), (Niño et al., 2016), (Zahid et al., 2018), (H. M. Chen, Kazman, et al., 2016)
Data granularity	As volume and diversity of data keep growing in some industries, the detail level also increases (Jose et al., 2017). In some cases, granularity varies among data sources (Gopalkrishnan et al., 2012). (F. L. Wang, Rischmoller, Reed, et al., 2018), (Günther et al., 2017)
Data monitoring	After the data is discovered, acquired, and integrated, a monitoring infrastructure is necessary (Huang et al., 2009). A change in the configuration in the source can make the acquisition and integration mechanism obsolete.

	(Brennan et al., 2018)

Table 4: The definitions and references for the primary dimension: Insights

Secondary Dimension	**Definition and references**
Presentation	Insights from data should be presented to domain analysts in the form of dashboards or reports (C. K. M. Lee et al., 2015). Useful visual representation of data with a friendly interface can increase the analytics adoption in the enterprise (Daradkeh, 2019). (Mikalef, Pappas, et al., 2018), (Golightly et al., 2018), (Liu & Shi, 2015), (F. L. Wang, Rischmoller, Reed, et al., 2018), (Phillips-Wren & Hoskisson, 2015), (Hazen et al., 2018), (Y. Wang, Kung, & Byrd, 2018), (S. Verma, 2017), (Assunção et al., 2015), (N. Shah et al., 2017), (Sá et al., 2015), (J. Y. Lee et al., 2017), (Gopalkrishnan et al., 2012), (Maleki et al., 2016), (Edge et al., 2018), (Y. Wang & Byrd, 2017), (Rikhardsson & Yigitbasioglu, 2018), (Mishra & Saini, 2016), (Alade, 2017), (Sahu et al., 2017), (Sivarajah et al., 2017), (Brichni & Guedria, 2018a), (Kasim et al., 2012), (Phillips-Wren & Hoskisson, 2014), (Bihl et al., 2016), (Pérez-González et al., 2019), (Coleman et al., 2016), (Lizotte-Latendresse & Beauregard, 2018), (Hamister et al., 2018), (Magee et al., 2016), (Vidgen et al., 2017), (Khatri, 2016), (Dutta & Bose, 2015), (Vo et al., 2017), (Mowrer et al., 2017), (Endert et al., 2014), (Ghose & Dam, 2014), (Smuc et al., 2008), (Ashraf & Khan, 2015), (Lavalle et al., 2011), (Olszak, 2016)
Insights to action	It is necessary to take timely action according to insights from data analytics (Mikalef, Pappas, et al., 2018). (Golightly et al., 2018), (Lu, 2018), (Liu & Shi, 2015), (Hazen et al., 2018), (Assunção et al., 2015), (Y. Wang & Byrd, 2017), (Sivarajah et al., 2017), (Pérez-González et al., 2019), (Coleman et al., 2016), (Seddon et al., 2017), (Dutta & Bose, 2015), (Lamba & Singh, 2018), (Flath & Stein, 2018), (Janssen et al., 2017)
Timeliness of results	In the big data era, some high-speed data must be processed timely as the value of insights diminish over a short period. Therefore, analytics systems should provide information promptly (S. Verma et al., 2018). (Jose et al., 2017), (Maturana & Asenjo, 2015), (Tucker et al., 2017), (Trieu, 2017), (G. Park et al., 2017)

Table 5: The definitions and references for the primary dimension: Design

Secondary Dimension	**Definition and references**
Architecture	Architecture refers to the physical and logical views of information and databases (Lavalle et al., 2011) or a blueprint representing the system's fundamental

	components and their interactions with each other (Sangupamba Mwilu et al., 2016). (J. Y. Lee et al., 2017), (Zorrilla & García-Saiz, 2013), (Kridel & Dolk, 2013), (Y. Wang, Kung, & Byrd, 2018), (Sá et al., 2015), (Muhammad Habib ur et al., 2016), (O'Donovan et al., 2016), (Maturana & Asenjo, 2015), (Ereth & Baars, 2015), (Miller, 2018), (Zafeiropoulos et al., 2018), (Glissmann et al., 2012), (Spruit & Sacu, 2015a), (Dwivedi & Kulkarni, 2008), (Khatri, 2016), (Pusala et al., 2016), (Vo et al., 2017), (C. K. M. Lee et al., 2015), (Watson, 2018), (Angelopoulos et al., 2016), (Zahid et al., 2018), (Aldea et al., 2018), (Zhao et al., 2018), (Baars & Ereth, 2016), (H. M. Chen, Kazman, et al., 2016), (H.-M. Chen, Kazman, et al., 2016)
Framework	Organizations can build or benefit from frameworks to understand best practices and techniques and reduce the required engineering effort (Zhu et al., 2016a). (Lu, 2018), (Yasin et al., 2018), (Llave, 2017), (Vera-Baquero et al., 2015), (Nalchigar & Yu, 2017), (Nalchigar & Yu, 2018), (Hee Yeong Kim et al., 2018a), (Sangupamba Mwilu et al., 2016), (Bhattacharya et al., 2009), (L. K. W. Fernando & Haddela, 2017), (Pape, 2016), (Nalchigar et al., 2014), (G. Park et al., 2017)
Standards	Enterprise-wide standards (for development processes, data acquisition, reporting, etc.) may be necessary (Spruit & Sacu, 2015b) to pay down avoid or settle data debt. (Zhu et al., 2016b), (Kretzer et al., 2014b), (Ardagna et al., 2016a), (Shanks et al., 2012a), (O 'donovan et al., 2016), (Brichni & Guedria, 2018b), (Hee Yeong Kim et al., 2018b), (Ali et al., 2016a)

Table 6: The definitions and references for the primary dimension: Requirements collection

Secondary Dimension	**Definition and references**
Business case evaluation	A thorough understanding of the business problem is vital (Dutta & Bose, 2015). Since data science projects deal with a degree of uncertainty (Riungu-Kalliosaari et al., 2017), exploratory analysis of available data at this stage may lead to counter-intuitive findings (Alexander & Lyytinen, 2017). (Altarturi et al., 2018), (Kowalczyk & Buxmann, 2015a), (Gopalkrishnan et al., 2012), (Maleki et al., 2016), (Božič & Dimovski, 2019), (Vidgen et al., 2017), (Lakoju & Serrano, 2018), (Li et al., 2019)
Requirements elicitation	Requirements elicitation activity focuses on how to define the requirements of the resulting data analytics system. This activity becomes more complex as traditional software engineering methodologies may not be sufficient for implementing big data systems (Altarturi et al., 2018).

	(Rohleder et al., 2014), (Batra, 2018), (Kretzer et al., 2014a), (Yasin et al., 2018), (Nalchigar & Yu, 2017), (Spruit & Sacu, 2015a), (Hazen et al., 2017), (Lavalle et al., 2011)
Use case identification	Organizations should identify valuable use cases before or in conjunction with checking available data sources (Yasin et al., 2018). (Power, 2015), (Golightly et al., 2018), (Rahman & Aldhaban, 2015), (Schüritz et al., 2017), (Altarturi et al., 2018), (Lavalle et al., 2011)
Partnership with stakeholders	In data-driven environments, stronger partnerships with customers and stakeholders lead to improved benefits (Zhan et al., 2018). (F. L. Wang, Rischmoller, Reed, et al., 2018), (Batra, 2017), (Rao et al., 2018), (Jose et al., 2017), (Dutta & Bose, 2015), (Truong & Dustdar, 2014), (Riungu-Kalliosaari et al., 2017)
Identification of data objects	This activity is interested in the data items that business functions should store & prioritize (Pape, 2016) based on functional and non-functional requirements (Yasin et al., 2018). (Golightly et al., 2018), (Gopalkrishnan et al., 2012), (Glissmann et al., 2012), (Goben & Raszewski, 2015), (Gončarovs & Grabis, 2017), (Schuff et al., 2018)

Table 7: The definitions and references for the primary dimension: Deployment

Secondary Dimension	**Definition and references**
Delivery	This activity is concerned with deploying the models to a production environment (Gončarovs & Grabis, 2017). (Golightly et al., 2018), (H. M. Chen, Kazman, et al., 2016), (J. Y. Lee et al., 2017), (Dutta & Bose, 2015), (Riungu-Kalliosaari et al., 2017)

4.3.2. Organization Theme

To create sustainable value with data-driven approaches, an organization needs various skills in many fields, such as data science, big data, and software engineering. Data scientists who build exploratory models on data are not as experienced as software engineers when building production services (Google, 2020). Additionally, there is likely an expertise gap between the analytics unit and business functions. One owns programmers who have the knowledge and experience in data science, while the other has domain specialists with expertise and in-depth knowledge in a particular domain (Gokalp et al., 2016). IT personnel should be capable of deploying and configuring distributed systems, storage solutions, and programming environments. Social skills, such as storytelling (Knaflic, 2015), play a critical role in communicating analytics results and relevant data products to end-users. Project management skills should support the iterative development process to mitigate risks while managing scope and expectations (Batra, 2017). The human resources department should develop strategies to acquire and retain a workforce if there is a shortage of cloud computing, big data, and data science skills in the market. For example, an easier way to retain data scientists is to make them work on exciting problems (Vidgen et al., 2017).

Adopting big data analytics necessitates an organizational transformation (Mikalef, van de Wetering, et al., 2018). Such a transformation usually involves changing the organization's structure and culture. For example, the organization can deploy analytics teams in different modes: centralized, a center of excellence, local, or hybrid (Lismont et al., 2017). Choosing the right one depends heavily on the organizational context. Since building systems with machine learning components involve risks, success also rests on a suitable culture whose traits are experimentation, continuous learning, data literacy, trust, and openness (Anderson, 2015).

Top management must lead the changing processes and practices through transformation (Golightly et al., 2018). Senior executives should override their intuition when evidence obtained from data analytics shows quite the opposite (Anderson, 2015). In data-driven environments where all units are supposed to share data, a shared understanding among stakeholders is critical, and leadership is necessary to manage conflicts (Batra, 2017). Finally, models do not make decisions but provide actionable insights to decision-makers to do so. Decisions based on models to gain insights for ill-intended purposes are not helpful (Hazen et al., 2018).

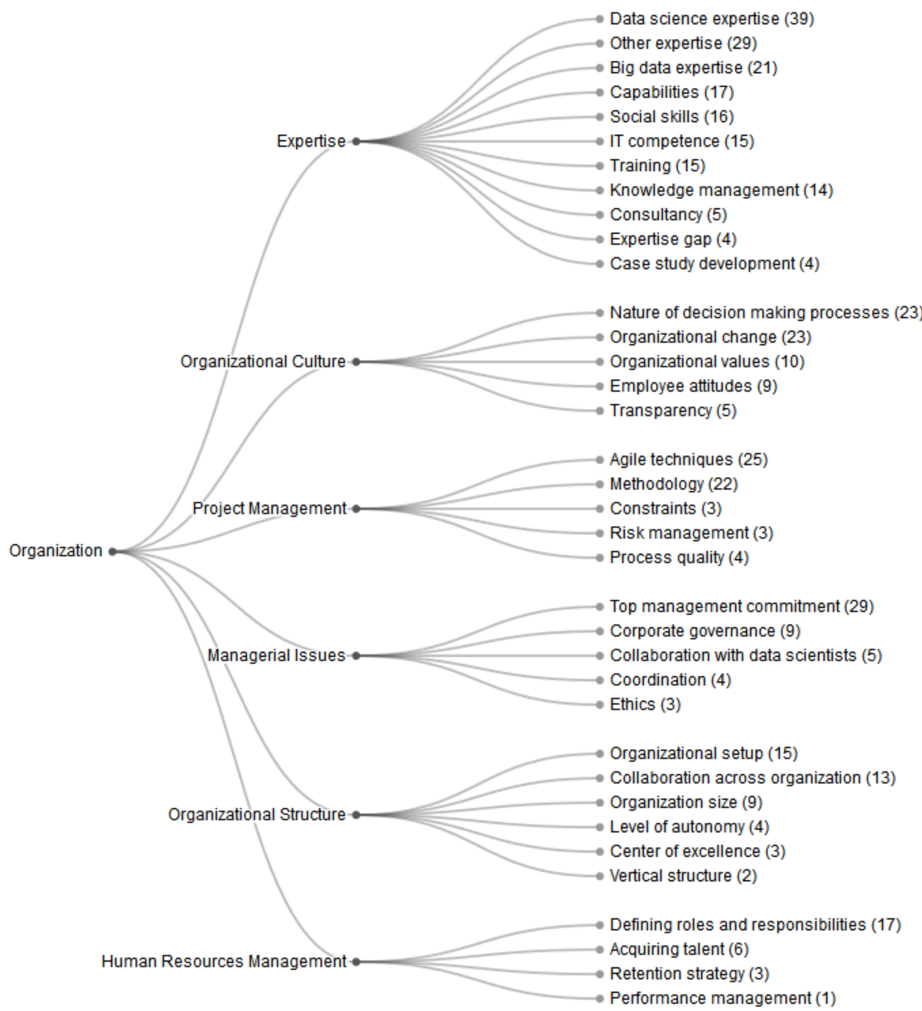

Figure 5: Secondary dimensions of organization theme and number of articles that reference each secondary dimension

Table 8: The definitions and references for the primary dimension: Expertise

Secondary Dimension	**Definition and references**
Data science expertise	Data science expertise refers to the analytical, statistical, and hacking knowledge and skills of people in the organization (Chatfield et al., 2014). (Mikalef, Pappas, et al., 2018), (Golightly et al., 2018), (Iqbal et al., 2018), (Liu & Shi, 2015), (F. L. Wang, Rischmoller, Reed, et al., 2018), (Ramanathan et al., 2017), (Kridel & Dolk, 2013), (S. Verma, 2017), (Koronios et al., 2014),

	(Krishnamoorthi & Mathew, 2018), (Božič & Dimovski, 2019), (H. Chen et al., 2012), (Malaka & Brown, 2015), (Tucker et al., 2017), (Miller, 2018), (Shanks & Sharma, 2011), (Brichni & Guedria, 2018a), (Aho & Uden, 2014), (Coleman et al., 2016), (Schüritz et al., 2017), (Akter et al., 2016), (Cegielski & Jones-Farmer, 2016), (Vidgen et al., 2017), (Lamba & Singh, 2018), (J. S. Saltz & Grady, 2017), (Garmaki et al., 2016), (Mikalef, Giannakos, et al., 2018), (Popovič et al., 2018), (Koelbl et al., 2018), (Baškarada & Koronios, 2017), (Schuff et al., 2018), (Trieu, 2017), (Ramesh & Ramakrishna, 2018), (Lismont et al., 2017)
Other expertise	Other expertise includes domain expertise, business expertise, and project management expertise. (Zorrilla & García-Saiz, 2013), (Liu & Shi, 2015), (F. L. Wang, Rischmoller, Reed, et al., 2018), (Batra, 2017), (Rao et al., 2018), (Kowalczyk & Buxmann, 2015a), (Kridel & Dolk, 2013), (Wamba et al., 2017), (Koronios et al., 2014), (Shanks et al., 2012b), (Božič & Dimovski, 2019), (H. Chen et al., 2012), (Miller, 2018), (Debortoli et al., 2010), (Roderick et al., 2017), (Carter & Sholler, 2015), (Chatfield et al., 2014), (Aho & Uden, 2014), (Coleman et al., 2016), (Akter et al., 2016), (Cegielski & Jones-Farmer, 2016), (Vidgen et al., 2017), (Mowrer et al., 2017), (Garmaki et al., 2016), (Mikalef, Giannakos, et al., 2018), (Koelbl et al., 2018), (Baškarada & Koronios, 2017), (Smuc et al., 2008), (K. Kim & Lee, 2018)
Big data expertise	Big data expertise refers to the skills required to process data in a distributed fashion. (Mikalef, Pappas, et al., 2018), (Iqbal et al., 2018), (Amankwah-Amoah & Adomako, 2019), (Wamba et al., 2017), (Ardagna et al., 2016b), (S. Verma, 2017), (F. Fernando & Engel, 2018), (Koronios et al., 2014), (Al-Hakimi, 2017), (Malaka & Brown, 2015), (Debortoli et al., 2010), (V. Grover et al., 2018), (Dubey & Gunasekaran, 2015), (Coleman et al., 2016), (Lamba & Singh, 2018), (Schüll & Maslan, 2018), (Mowrer et al., 2017), (Garmaki et al., 2016), (Mikalef, Giannakos, et al., 2018), (Aldea et al., 2018), (Adrian et al., 2017)
Capabilities	Organizational capabilities are a collection of routines purposely built by focusing on complex interactions. The resource-based view posits that organizations achieve a competitive advantage based on the capabilities under their control. In contrast, the dynamic capabilities view posits firms achieve that advantage by adapting to changing environments (Mikalef, Pappas, et al., 2018). (Miller, 2018), (V. Grover et al., 2018), (Seddon et al., 2017), (Raguseo & Vitari, 2018), (Ji-fan Ren et al., 2017), (Bekmamedova & Shanks, 2014), (Kurniawati et al., 2013), (Popovič et al., 2018), (Shanks et al., 2010), (Cao et al., 2019), (Seddon & Constantinidis, 2012), (Y. Wang & Hajli, 2017), (Shuradze & Wagner, 2016), (Shanks & Bekmamedova, 2012a), (D. Q. Chen et al., 2015), (Sena et al., 2019)

Social skills	In addition to technical skills, softer skills (e.g., communication, story-telling) play a role when the analytics process and its products are communicated to end-users (Khachatryan & Karst, 2017). (Mikalef, Pappas, et al., 2018), (Liu & Shi, 2015), (Batra, 2018), (Koronios et al., 2014), (H. Chen et al., 2012), (Willcocks et al., 2012), (Tucker et al., 2017), (Chatfield et al., 2014), (Thirathon et al., 2018), (Magee et al., 2016), (Vidgen et al., 2017), (J. S. Saltz et al., 2016), (Endert et al., 2014), (Baškarada & Koronios, 2017), (Riungu-Kalliosaari et al., 2017)
IT competence	The firm should possess personnel capable of deploying and configuring necessary platforms and services such as cloud computing, storage solutions, and programming environments. (Mikalef, Pappas, et al., 2018), (Golightly et al., 2018), (Liu & Shi, 2015), (Ramanathan et al., 2017), (Batra, 2018), (Wamba et al., 2017), (Koronios et al., 2014), (H. Chen et al., 2012), (Willcocks et al., 2012), (Miller, 2018), (Debortoli et al., 2010), (Pérez-González et al., 2019), (Cegielski & Jones-Farmer, 2016), (Garmaki et al., 2016), (Baškarada & Koronios, 2017)
Training	Significant returns on big data are not possible without appropriate training (Dubey & Gunasekaran, 2015). On the other hand, data scientists should be aware of the term continuous learning, in which they iteratively gain an understanding of the business problem and application domain (Riungu-Kalliosaari et al., 2017). (Phillips-Wren & Hoskisson, 2015), (Tan & Haji, 2017), (N. Shah et al., 2017), (Maleki et al., 2016), (Malaka & Brown, 2015), (Brichni & Guedria, 2018a), (Hee Yeong Kim et al., 2018a), (Günther et al., 2017), (Phillips-Wren & Hoskisson, 2015), (Schüritz et al., 2017), (Carillo, 2017), (Alexander & Lyytinen, 2017), (Giacumo et al., 2018)
Knowledge management	Capturing and maintaining knowledge about the complexity of solutions, analysis, and underlying IT resources is critical for the long-term utilization of analytics results (Golightly et al., 2018). (Depeige & Doyencourt, 2015), (Kretzer et al., 2014a), (Gunasekaran et al., 2017), (Koronios et al., 2014), (Gudfinnsson & Strand, 2018), (Tucker et al., 2017), (Bhimani & Willcocks, 2014), (Spruit & Sacu, 2015a), (Schüritz et al., 2017), (Intezari & Gressel, 2017), (Kamoun-Chouk et al., 2017), (Janssen et al., 2017), (Del Vecchio et al., 2018)
Consultancy	Data analytics consulting services can be provided by consulting firms (Iqbal et al., 2018) or academia. (H. M. Chen, Kazman, et al., 2016), (Bhimani & Willcocks, 2014), (Bekmamedova & Shanks, 2014), (H. M. Chen, Schutz, et al., 2016)

Expertise gap	The expertise gap is prominent in managerial thinking and domain knowledge that data scientists should possess (Mikalef, Giannakos, et al., 2018). (Kowalczyk & Buxmann, 2015a), (Kridel & Dolk, 2013), (Riungu-Kalliosaari et al., 2017)
Case study development	Development of internal (& incorporation of external) representative case studies and success stories can be stimulating and trend-setting for organizations (Iqbal et al., 2018). (Rikhardsson & Yigitbasioglu, 2018), (Coleman et al., 2016), (Zimmermann et al., 2016)

Table 9: The definitions and references for the primary dimension: Organizational culture

Secondary Dimension	**Definition and references**
Nature of decision-making processes	Models do not make decisions but provide insights to decision-makers to do so. Decisions based on models used to gain insight on ill-intended problems would not be useful (Hazen et al., 2018). Besides, decisions should not be based on superiority or experience. They should instead be fact-based decisions aligned with business goals (Giacumo et al., 2018). (Power, 2015), (Kowalczyk & Buxmann, 2015a), (Wamba et al., 2017), (S. Verma, 2017), (Kowalczyk & Buxmann, 2014), (Krishnamoorthi & Mathew, 2018), (V. Grover et al., 2018), (King, 2016), (Plomp, 2017), (Thirathon et al., 2018), (Cao & Duan, 2017), (Intezari & Gressel, 2017), (Zimmermann et al., 2016), (N. Verma & Voida, 2016), (Kowalczyk & Buxmann, 2015b), (Popovič et al., 2018), (Koelbl et al., 2018), (Jaklič et al., 2018), (Grund & Meier, 2016), (Shanks & Bekmamedova, 2012a), (Lavalle et al., 2011)
Organizational change	There are several factors to accomplish and sustain organizational change (N. Shah et al., 2017). This activity is related to organizational learning, organizational readiness (Giacumo et al., 2018), organizational inertia, and new knowledge diffusion (Mikalef, van de Wetering, et al., 2018). (Mikalef, Pappas, et al., 2018), (Shanks & Bekmamedova, 2012b), (Ramanathan et al., 2017), (Phillips-Wren & Hoskisson, 2015), (Brock & Khan, 2017), (Gunasekaran et al., 2017), (Al-Hakimi, 2017), (Božič & Dimovski, 2019), (V. Grover et al., 2018), (Phillips-Wren & Hoskisson, 2014), (Seddon et al., 2017), (Ahmad et al., 2016), (Lam et al., 2017), (Vidgen et al., 2017), (Brennan et al., 2018), (Caesarius & Hohenthal, 2018), (Koelbl et al., 2018), (Seddon & Constantinidis, 2012), (Fosso Wamba et al., 2015), (D. Q. Chen et al., 2015)
Organizational values	Values are organizational norms, values, and behavioral patterns resulting in systematic ways of analyzing data and making insights available for the right

	audience (Krishnamoorthi & Mathew, 2018). Goals-first, trusting questioning, and iterative approaches are valuable in the data-driven paradigm (Anderson, 2015). (Batra, 2017), (N. Shah et al., 2017), (Tucker et al., 2017), (Spruit & Sacu, 2015a), (Al-Hitmi & Sherif, 2018), (Brennan et al., 2018), (Bekmamedova & Shanks, 2014), (Jaklič et al., 2018)
Employee attitudes	Employees as individuals are directly dependent upon themselves to accomplish organizational transformation (N. Shah et al., 2017). (Golightly et al., 2018), (S. Verma et al., 2018), (Brock & Khan, 2017), (Gunasekaran et al., 2017), (Carter & Sholler, 2015), (Al-Hitmi & Sherif, 2018), (Mushore & Kyobe, 2017), (Daradkeh, 2019)
Transparency	Organizations should be transparent about how data is used to generate trust (Vidgen et al., 2017). (Batra, 2017), (Kretzer et al., 2014a), (Gopalkrishnan et al., 2012), (Fosso Wamba et al., 2015)

Table 10: The definitions and references for the primary dimension: Project management

Secondary Dimension	**Definition and references**
Agile techniques	Companies routinely launch new products as quickly as possible, gather feedback, and quickly relaunch in an iterative cycle. In data-driven environments, feedback from stakeholders can be collected quickly (Zhan et al., 2018). Hence agile techniques can be incorporated into data-driven workflows easily. (Golightly et al., 2018), (Liu & Shi, 2015), (F. L. Wang, Rischmoller, Reed, et al., 2018), (Batra, 2017), (Batra, 2018), (Sharma et al., 2017), (Rao et al., 2018), (Koronios et al., 2014), (Willcocks et al., 2012), (Ereth & Baars, 2015), (Kisielnicki & Misiak, 2016), (Shcherbakov et al., 2014), (Vidgen et al., 2017), (Paschke, 2016), (Koch & Peters, 2017), (Fabijan et al., 2017), (Serrato & Ramirez, 2016), (Gončarovs & Grabis, 2017), (Smuc et al., 2008), (Riungu-Kalliosaari et al., 2017), (Fosso Wamba et al., 2015), (Ramesh & Ramakrishna, 2018), (J. S. Saltz & Shamshurin, 2018), (H. M. Chen, Kazman, et al., 2016)
Methodology	Methodology refers to proper planning, process definition, or roadmap for developing data science capabilities in organizations. (Belfo & Andreica, 2018), (Dittert et al., 2018), (J. Saltz et al., 2017), (Ghabri et al., 2018), (Sharma et al., 2017), (Rao et al., 2018), (Kridel & Dolk, 2013), (H. M. Chen, Kazman, et al., 2016), (Nino et al., 2015), (O'Donovan et al., 2015), (Hindle & Vidgen, 2018), (Spruit & Sacu, 2015a), (Bhattacharya et al., 2009), (Shcherbakov et al., 2014), (Dutta & Bose, 2015), (Rane & Mishra, 2018),

	(Gončarovs & Grabis, 2017), (Schuff et al., 2018), (Zahid et al., 2018), (Nalchigar & Yu, 2013), (Ramesh & Ramakrishna, 2018), (Braganza et al., 2017)
Constraints	Constraints can be budget, schedule, quality, and other novelty and complexity factors (Batra, 2018). (Miller, 2018), (S. Shah et al., 2018)
Risk management	This activity copes with identifying, monitoring, and controlling risk (Rekha & Parvathi, 2015). Given the complexity of data science projects, risk management should be implemented as part of the organization's project management. (Golightly et al., 2018), (Batra, 2018)
Process quality	High-quality processes are essential for the success of data science initiatives. For example, Malaka and Brown (Malaka & Brown, 2015) report, participants in their case study main business challenge to the adoption of big data analytics arises from unclear communication channels, unclear distribution of data within the organization, and the unclear process of getting data into a big data solution., (Spruit & Sacu, 2015a), (Cao & Duan, 2017), (Tao & Gao, 2016)

Table 11: The definitions and references for the primary dimension: Managerial issues

Secondary Dimension	**Definition and references**
Top management commitment	Senior people (especially at the top) must support the delivery of technology and lead the changing processes and practices (Golightly et al., 2018). (Mikalef, Pappas, et al., 2018), (Shanks & Bekmamedova, 2012b), (Batra, 2017), (Ramanathan et al., 2017), (Batra, 2018), (Rao et al., 2018), (Phillips-Wren & Hoskisson, 2015), (F. Fernando & Engel, 2018), (Gunasekaran et al., 2017), (N. Shah et al., 2017), (Koronios et al., 2014), (Al-Hakimi, 2017), (Krishnamoorthi & Mathew, 2018), (Shanks et al., 2012b), (Miller, 2018), (Phillips-Wren & Hoskisson, 2014), (Spruit & Sacu, 2015a), (Jobs et al., 2016), (Lamba & Singh, 2018), (Schüll & Maslan, 2018), (Popovič et al., 2018), (Koelbl et al., 2018), (Barahona et al., 2016), (Seddon & Constantinidis, 2012), (Lai et al., 2018), (Shanks & Bekmamedova, 2012a), (Ramesh & Ramakrishna, 2018), (D. Q. Chen et al., 2015)
Corporate governance	Corporate governance is concerned with specifying accountability and decision rights to ensure that value is obtained from IT investments (Shanks et al., 2012b). (S. Verma, 2017), (V. Grover et al., 2018), (Hee Yeong Kim et al., 2018a), (Lamba & Singh, 2018), (Koelbl et al., 2018), (Mikalef, Pappas, et al., 2018), (Yu et al., 2012), (Ramesh & Ramakrishna, 2018)

Collaboration with data scientists	Collaboration among analysts and decision-makers is key to creating a value chain (Janssen et al., 2017). Higher quantitative skills of managers and particular soft interaction skills of analysts create an incentive for managers to make analytics-based decisions (Thirathon et al., 2018). (Kowalczyk & Buxmann, 2015a), (Kowalczyk & Buxmann, 2015b), (Bygstad et al., 2019)
Coordination	Actions should be taken to create a shared vision within the enterprise and ensure coordination between different silos or business units. Coordination is crucial for data-driven environments where all units are supposed to be sharing data (Batra, 2017). (Wamba et al., 2017), (Akter et al., 2016), (Riungu-Kalliosaari et al., 2017)
Ethics	Dishonesty, deception, lack of intellectual integrity, and failure to exercise judgment have implications for the information systems discipline (Pauleen et al., 2017). (Vidgen et al., 2017), (Angelopoulos et al., 2016)

Table 12: The definitions and references for the primary dimension: Organizational structure

Secondary Dimension	**Definition and references**
Organizational setup	Organizational setup is concerned with the way communities and teams are formed for reasoning, decision making, development, and other activities. For example, a centralized analytics team can offer economies of scale and scope (Lismont et al., 2017). (Mikalef, Pappas, et al., 2018), (Sharma et al., 2017), (O'Donovan et al., 2015), (Debortoli et al., 2010), (Shanks & Sharma, 2011), (Roderick et al., 2017), (Ereth, 2018), (Günther et al., 2017), (Spruit & Sacu, 2015a), (Coleman et al., 2016), (Vidgen et al., 2017), (J. S. Saltz et al., 2016), (Baškarada & Koronios, 2017), (Bygstad et al., 2019)
Collaboration across organization	Collaboration across the organization, such as creating consistent shared standardized reports and data repositories, can accelerate the core business's ability to perform a full range of analytics (Shanmuganathan, 2019). (Zhan et al., 2018), (Batra, 2017), (Kowalczyk & Buxmann, 2015a), (N. Shah et al., 2017), (Koronios et al., 2014), (Gopalkrishnan et al., 2012), (Božič & Dimovski, 2019), (Tucker et al., 2017), (Vidgen et al., 2017), (Dutta & Bose, 2015), (Alexander & Lyytinen, 2017), (Pickering, 2013)

Organization size	Start-ups, small and medium-sized enterprises, and larger firms may possess and need different analytics competencies and resources (Amankwah-Amoah & Adomako, 2019). (Tan & Haji, 2017), (Llave, 2017), (Gudfinnsson & Strand, 2018), (Thirathon et al., 2018), (Coleman et al., 2016), (S. Shah et al., 2018), (Trieu, 2017), (Hartmann et al., 2016)
Level of autonomy	Greater independence and autonomy for teams can play an essential role in stimulating innovation (Zhan et al., 2018). (Batra, 2017), (Shanks & Sharma, 2011), (Shcherbakov et al., 2014)
Center of excellence	Analytics competency centers are considered a practical approach to deal with the shortage of analytical skills, housing expertise, and providing service to business units (Günther et al., 2017). (Schüritz et al., 2017), (Lismont et al., 2017)
Vertical structure	Ever the last decade, horizontal rather than vertical structures are gaining acceptance in organizations (Lamba & Singh, 2018). (Amankwah-Amoah & Adomako, 2019)

Table 13: The definitions and references for the primary dimension: HR Management

Secondary Dimension	**Definition and references**
Defining roles and responsibilities	This activity aims to define the roles and responsibilities such as deep knowledge analyst, data-savvy user, technology support specialist, and data science researcher (Cegielski & Jones-Farmer, 2016). In this context, DataOps is an emerging discipline that takes cues from DevOps that aims to improve quality, speed, and collaboration in data-driven environments (Ereth, 2018). (Golightly et al., 2018), (Altarturi et al., 2018), (Kowalczyk & Buxmann, 2015a), (H. M. Chen, Kazman, et al., 2016), (N. Shah et al., 2017), (Maleki et al., 2016), (J. Saltz et al., 2017), (Koelbl et al., 2018), (Baškarada & Koronios, 2017), (Riungu-Kalliosaari et al., 2017), (Zahid et al., 2018), (Bygstad et al., 2019), (Zhao et al., 2018), (H.-M. Chen, Kazman, et al., 2016)
Acquiring talent	There exists a skills shortage for cloud computing, big data, and data science skills (Willcocks et al., 2012). Organizations have to develop strategies to acquire talent in such an environment to develop the necessary expertise. (Koronios et al., 2014), (Günther et al., 2017), (Coleman et al., 2016), (Vidgen et al., 2017), (H. M. Chen, Schutz, et al., 2016)

Retention strategy	Once acquired, organizations need to retain their workforce. For example, data scientists are retained if given exciting problems and have career paths (Vidgen et al., 2017). (Willcocks et al., 2012), (Giacumo et al., 2018)
Performance management	Performance management is the process of discovering crucial human performance gaps and designing & implementing cost-effective & ethically justifiable interventions to close these gaps (Giacumo et al., 2018).

4.3.3. Technology Theme

The technology architecture must meet requirements such as timely response, flexibility, scalability, and multi-tenancy (Maturana & Asenjo, 2015). Open-source tools and platforms make data analytics technologies more accessible to organizations (Rao et al., 2018), but the configuration and orchestration of various distributed technologies are complicated. There is an abundance of open-source tools and application frameworks for processing big data (Gökalp et al., 2019). Organizations should build their technology architecture with suitable technologies that satisfy data science requirements.

Ensuring security and privacy is particularly critical when dealing with sensitive data. Due to regulations, there are industries where it may be risky to store sensitive data on the public cloud. In this case, setting up and managing on-premises resources is a viable option. Organizations should still be aware of architecture and infrastructure lock-in in all setups (Pérez-González et al., 2019) and the IT debt associated with changing the underlying technologies while the operations persist.

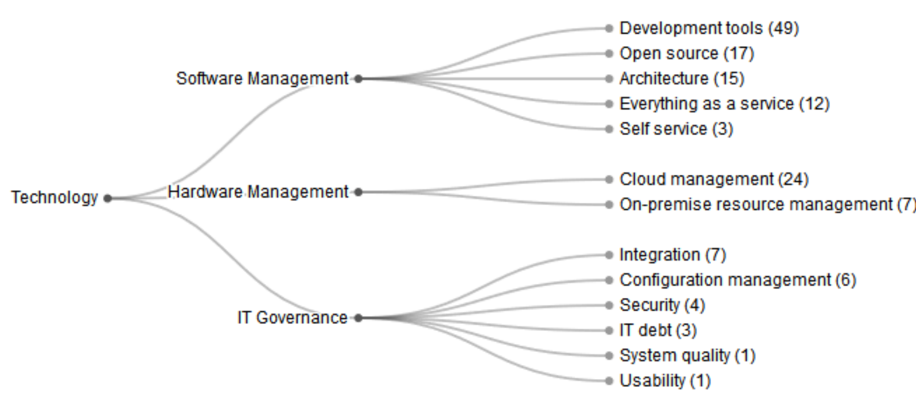

Figure 6: Secondary dimensions of technology theme and number of articles that reference each secondary dimension

Table 14: The definitions and references for the primary dimension: Software management

Secondary Dimension	Definition and references
Development tools	Development tools include storage platforms such as Hadoop Distributed File System (Rubab et al., 2019), big data processing tools such as Spark, programming languages such as Python, machine learning libraries such as BigML, and source code management platforms such as GIT (Zahid et al., 2018). (Batra, 2017), (Ramanathan et al., 2017), (Kridel & Dolk, 2013), (Wu et al., 2016), (Ardagna et al., 2016b), (Choi et al., 2018), (Chong & Shi, 2015), (P. Grover & Kar, 2017), (S. Verma, 2017), (F. Fernando & Engel, 2018), (H. M. Chen, Kazman, et al., 2016), (Assunção et al., 2015), (Muhammad Habib ur et al., 2016), (Nanjappan et al., 2017), (Maleki et al., 2016), (Krishnamoorthi & Mathew, 2018), (Bedeley et al., 2018), (Y. Wang & Byrd, 2017), (Božič & Dimovski, 2019), (Alade, 2017), (Ricardo et al., 2008), (Sahu et al., 2017), (Brichni & Guedria, 2018a), (Hee Yeong Kim et al., 2018a), (Bihl et al., 2016), (Zafeiropoulos et al., 2018), (Rosenthal et al., 2015), (Spruit & Sacu, 2015a), (Gupta et al., 2013), (Cegielski & Jones-Farmer, 2016), (Khatri, 2016), (Pusala et al., 2016), (Roy et al., 2014), (Vo et al., 2017), (Chung & Chung, 2013), (Milosevic et al., 2016), (Janković et al., 2018), (Convertino & Echenique, 2017), (Rekha & Parvathi, 2015), (Ali et al., 2016b), (Schuff et al., 2018), (Zhao et al., 2018), (Ramesh & Ramakrishna, 2018), (Lavalle et al., 2011), (G. Park et al., 2017), (H.-M. Chen, Kazman, et al., 2016)
Open-source	Open-source platforms make data analytics technologies more accessible to organizations with fewer resources, and they enhance agility by solution experimentation (Rao et al., 2018). Hence, open-source has become the de-facto processing platform for big data (Davenport & Dyché, 2013). (Wu et al., 2016), (H. M. Chen, Kazman, et al., 2016), (N. Shah et al., 2017), (Maturana & Asenjo, 2015), (Ricardo et al., 2008), (Brichni & Guedria, 2018a), (Zafeiropoulos et al., 2018), (Gupta et al., 2013), (Akter et al., 2016), (Roy et al., 2014), (Schüll & Maslan, 2018), (Rubab et al., 2019), (Ghose & Dam, 2014), (Zahid et al., 2018)
Architecture	As data collected from machines and applications scale-up, a proper architecture should satisfy timely response, flexibility, scalability, and reusability (Maturana & Asenjo, 2015). (Mikalef, Pappas, et al., 2018), (Batra, 2017), (Rao et al., 2018), (S. Verma et al., 2018), (Wamba et al., 2017), (Božič & Dimovski, 2019), (Malaka & Brown, 2015), (Willcocks et al., 2012), (Ereth & Baars, 2015), (Brichni & Guedria, 2018a), (Kasim et al., 2012), (Glissmann et al., 2012), (Akter et al., 2016), (Janssen et al., 2017)

Everything-as-a-service	Everything-as-a-service is about presenting cloud-based services such as Business Intelligence as a Service to lower costs and improve service quality (Sangupamba Mwilu et al., 2016). (Depeige & Doyencourt, 2015), (Rao et al., 2018), (Kridel & Dolk, 2013), (Ardagna et al., 2016b), (H. M. Chen, Kazman, et al., 2016), (Assunção et al., 2015), (Tucker et al., 2017), (Kasim et al., 2012), (Bhattacharya et al., 2009), (Dwivedi & Kulkarni, 2008), (Truong & Dustdar, 2014)
Self-service	Self-service BI&A are concerned with giving business users the analysis and reporting tools without requiring IT intervention. This concept is also related to democratizing analytics (Schuff et al., 2018). (Zorrilla & García-Saiz, 2013), (F. L. Wang, Rischmoller, Reed, et al., 2018)

Table 15: The definitions and references for the primary dimension: Hardware management

Secondary Dimension	**Definition and references**
Cloud management	Characteristics of cloud computing make it the most accessible infrastructure, especially for small and medium-sized enterprises to implement big data analytics (Wu et al., 2016). (Banda & Ngassam, 2017), (Chaoui & Makdoun, 2017), (Ardagna et al., 2016b), (Y. Wang, Kung, & Byrd, 2018), (H. M. Chen, Kazman, et al., 2016), (Assunção et al., 2015), (Sá et al., 2015), (Muhammad Habib ur et al., 2016), (Maleki et al., 2016), (Llave, 2017), (Maturana & Asenjo, 2015), (Willcocks et al., 2012), (Ereth & Baars, 2015), (Kasim et al., 2012), (Sangupamba Mwilu et al., 2016), (Pérez-González et al., 2019), (Gupta et al., 2013), (Coleman et al., 2016), (J. Y. Lee et al., 2017), (Paschke, 2016), (Milosevic et al., 2016), (Watson, 2018), (Radha et al., 2015)
On-premise resource management	For some instances, on-premise deployment is more suitable. In this case, organizations face various choices for BI&A deployment (Banda & Ngassam, 2017). (Chong & Shi, 2015), (Koronios et al., 2014), (Shanks et al., 2012b), (Spruit & Sacu, 2015a), (Coleman et al., 2016), (Riungu-Kalliosaari et al., 2017)

Table 16: The definitions and references for the primary dimension: IT governance

Secondary Dimension	**Definition and references**
Integration	Integration is about IT integration with the organizational processes (Lamba & Singh, 2018) and other IT systems (S. Verma et al., 2018).

	(Koronios et al., 2014), (Willcocks et al., 2012), (Dutta & Bose, 2015), (Ji-fan Ren et al., 2017), (Atyeh et al., 2017)
Configuration management	Proper configurations should be set for the cloud infrastructure (Banda & Ngassam, 2017) and the development tools used for analytics. (Derguech et al., 2014), (Huang et al., 2009), (Willcocks et al., 2012), (Paschke, 2016), (C. K. M. Lee et al., 2015)
Security	Security management is critical, and it has several components: authentication, authorization, tool-based security, and role-based security (Spruit & Sacu, 2015a). (C. K. M. Lee et al., 2015), (Banda & Ngassam, 2017), (Miller, 2018)
IT debt	Organizations should be aware of architecture and infrastructure lock-in (Pérez-González et al., 2019). IT debt is the cost associated with changing the underlying technologies while operations persist. (Banda & Ngassam, 2017), (Huang et al., 2009)
System quality	System quality involves reliability, adaptability, accessibility, and response time (Ji-fan Ren et al., 2017).
Usability	Usability is concerned with the simplicity and ease of using a tool or platform (Atyeh et al., 2017).

4.3.4. Strategy Theme

Exploiting data can enable an organization to offer new products and services, save costs, and improve decision-making (Davenport, 2014). However, blindly collected and analyzed data without an appropriate business strategy has no value (Liu & Shi, 2015). The business strategy should explain the potential benefits and how to realize them together with well-defined key performance indicators (KPIs) (Mikalef, Pappas, et al., 2018). Senior management can then prioritize data science initiatives according to the business strategy.

Finally, the market structure (e.g., regulations, competition intensity, disruption potential, macro-uncertainties) governs the full value realization from data (Fosso Wamba et al., 2015). Enterprises are facing digital transformation as a substantial paradigm shift that reshapes industries and disrupts businesses. This phenomenon affects business models and cross-organizational relationships. The boundaries between sectors, partners, and competitors are blurring (Rogers, 2016). Thus, realizing value from data is only possible through data flow across the value chain, which depends on collaborative relationships (Golightly et al., 2018).

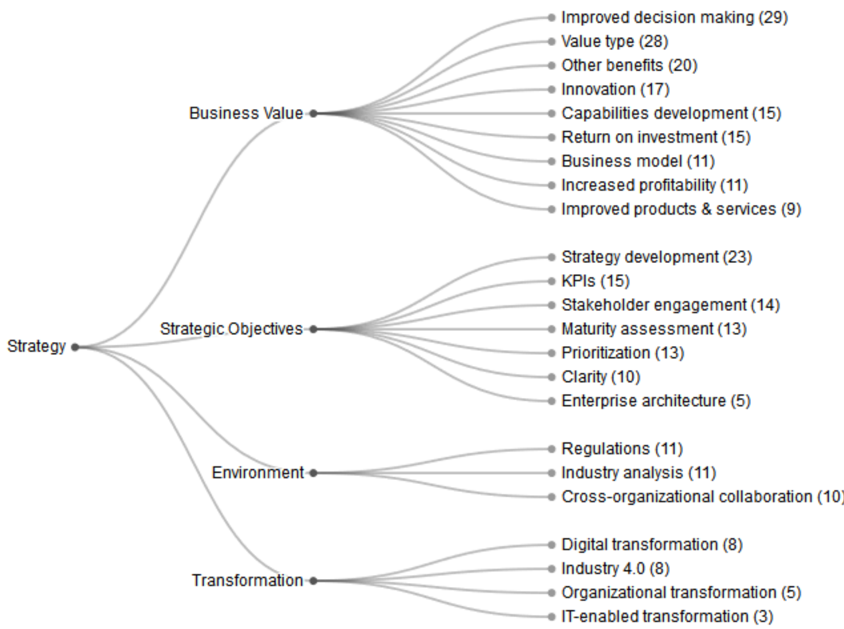

Figure 7: Secondary dimensions of "strategy theme" and number of articles that reference each secondary dimension

Table 17: The definitions and references for the primary dimension: Business value

Secondary Dimension	**Definition and references**
Improved decision-making	Faster and better decisions can be made if semi-structured and unstructured data can be incorporated into the decision-making process (Davenport, 2014). (Altarturi et al., 2018), (Liu & Shi, 2015), (Holsapple et al., 2014), (Rialti et al., 2018), (Kowalczyk & Buxmann, 2014), (Tan & Haji, 2017), (Acharya et al., 2018), (Esswein & Chamoni, 2018), (Y. Wang & Byrd, 2017), (Božič & Dimovski, 2019), (Goul, 2010), (Gudfinnsson & Strand, 2018), (Furtado et al., 2016), (V. Grover et al., 2018), (Muhammad et al., 2010), (Segooa & Kalema, 2018), (Mushore & Kyobe, 2017), (Brennan et al., 2018), (Bayamlıoğlu & Leenes, 2018), (Cao et al., 2019), (Jelonek et al., 2019), (Serrato & Ramirez, 2016), (Seddon & Constantinidis, 2012), (Shollo, 2011), (Fosso Wamba et al., 2015), (Ji-fan Ren et al., 2017), (Troilo et al., 2016), (Engelseth & Wang, 2018)
Value type	Data analytics methods can be classified as descriptive (what happened), diagnostic or inquisitive (why it happened), predictive (what is likely to happen),

	or prescriptive or pre-emptive (what actions to take) (Sivarajah et al., 2017). Accordingly, the analytics activity's value can be a piece of information, an insight, a decision, or an action. (Holsapple et al., 2014), (Derguech et al., 2014), (Y. Wang, Kung, Wang, et al., 2018), (Arora & Malik, 2015), (Kridel & Dolk, 2013), (Hazen et al., 2018), (Kowalczyk & Buxmann, 2014), (Assunção et al., 2015), (O'Donovan et al., 2015), (Al-Hakimi, 2017), (Aydiner et al., 2019), (Bedeley et al., 2018), (Raffoni et al., 2018), (Malaka & Brown, 2015), (Günther et al., 2017), (Phillips-Wren & Hoskisson, 2014), (Bihl et al., 2016), (Cao & Duan, 2017), (Pusala et al., 2016), (Daily & Peterson, 2017), (Nair, 2015), (Hosoya & Kamioka, 2018), (Tanwar et al., 2015), (Ramesh & Ramakrishna, 2018), (Lavalle et al., 2011), (H.-M. Chen, Kazman, et al., 2016), (H. M. Chen, Schutz, et al., 2016)
Other benefits	Incorporation of data-driven approaches might have other benefits such as IT infrastructure benefits (e.g., building flexible architecture), operational benefits (e.g., cycle time reduction, risk management), managerial benefits (e.g., improved decision making), strategic benefits, and organizational benefits (e.g., knowledge creation, organizational learning) (Y. Wang, Kung, & Byrd, 2018). (Y. Wang, Kung, Wang, et al., 2018), (Wamba et al., 2017), (F. Fernando & Engel, 2018), (Gunasekaran et al., 2017), (Tan & Haji, 2017), (Krishnamoorthi & Mathew, 2018), (Y. Wang & Byrd, 2017), (Goul, 2010), (V. Grover et al., 2018), (Melville, 2015), (Seddon et al., 2017), (Kurniawati et al., 2013), (Müller et al., 2018), (Popovič et al., 2018), (Ziora, 2015), (Hosoya & Kamioka, 2018), (Y. Wang & Hajli, 2017), (Fosso Wamba et al., 2015), (Engelseth & Wang, 2018)
Innovation	Davenport (Davenport, 2014) describes the most ambitious thing an organization can do with big data is to deliver new product and service offerings based on data. (Zhan et al., 2018), (Y. Wang, Kung, & Byrd, 2018), (H. M. Chen, Kazman, et al., 2016), (Tan & Haji, 2017), (Johnson et al., 2017), (Muhammad Habib ur et al., 2016), (Božič & Dimovski, 2019), (Schuh et al., 2015), (Gudfinnsson & Strand, 2018), (Furtado et al., 2016), (Melville, 2015), (Cao & Duan, 2017), (Raguseo & Vitari, 2018), (Ziora, 2015), (Lehrer et al., 2018), (Fosso Wamba et al., 2015)
Capabilities development	Capabilities development is about creating enhanced capabilities and know-how. (Shanks & Bekmamedova, 2012b), (Rao et al., 2018), (Y. Wang, Kung, Wang, et al., 2018), (Wamba et al., 2017), (Y. Wang, Kung, & Byrd, 2018), (Gunasekaran et al., 2017), (Johnson et al., 2017), (Acharya et al., 2018), (Krishnamoorthi & Mathew, 2018), (Božič & Dimovski, 2019), (Melville, 2015), (Ji-fan Ren et al., 2017), (Y. Wang & Hajli, 2017), (D. Q. Chen et al., 2015), (Braganza et al., 2017)
Return on investment	Investment in data-related activities involves costs of various processes, and very few companies take steps to measure return on investment for their data analytics efforts (V. Grover et al., 2018).

	(Shanks & Bekmamedova, 2012b), (Rao et al., 2018), (Phillips-Wren & Hoskisson, 2015), (Bumblauskas et al., 2017), (Koronios et al., 2014), (Gopalkrishnan et al., 2012), (Krishnamoorthi & Mathew, 2018), (Malaka & Brown, 2015), (Nalchigar & Yu, 2017), (Phillips-Wren & Hoskisson, 2014), (Rohleder et al., 2014), (Mikalef, Pappas, et al., 2018), (Ramesh & Ramakrishna, 2018), (Baesens et al., 2016)
Business model	Business models express how an organization creates and delivers value. Companies only recently started to make use of data sources (Hartmann et al., 2016). Hence, data-driven business models are needed. For example, Nino et al. (Nino et al., 2015) introduce the case of a manufacturing company trying to change its business model leveraging the power of data analytics and big data technologies. (Muhammad Habib ur et al., 2016), (Maleki et al., 2016), (Günther et al., 2017), (Hindle & Vidgen, 2018), (Schüritz et al., 2017), (Ji-fan Ren et al., 2017), (Schuritz & Satzger, 2016), (Ziora, 2015), (Serrato & Ramirez, 2016)
Increased profitability	Cost reduction can be achieved by implementing distributed storage clusters (Davenport, 2014), or differentiated products/services can be sold at higher prices (F. Fernando & Engel, 2018), leading to increased profitability. (Altarturi et al., 2018), (Gunasekaran et al., 2018), (Furtado et al., 2016), (Miller, 2018), (Jonas, 2018), (Muhammad et al., 2010), (Lam et al., 2017), (Daily & Peterson, 2017), (Sena et al., 2019)
Improved products and services	Through data-enabled service innovation, tailored services can be provided consistently in response to triggers (e.g., customer's current location) (Lehrer et al., 2018). (Gunasekaran et al., 2018), (Muhammad et al., 2010), (Raguseo & Vitari, 2018), (Lam et al., 2017), (Ji-fan Ren et al., 2017), (Kurniawati et al., 2013), (H. M. Chen, Schutz, et al., 2016), (Del Vecchio et al., 2018)

Table 18: The definitions and references for the primary dimension: Strategic objectives

Secondary Dimension	**Definition and references**
Strategy development	Without an appropriate business strategy, blindly collected and analyzed data is without value (Liu & Shi, 2015). (Mikalef, Pappas, et al., 2018), (Foshay et al., 2015), (Rao et al., 2018), (Phillips-Wren & Hoskisson, 2015), (Johnson et al., 2017), (Koronios et al., 2014), (Maleki et al., 2016), (Krishnamoorthi & Mathew, 2018), (Esswein & Chamoni, 2018), (Shanks et al., 2012b), (Goul, 2010), (Nalchigar & Yu, 2017), (Phillips-Wren & Hoskisson, 2014), (Melville, 2015), (Cao & Duan, 2017), (Vidgen et al., 2017),

	(Lamba & Singh, 2018), (Ji-fan Ren et al., 2017), (Fraser, 2017), (Popovič et al., 2018), (Li et al., 2019), (Duke & Ashraf, 2019)
KPIs	Having a good understanding of and knowing how to measure and improve key performance indicators (KPIs) is essential (Mikalef, Pappas, et al., 2018). (Zorrilla & García-Saiz, 2013), (Bumblauskas et al., 2017), (Vera-Baquero et al., 2015), (Ricardo et al., 2008), (Lizotte-Latendresse & Beauregard, 2018), (C. K. M. Lee et al., 2015), (Pape, 2016), (Popovič et al., 2018), (Zahid et al., 2018), (Veneberg et al., 2014), (G. Park et al., 2017)
Stakeholder engagement	Partnership with stakeholders and the ability to understand and co-create with customers lead to accelerated innovation in data-driven environments (Zhan et al., 2018). (Rao et al., 2018), (F. Fernando & Engel, 2018), (Troisi et al., 2018), (Božič & Dimovski, 2019), (Miller, 2018), (Günther et al., 2017), (Alexander & Lyytinen, 2017), (Fromm et al., 2012), (Koch & Peters, 2017), (Y. Chen et al., 2008), (Mohapatra & Ghosh, 2016), (H. M. Chen, Schutz, et al., 2016), (Del Vecchio et al., 2018)
Maturity assessment	Maturity assessment identifies the current state of where the organization stands in terms of data science capabilities (Gökalp et al., 2020). (Rao et al., 2018), (Al-Hakimi, 2017), (Krishnamoorthi & Mathew, 2018), (Llave, 2017), (Gudfinnsson & Strand, 2018), (Spruit & Sacu, 2015a), (Coleman et al., 2016), (Comuzzi & Patel, 2016), (Brennan et al., 2018), (Rane & Mishra, 2018), (Shanks et al., 2010), (Lismont et al., 2017), (Olszak, 2016)
Prioritization	Project selection should be done according to business objectives like efficiency goals and growth objectives (Krishnamoorthi & Mathew, 2018). (Mikalef, Pappas, et al., 2018), (Liu & Shi, 2015), (Batra, 2017), (Rao et al., 2018), (Malaka & Brown, 2015), (Gudfinnsson & Strand, 2018), (Hee Yeong Kim et al., 2018a), (Glissmann et al., 2012), (Shollo, 2011), (Shanks & Bekmamedova, 2012a), (Ramesh & Ramakrishna, 2018), (Gear et al., 1982)
Clarity	An analytics strategy should clearly explain how and where the value will be created (Vidgen et al., 2017). (Altarturi et al., 2018), (Kowalczyk & Buxmann, 2015a), (Hazen et al., 2018), (F. Fernando & Engel, 2018), (Koronios et al., 2014), (Gopalkrishnan et al., 2012), (Miller, 2018), (Rohleder et al., 2014), (Lakoju & Serrano, 2018)
Enterprise architecture	Enterprise architecture is a set of frameworks, methods, models, and tools to help organizations deal with IT capabilities and understand the impacts of changes related to IT on business strategy & performance (Aldea et al., 2018). Long-term

	goals (strategy) can be conceptualized as an enterprise architecture (Shanks & Bekmamedova, 2012a). (Veneberg et al., 2014), (Yu et al., 2012), (Nalchigar & Yu, 2013)

Table 19: The definitions and references for the primary dimension: Environment

Secondary Dimension	**Definition and references**
Regulations	Developers, lawyers, and subjects of decisions by data-driven automated decision-making systems should pay attention to the rule of law and regulations (Bayamlıoğlu & Leenes, 2018). (Iqbal et al., 2018), (Ramanathan et al., 2017), (Ardagna et al., 2016b), (Pauleen et al., 2017), (Krishnamoorthi & Mathew, 2018), (Tucker et al., 2017), (M.-K. Kim & Park, 2016), (Vidgen et al., 2017), (Trieu, 2017), (Li et al., 2019)
Industry analysis	The value realization from data will be subject to the industry structure (e.g., the intensity of competition, performance transparency, disruption potential, macro uncertainties) (Fosso Wamba et al., 2015). (Ramanathan et al., 2017), (Amankwah-Amoah & Adomako, 2019), (Johnson et al., 2017), (Pauleen et al., 2017), (Krishnamoorthi & Mathew, 2018), (Schüll & Maslan, 2018), (Ghose & Dam, 2014), (Trieu, 2017), (D. Q. Chen et al., 2015), (Baesens et al., 2016)
Cross-organizational collaboration	There are fragmented business functions across organizations as well as partners and suppliers. Development & deployment of predictive and prescriptive activities requires a flow of information between partners. Knowledge of design and functional elements should be spread across the supply chain (Golightly et al., 2018). (Zhan et al., 2018), (Tan & Haji, 2017), (Günther et al., 2017), (M.-K. Kim & Park, 2016), (Hamister et al., 2018), (Vidgen et al., 2017), (Janssen et al., 2017), (Ramesh & Ramakrishna, 2018), (Braganza et al., 2017)

Table 20: The definitions and references for the primary dimension: Transformation

Secondary Dimension	**Definition and references**
Digital transformation	Digital transformation is a more comprehensive organizational transformation as its complexity exceeds IT-enabled transformation, and the range of potential impact and benefits is higher (Ismail et al., 2017). Data is suggested to be one of Rogers's five domains of digital transformation (Rogers, 2016).

	(Brichni & Guedria, 2018a), (Jonas, 2018), (Bhimani & Willcocks, 2014), (Zimmermann et al., 2016), (Serrato & Ramirez, 2016), (Aldea et al., 2018)
Industry 4.0	Different sensors implanted in different machines throughout the entire production value chain produce data. Data analytics is the essential part of the next industrial revolution, where machines will be transformed into fully integrated and automated facilities (Brichni & Guedria, 2018a). (Yasin et al., 2018), (Nino et al., 2015), (O'Donovan et al., 2015), (De Oliveira Cordeiro et al., 2017), (Rubab et al., 2019), (Niño et al., 2016), (Aldea et al., 2018)
Organizational transformation	Business analytics systems support alignment between business and IT strategy, enabling organizational transformation (Shanks et al., 2012b). (Schuh et al., 2015), (Fraser, 2017), (Lavalle et al., 2011), (Baesens et al., 2016)
IT-enabled transformation	Data analytics capabilities can be linked to IT-enabled transformation practices at both evolutionary and revolutionary levels (Y. Wang, Kung, Wang, et al., 2018). (Mikalef, van de Wetering, et al., 2018), (N. Shah et al., 2017)

4.4. Discussion

Amassing over the body of knowledge to answer RQ1 and RQ2, DOTS research framework comprises 20 primary and 105 secondary dimensions. A summary capturing the key points and the trend is presented in this section. Then the decisions while developing the DOTS research framework are discussed.

As paradigm shifts force organizations to become data-driven, focusing solely on data analysis is not enough for an organization to transform to become data-driven. To implement successful data-driven operations to create business value, an organization must also concentrate on other factors and solve different challenges. These factors and challenges include strategic, environmental, cultural, human, and technological elements. Nevertheless, the prominent dimensions are data quality, data discovery and acquisition, data science expertise, and the development tools used for data science.

The given themes are not mutually exclusive. For example, stakeholder engagement exists under data and strategy themes as a secondary dimension. Software architecture, development tools, and tool selection are categorized under data and technology themes with their respective descriptions. The analytics level (descriptive, diagnostic, predictive, and prescriptive) is categorized under the strategy theme to represent the value type.

However, it can also be considered under data theme, under development subdimension. Hence, the four themes interact to complement one another towards successful data-driven organizational transformation.

The categorization presented in this study identified four themes, but there may be more than four themes, such as data governance can be separated from data analytics. In the book Analytics at Work, Davenport et al. (2010) describe the DELTA model in which five generic themes are Data, Enterprise, Leadership, Targets, and Analysts. The DOTS research framework considers enterprise, leadership, and analysts under the organization theme. It is intended to keep it simple and memorable for practitioners.

4.5. Limitations

The following limitations are identified: (1) The review and coding processes are based on human judgment. This limitation is mitigated by developing a solid research protocol and regularly meeting with contributing researchers between iterations. (2) There may be factors and challenges that are not published in the literature or detected by the systematic method. (3) Application of the framework and relevance of dimensions can differ among industries since there may be more domain-specific subdimensions.

This chapter introduces a comprehensive research framework to guide data-driven transforming organizations. However, the DOTS research framework addresses the "what" questions (see RQ1 & RQ2), but not the "how" questions. For example, not every item in the framework is equally important for every organization. Furthermore, each item's prominence can change during the organization's transformation journey. In Chapter 5, the DSR framework development is explored, synthesizing the DOTS research framework with the widely adopted TRM to address the "how" question.

CHAPTER 5

DEVELOPMENT OF DSR FRAMEWORK

5.1. Customizing TRM for Data Science

TRM is a flexible approach; however, a successful application requires tailoring to a particular situation (Phaal et al., 2013). The roadmap architecture and the roadmapping process must provide a framework for structuring knowledge. Some studies guide and facilitate customization activity (S. Lee & Park, 2005; Phaal et al., 2004b). However, this requirement makes it challenging for organizations to start such initiatives because they lack expertise in roadmapping. To tailor-make the roadmap architecture and the standard T-Plan process according to the data science context, we follow the guidelines recommended by Phaal et al. (2004a):

a) **Context:** Context refers to the nature of the issue that triggered roadmapping (Phaal et al., 2004b). It accounts for the scope, focus, aims, and roles. The aims and scope heavily depend on the organizational context, and they are topics of discussion for kickoff. We provide practitioners with specifications and templates (Tables 21, 22, 23, and 24) to help with these discussions using DOTS themes.

b) **Architecture:** The timeframe (in the horizontal axis) and the layers (in the vertical axis) define the roadmap's structure. The layers comprise the purpose, value, and resources. The roadmap also shows the linkages between these layers. Accordingly, we modify the business reconfiguration template (Phaal et al., 2004b) to explore the strategic, data-related, technological, and organizational migration paths and bridge the gaps.

c) **Process:** The roadmapping process contains the macro-level and micro-level sets of staged activities (Phaal et al., 2004b). The former accounts for planning, workshops, and review activities, whereas the latter emphasizes the workshops' agenda. This study leverages two industry best practices to customize the roadmapping process for DSR: CRISP-DM and workshop-based T-Plan. Consequently, the process backs up practitioner knowledge with quantitative evidence throughout the workshops. The DOTS themes determine the consecutive workshops' agenda. Inspired by Kerr et al. (2019), we also inserted a pre-workshop step to incorporate data preparation and analysis activities into the macro-process. Overall, extensible process models, which account for organizational stakeholders, macro-and micro-level tasks, and data flow between these tasks, emerged. BPMN (White, 2004) enabled the graphical representation of business processes and made it easier to obtain feedback from industry experts and apply changes throughout the applications. Figures 11-17 depict the micro-level processes.

5.2. DSR Context

The context defines the nature of the issue that triggered the need to develop a roadmap, and it accounts for the focus, scope, objectives, and participants. Accordingly, this section explains the DSR context and provides practitioners with specifications and templates (Tables 21, 22, 23, and 24) to help them employ the DSR methodology in their businesses.

5.2.1. Focus and Scope

The outputs of data science activities, which are data products and services, should align with the organization's business strategy. Data-related, technological, and organizational resources should support data science processes throughout the roadmap.

The DOTS themes determine the roadmapping initiative boundaries. Not all topics must be within the scope of a roadmapping initiative. For example, a pilot study that involves the data and technology layers can help the organization understand how roadmapping can deliver value, gain stakeholder buy-in, and further customize the process to better suit the context. Table 21 can help the scope of an initiative at the kickoff. The scope should comprise the requirements of data science processes that potentially create business value.

Table 21: Template for scoping DSR

Theme	**Topic**
Data	Data science processes
	Data and metadata objects
	Data sources
	Data governance
Organization	Skills and competences
	Managerial issues

	Human resources management
	Project management
	Organizational structure
	Organizational culture
	Financial resources
Technology	Technology architecture
	Infrastructure management
	Technology governance
Strategy	Business trends and drivers
	Market trends and drivers
	Data science maturity

5.2.2. Aims

The aims of roadmapping initiatives differ depending on the context and organization (Kerr & Phaal, 2019). The aims and expectations are the kickoff activity discussions, and Table 22 provides some short-, mid-, and long-term sample objectives for a DSR initiative.

Table 22: Sample objectives for DSR

Horizon	**Objective**
Short-term	• Identify business problems and opportunity scenarios created by data science. • Analyze data products that can address business and market drivers/trends. • Pinpoint strategic, data-related, technological, and organizational gaps. • Determine and prioritize valuable internal and external data sources. • Find and experiment with tools that can support data science projects. • Support communication between different business functions.

	• Prioritize data science projects and collect the quick-wins. • Get stakeholder buy-in.
Mid-term	• Find sponsorship and funding to make data science projects sustainable. • Build the right technological and organizational capabilities that align with the business strategy. • Understand the organizational bottlenecks that hinder the value of data products. • Adjust the organizational culture and structure to support the use of data-driven approaches.
Long-term	• Get ready for paradigm shifts. • Add data-driven features to products and services. • Explore data-driven business models. • Enable domain experts to create data science workflows through self-service architecture. • Embrace a data-driven culture.

5.2.3. Roles and Responsibilities

Who takes part in a roadmapping initiative heavily depends on the organizational context. To this end, we guide organizations to plan initiatives by defining their roles. Tables 23 and 24 explain the roles and show their involvement in teams throughout roadmapping.

Table 23: Groups, group members, and their responsibilities in the roadmapping process

Group	**Members**	**Group's responsibilities**
Executive committee	CEO, Head of Analytics/AI, Head of IT, and COO/Head of Human Resources (HR)	To scope the initiative, assign the roadmap champion, and review the results.
Technical and domain experts	IT experts, HR specialists, data owners, domain experts, and end users of data products/services	To provide pre-workshop inputs and attend respective workshops.

Data experts	Data scientists/analysts, data engineers, and machine learning engineers	To attend pre-workshop and respective workshops.
Roadmap champion	Head of Analytics/AI, Head of Digital Transformation, or similar	To own the business problem, the roadmapping process, and the roadmap. To keep the roadmap up-to-date after roll-out. The champion has the span of influence to get active involvement from a range of business units.
Roadmapping experts	Consultants	To coordinate the initiative and facilitate workshop activities.

Table 24: Members of the roadmapping board

Team		**Members**
Kickoff team		Executive committee Roadmapping experts
Planning team		Roadmap champion Roadmapping experts Head of Analytics/AI
Pre-workshop team		Roadmap champion Roadmapping experts IT experts Data experts Domain experts Data owners
Workshop teams	Strategy workshop team	Roadmap champion Roadmapping experts Executive committee

		Nice to have: Data experts
	Data workshop team	Roadmap champion Roadmapping experts Head of Analytics/AI Head of IT Data experts End users of data products/services Nice to have: IT experts
	Technology workshop team	Roadmap champion Roadmapping experts Head of Analytics/AI Nice to have: Data experts
	Organization workshop team	Roadmap champion Roadmapping experts COO/Head of HR Head of Analytics/AI HR specialists Nice to have: Data experts
Post-workshop team		Roadmap champion Roadmapping experts All stakeholders of roadmapping board
Review committee		Executive committee Roadmap champion

	Roadmapping experts

5.3. DSR Architecture

On the horizontal axis, timeframes depend on the organizational context and the scope of the initiative. Suppose the scope spans organization and strategy themes. In that case, as Phaal and Muller (2009) suggest, five timeframes may be appropriate: the past, 1-year (short-term budget horizon), 3-year (mid-term strategy horizon), 10-year (long-term radar horizon), and vision. The business reconfiguration template (Phaal et al., 2004b) is suitable to explore the migration paths to bridge the gap between the company's current position and its strategic vision. Since successful adoption and diffusion of data-driven approaches necessitate organizational transformation (Mikalef, van de Wetering, et al., 2018), the business reconfiguration template can serve as a guideline. However, it needs to be modified to explore strategic data-related, technological, and organizational migration paths according to data science processes and data management strategies.

Strategy, data, technology, and organizational layers cover purpose, value, and resource perspectives in the vertical axis. The strategy constitutes the purpose layer in the roadmap architecture (Figure 8), comprising business and market trends and drivers. Data science processes consume data and metadata objects created by numerous data sources to produce data products and services, aligning data layers with business strategies. Therefore, the data layer crosscuts the value and resource layers. The technology layer comprises the infrastructure, platforms running on the infrastructure that provide a range of functionalities, and tools not part of those platforms providing specific functionalities. Public or private clouds, workstations, edge devices, hardware acceleration, interconnects, storage devices, high-performance computing centers may be part of the infrastructure depending on the data management strategy, data science process requirements, and IT. Platforms run on top of the infrastructure and implement virtualization, containerization, container orchestration, resource management, model lifecycle management, version control, master data management, batch and stream processing, data ingestion, etc. Chosen tools may provide additional functionalities, such as time-series analysis and visualization libraries. Finally, the organizational gaps and migration paths are depicted at the bottom. Organizational structure, capabilities, values, project management, and partnerships need to change to create sustainable business value from data science activities. Section 5.4 explains the macro-and micro-level set of activities (Figures 10-17), bringing together members of the roadmapping board (Table 24) in a series of workshops, exploring the gaps, and creating migration paths.

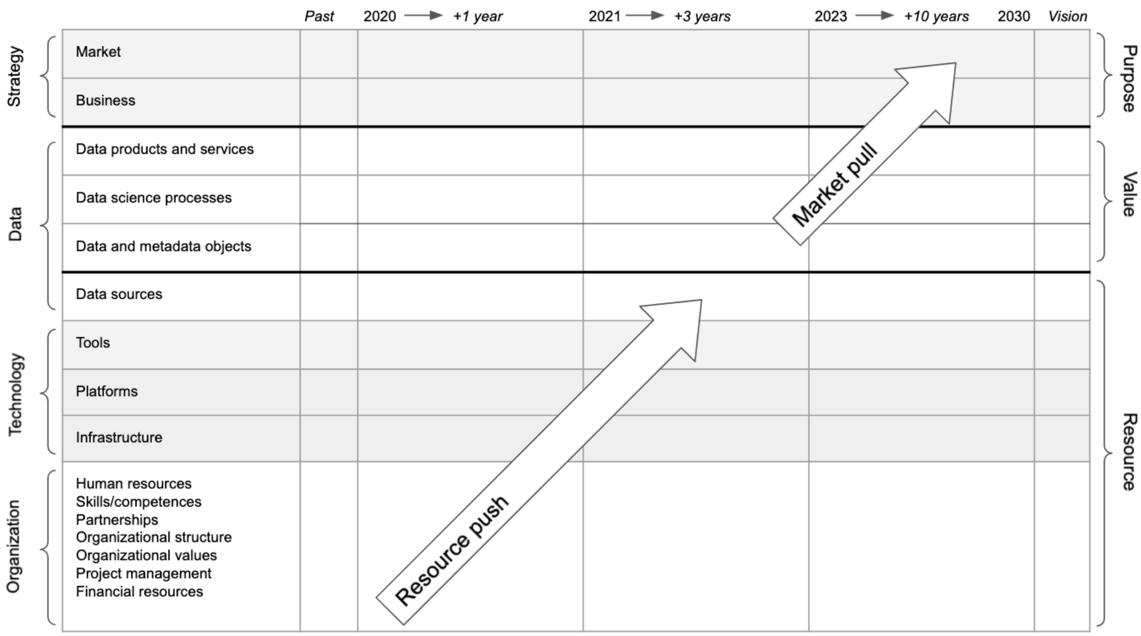

Figure 8: Modified roadmap architecture for DSR [based on (Phaal et al., 2004a)]

5.4. DSR Process

Workshops are recommended for applying a strategic management toolkit (Kerr et al., 2013). They provide a structured setting where people solve inherently complex problems through the combined effort and mutual reliance. In the context of roadmapping for data-driven organizational transformation, workshops can facilitate communication among business functions, generating consensus among all stakeholders (Table 24). Therefore, DSR modifies the workshop-based T-Plan process according to the data science context and adds activities to back up practitioner knowledge with quantitative evidence, creating a hybrid roadmap development method.

The T-Plan has a three-stage process: planning, workshops, and reviewing outcomes (Phaal et al., 2004a). Similarly, the S-Plan, which brings together large groups in a one-or two-day workshop to explore and prioritize strategic issues, also has a three-stage process. Customizing the S-Plan, Kerr et al. (2019) inserted the pre-workshop and post-workshop activities. By capturing participant perspectives before the workshops, the pre-workshop work helped the pre-population of the roadmap landscape. Before the review meeting, the data captured in the workshops were analyzed and synthesized during the post-workshop activities. We argue that pre-workshop work is much more relevant for DSR. The three initial steps of the CRISP-DM process model (Wirth & Hipp, 2000) are business understanding, data understanding, and data preparation. Executing these three steps in the pre-workshop step would identify quick-win opportunities and resource requirements before the workshops begin (Figure 13). Moreover, CRISP-DM also provides a guideline for data workshop deep-dive activity integrating build-measure-learn cycles (Ries, 2011) with the roadmap development process (Figure 15). Figures 9 and 10 illustrate the macro-process for DSR, and the following sections elaborate on the steps as needed.

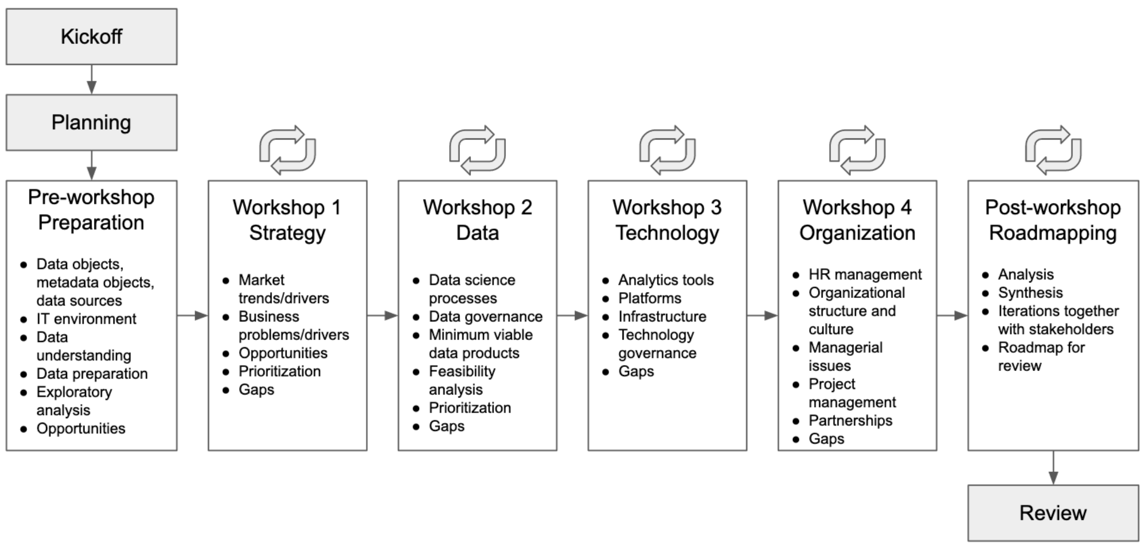

Figure 9: The agenda for strategy, data, technology, and organization workshops [based on (Kerr et al., 2019; Phaal et al., 2004a)]

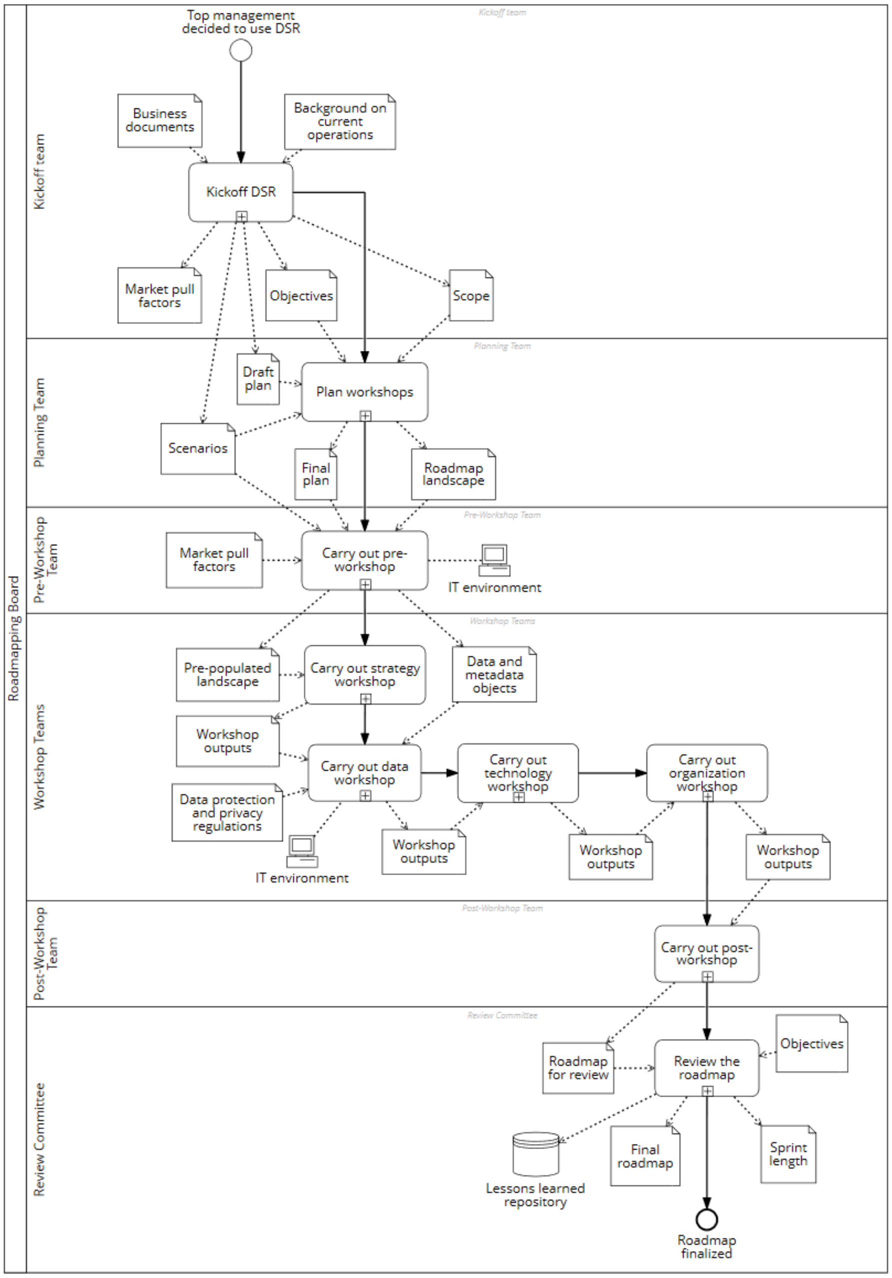

Figure 10: The macro-level process model for carrying out DSR

5.4.1. *Kickoff and Planning*

Kerr and Phaal (2019) recommend starting roadmapping with the question, "why is roadmapping needed?." At the DSR kickoff, it is crucial to establish the rationale for using roadmapping as a management tool. If the underlying reason is "seeing what it can do," we recommend adopting an agile approach by keeping the scope small for a pilot study. Data experts can identify short-term problems and focus on a few data science problems to provide immediate value with roadmapping. This approach would obtain stakeholder buy-in for the next iteration.

The roadmap champion and roadmapping experts prepare the roadmap landscape at the beginning of the planning activity according to the scope. They discuss internal and external data sources they can use in the pre-workshop and data workshop steps, determine the remaining stakeholders of the roadmapping board, and decide on the workshops' schedule. They should also review managerial and technical tools that can support this initiative. These tools are necessary for exploratory data analysis, meetings, and collaborative decision-making. When the final plan is ready, the planning team briefs all the stakeholders. Upon feedback from stakeholders, the plan may need to be revised.

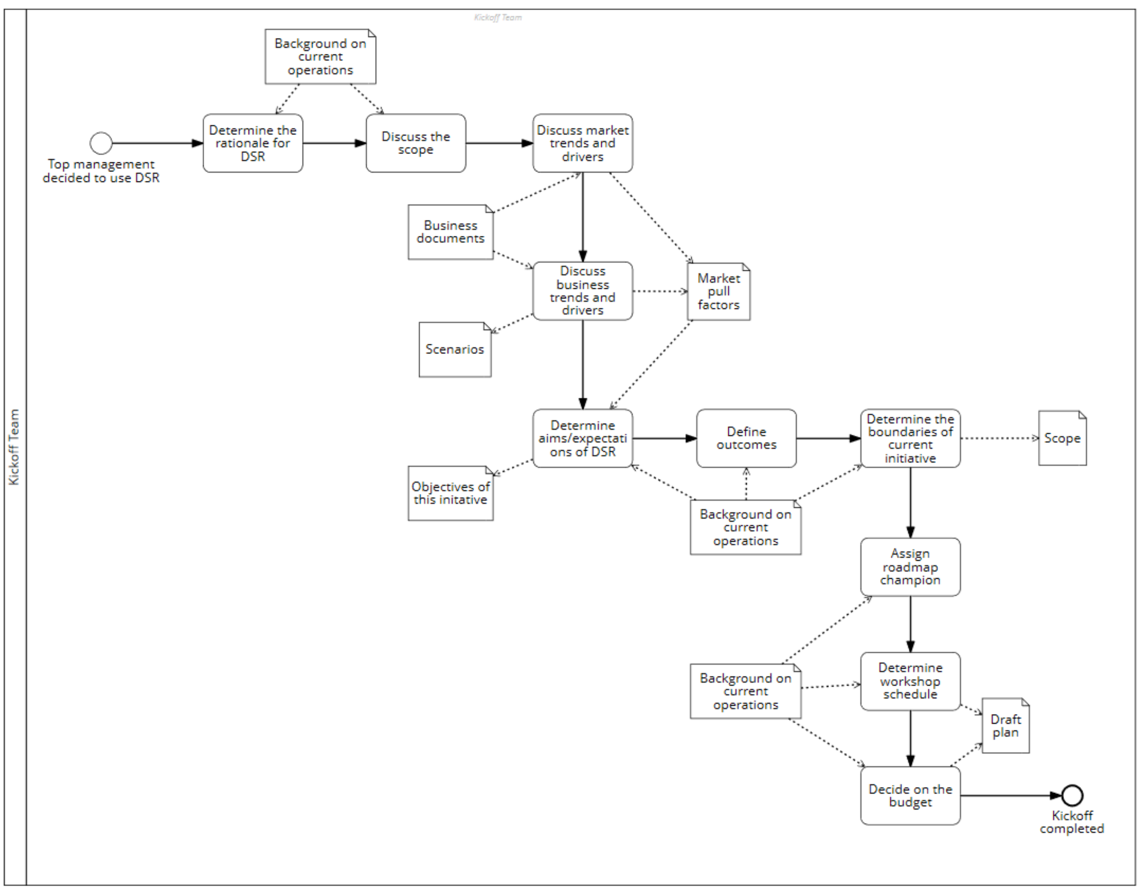

Figure 11: The recommended process model for kicking off DSR

5.4.2 Pre-Workshop

The pre-workshop is a critical step that governs the successful application of the DSR roadmapping initiative. The pre-workshop team should answer two fundamental questions in this step: can they create business value with the data they have, and are there any opportunities that they could not see before. Accordingly, the data experts meet with the data owners to determine data objects, metadata objects, and data sources according to possible scenarios. They should gather internal and external data and access the IT environment before the workshops begin. They make sense of data, do some data preparation, and create exploratory models. Executing much of the first three phases defined by CRISP-DM is a good idea since they are cumbersome for a single workshop. This pre-work maximizes the value of time spent together in strategy, data, and technology workshops. Figure 13 shows the recommended process model for the pre-workshop step.

5.4.3 Workshops

The four workshops in the T-Plan process are market, product, technology, and roadmapping, with a product workshop focusing on the value layer. It is appropriate for data science initiatives to consider data sources and related data products in a single workshop. Therefore, the customized process for DSR comprises of strategy, data, technology, and organization workshops.

The landscaping and landmark exploration activities are suitable to define workshop structures (Kerr et al., 2019). We recommend starting a workshop with all the participants reviewing the landscape while brainstorming and prioritizing landmarks across the landscape. Small groups can deep-dive into landmarks and create business cases. The participants can then come together to discuss business cases and agree on a way forward. Turning workshops into small one-to-two-hour meetings can make the process much more agile by leveraging iterations between deep-dive and landscaping activities (Figure 14). A single workshop can comprise the following parts, especially if the workshops are online.

1. All participants review the landscape and brainstorm landmarks on the landscape.
2. Small groups deep-dive into landmarks.
3. All participants come together to discuss the business cases and update the landscape.
4. Repeat steps two and three as many times as necessary, as suggested by Pearson et al. (2020).

The deep-dive activity for the data workshop (Figure 15) requires particular consideration. This may be straightforward, depending on the quality and completeness of the pre-workshop. However, the agreed pathways in the strategy workshop may require further data analysis. If this is the case, the data experts can leverage the CRISP-DM process model during deep-dive activities, creating minimum viable data products between landscaping meetings. Data owners and data experts also need to create a data management plan vital for the next technology and organization workshops.

5.4.4 Post-Workshop, Review Meeting, and Keeping the Roadmap Alive

Up to this point, workshops should generate many outputs, but not the roadmap. A series of post-workshop activities is necessary to develop a proper roadmap for the final review meeting. (Kerr et al., 2019). These activities include analyzing workshop data, synthesizing the roadmap draft, and iterating the roadmap with stakeholders' feedback. Finally, the review committee evaluates the roadmap regarding the objectives of this initiative. The committee may require updates from the post-workshop team according to the comments until the roadmap is final. After the roadmap is ready for publication, they should also discuss how to keep the roadmap alive, open issues for the next iteration, the responsible party for keeping the roadmap alive, and the time (sprint length) to meet for the next iteration. The recommended sprint length would be three to six months, or meetings can be held upon request. The planning team then publishes the roadmap and runs a retrospective session. They document and store lessons learned about the roadmapping experience and the organization for the next iteration.

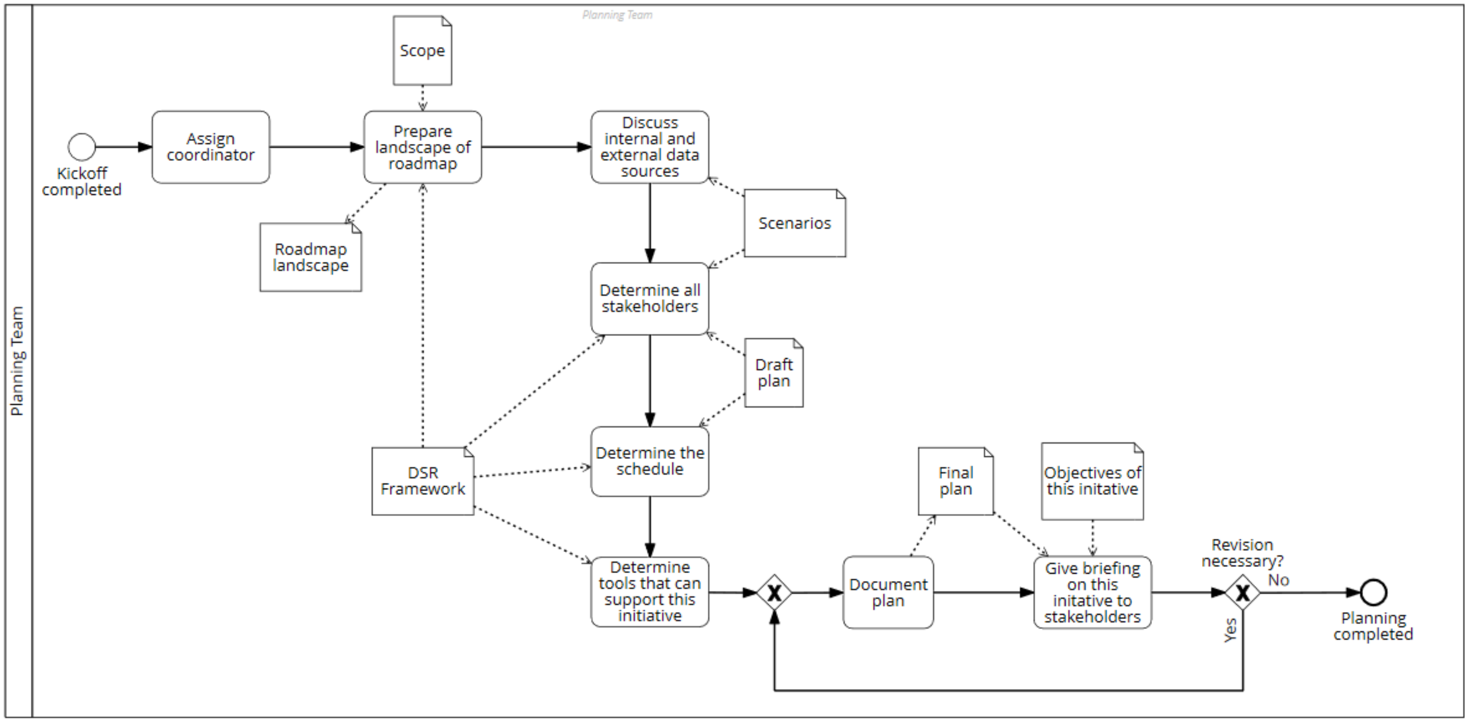

Figure 12: The recommended process model for planning workshops

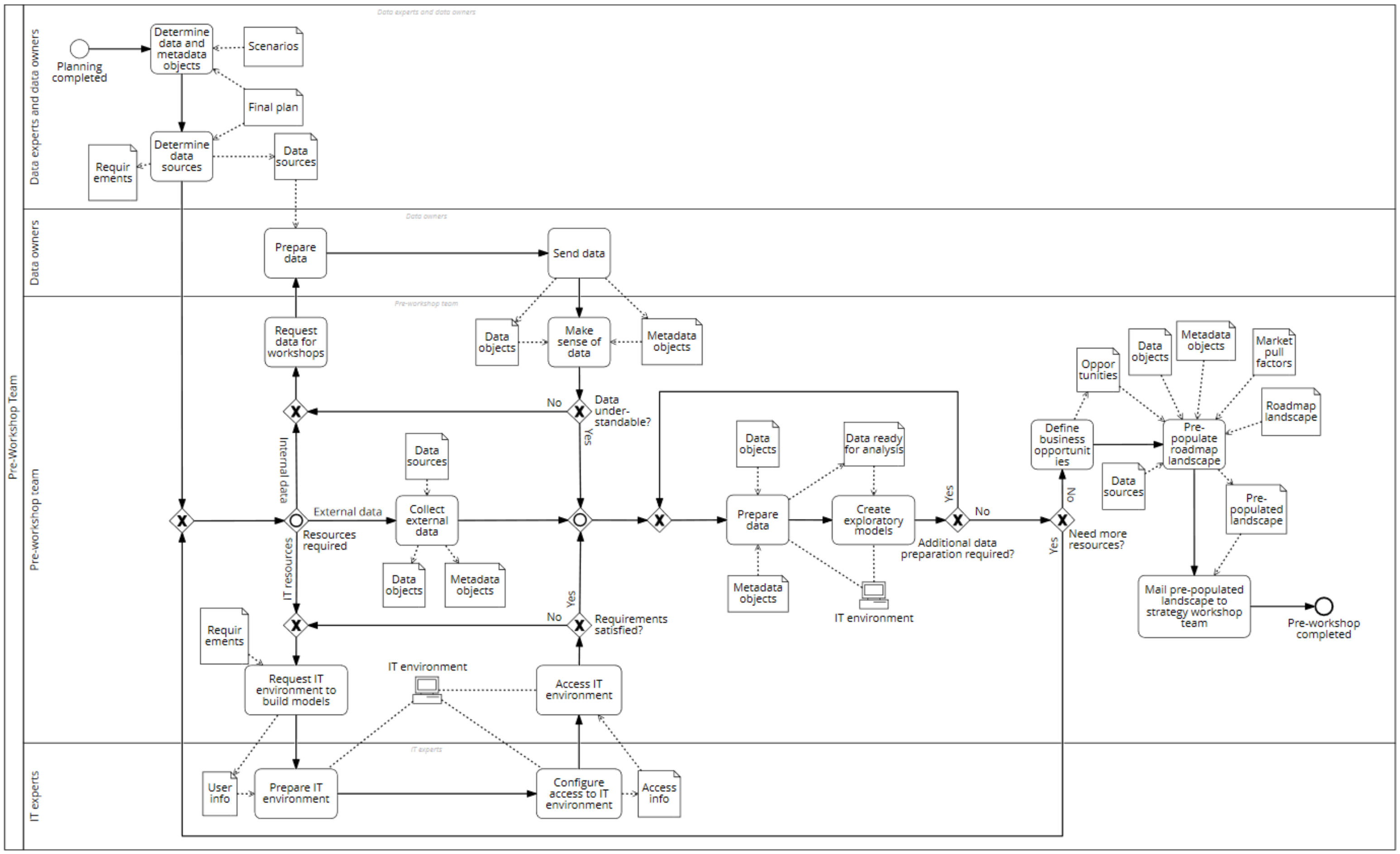

Figure 13: The recommended process model for carrying out pre-workshop

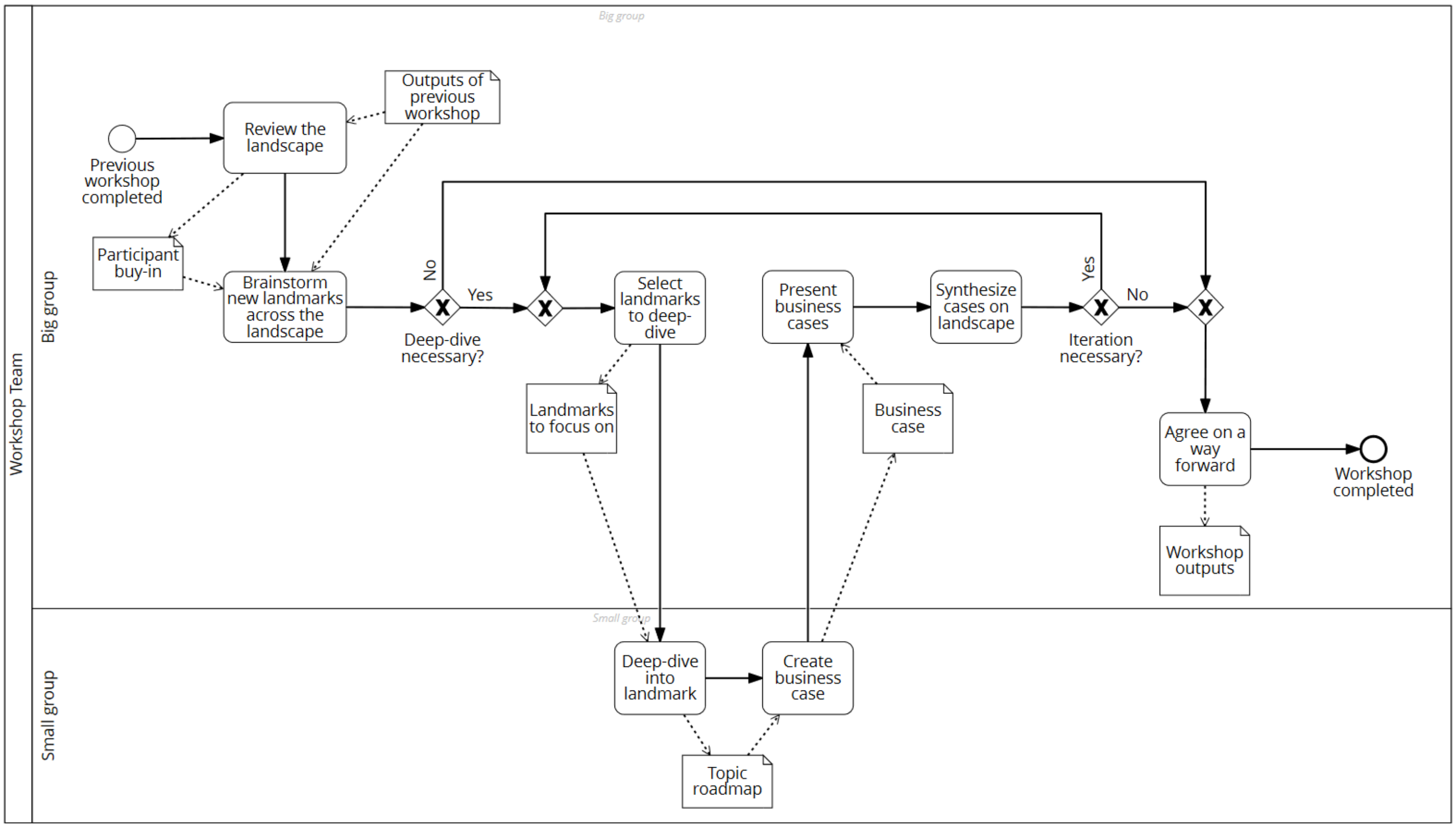

Figure 14: The recommended process model for carrying out strategy, technology, and organization workshops

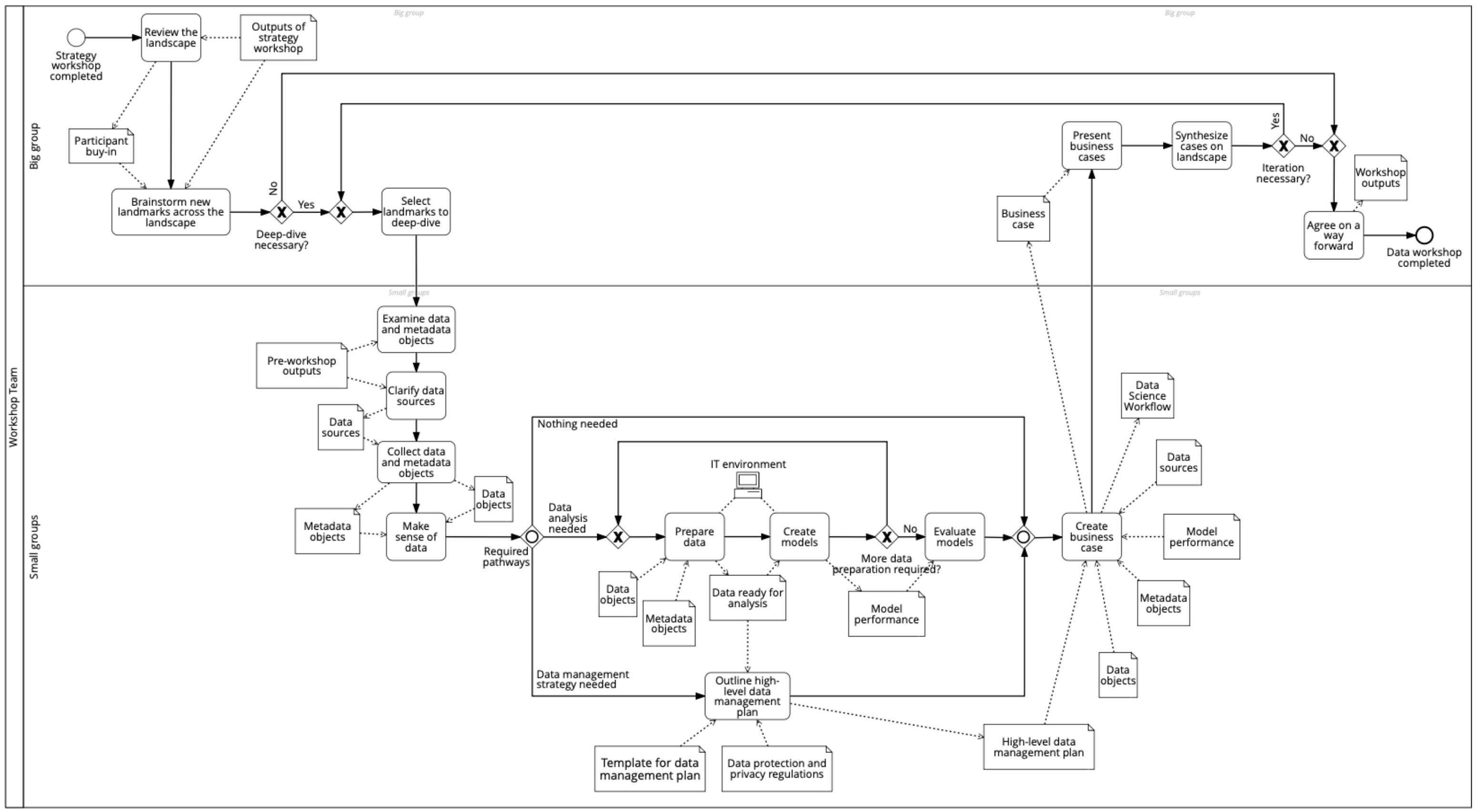

Figure 15: The recommended process model for carrying out data workshop

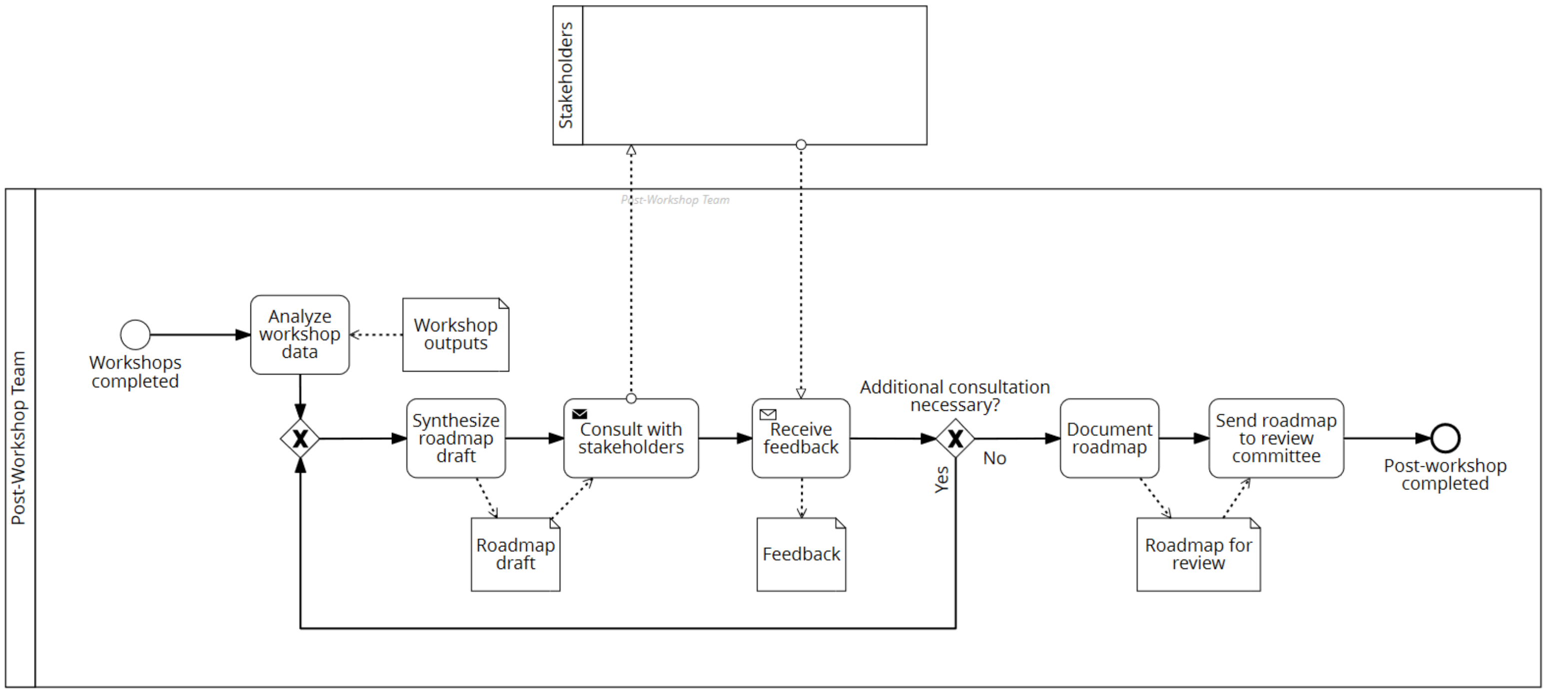

Figure 16: The recommended process model for carrying out post-workshop

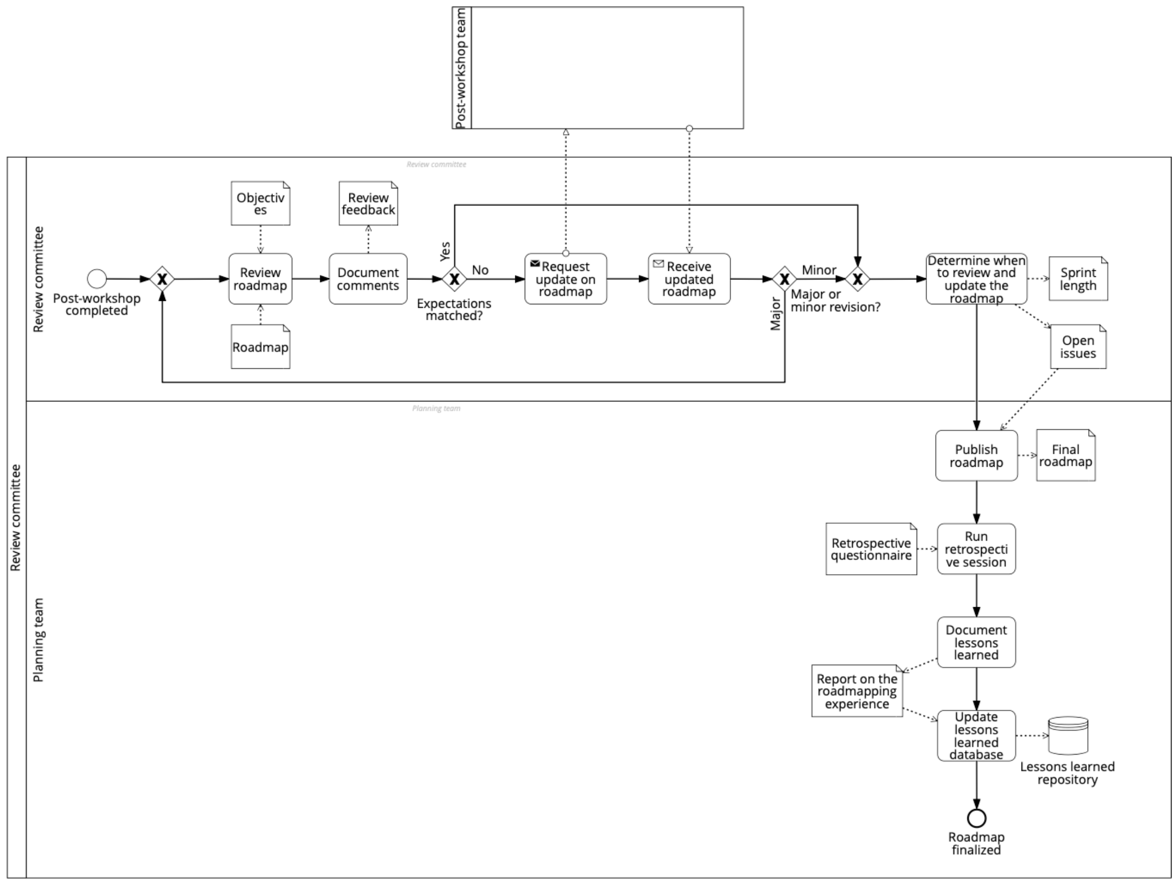

Figure 17: The recommended process model for reviewing the roadmap

CHAPTER 6

APPLICATION OF DSR FRAMEWORK

6.1. Action Research Design

This study adopts the action research design to refine the process models and validate DSR's applicability and usefulness. Using action research design is suitable when the research question is about understanding the process of change or improvement to learn from it (Coughlan & Coghlan, 2002). The underlying beliefs in action research designs are (Easterby-Smith et al., 2015): (1) The best way to learn about an organization is by attempting to change it, and (2) the people implementing these changes should become part of the research process itself. Accordingly, action research is appropriate for developing practical tools such as roadmapping that require working together on "live" management problems and challenges (Kerr et al., 2019).

We collaborated with a cross-disciplinary research group and an oil and gas downstream company through three action research cycles (Figure 18). The research group first applied DSR in a pilot study to understand the shortcomings of the process models and see if DSR can deliver a quick win. Following the adjustments, the group utilized DSR in a second iteration with a broader scope, refining the framework. Lastly, the oil and gas company employed DSR to create its first-pass data science roadmap. The following section elaborates on the backgrounds of both organizations.

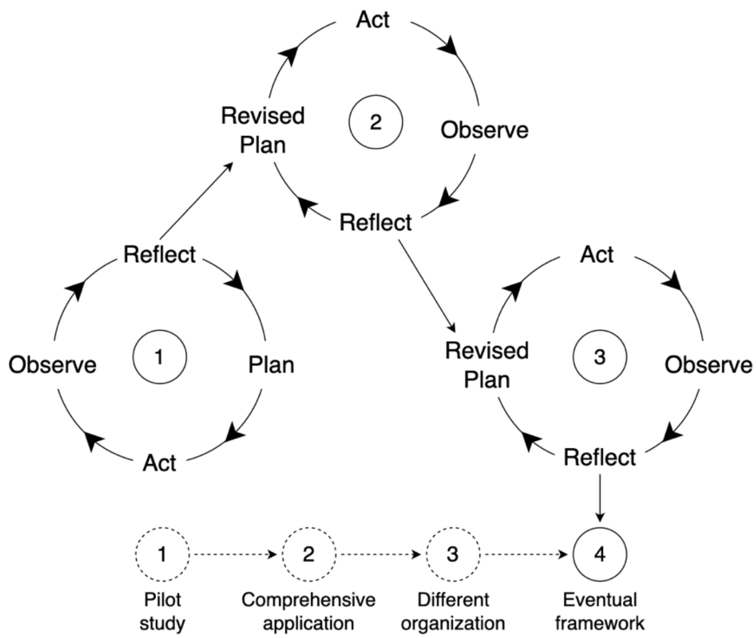

Figure 18: The action research cycle spiral [based on (Altrichter et al., 2002)]

6.2. Organization Backgrounds

6.2.1. The Cross-Disciplinary Research Group

The research group studies Industry 4.0, digital transformation, and data science domains in Turkey. The goal is to produce high-quality research output while addressing real industrial problems. They also provide high-quality courses for M.S. and Ph.D. students. With these courses, the research group can attract skillful students, some of whom are research assistants. The group has ongoing industry collaborations in which they use data science to create value for organizations. The group has a technology infrastructure for research projects and courses. Furthermore, the group's research lab has various IoT devices, including Raspberry Pi[1] units and sensors.

The group was suffering from the lack of a master research plan: when a new student wanted to join the group, they could not immediately suggest a research topic. When there was a potential research topic, the group could not immediately assign it to a student. Furthermore, the group responded to industry problems in an ad hoc manner. Usually, the

[1] https://www.raspberrypi.org/

resulting research projects do not align academic interests with industry problems. Finally, the technology infrastructure and IoT devices were underutilized since they did not meet the research project requirements. Using DSR, the research group wanted to create a master plan to address these shortcomings.

6.2.2. The Oil and Gas Downstream Company

The oil and gas company refines crude oil and distributes gasoline and diesel products. The company has data assets in heterogeneous data warehouses, and the data sources include the Industrial Internet of Things (IIoT) and business processes (e.g., sales and logistics). The company established data analytics centers in two university technoparks in 2018 to become a data-driven organization in collaboration with technology provider SMEs and academia. Since then, the analytical activities have increased, but the data experts have faced challenges at a large and traditional organization operating with siloed business units. The data analytics team wanted to employ DSR to create a first-pass roadmap to plan two to three projects and see the results.

6.3. First Iteration: Pilot Study for Planning Newsletters for the Research Group Members

6.3.1. Kickoff and Planning

This iteration's motivation was to create automatic reports identifying and categorizing critical data objects and capturing data sources. Based on these, the group would determine the data products (newsletters). These newsletters could provide academic and industry trends, drawing a holistic picture for the next iteration (Section 6.4). The executive committee also wanted the group members to experiment with open-source tools to harvest data from sources. Accordingly, this initiative's objectives were as follows:

- Understand the key professionals in the industry and researchers in academia.
- Understand the top conferences and journals in respective fields.
- Capture the data science technology landscape, particularly open-source.
- Capture the data sources in the roadmap landscape. Agree on how to use these data sources, such as how to remove noise and clutter.
- Agree on the automated reports (newsletters) and prioritize the development of these newsletters.
- Perform experiments with open-source tools that will collect the data.
- Reach a common understanding among all the group members.
- Document the roadmap so that all stakeholders can objectively review the plan.

The planning team agreed on the roadmap architecture to include the following layers from top to bottom: newsletters, newsletter projects, data objects, data sources, tools, and platforms. The timeframe would be six months, and the tools that would support this initiative were as follows:

- **Zoom**[2]**:** It was impossible to arrange physical workshops because of the COVID-19 pandemic[3]. Instead, the group relied on virtual arrangements and real-time online tools, where physical meetings were impossible.
- **Coda**[4]**:** The group needed a flexible sticky-note-like capability during the workshops. Coda enabled the group to prepare swimlane-like roadmaps using card components.
- **diagrams.net**[5]**:** This tool allowed the post-workshop team to produce the roadmap properly during the post-workshop.
- **Huginn**[6]**:** Huginn is an open-source tool that enables users to build agents that perform automated tasks online. The agents create and consume events, propagating them on a directed graph.
- **Signavio**[7]**:** Signavio allowed the facilitator team to graphically represent the macro-level and micro-level roadmapping processes (Figures 10-17), incorporating roles in Table 24 and the data flow between steps. The graphical representations made it easier to get feedback during the application, custom-make the process models in the pilot, and save revisions together with comments as lessons learned.

6.3.2. Pre-Workshop and Workshop Activities

The pre-workshop team determined a representative sample of data objects and data sources based on kickoff and planning discussions. The team also deployed Huginn and

---

[2] https://zoom.us/

[3] https://www.who.int/health-topics/coronavirus

[4] https://coda.io/

[5] https://www.diagrams.net/

[6] https://github.com/huginn/huginn

[7] https://www.signavio.com/

accessed the Huginn dashboard to create sample scenarios that aggregated and propagated new publications and events from selected data sources. Finally, they pre-populated the roadmap landscape based on this proof-of-concept implementation.

The pilot study's scope comprised only data and technology workshops. The workshop teams were able to conduct them consecutively on the same day. In both workshops, the workshop teams first reviewed the roadmap landscape, adding and arranging landmarks. Some data sources required more focus, such as WikiCFP[8], which publishes calls for papers. Since category-based scraping would produce many events causing cluttering in the reports, the team selected those conferences that mattered most.

6.3.3. Results of the Pilot Study

One week later, the review committee met again to evaluate the roadmap produced during the post-workshop. The review committee reflected on questions: what was delivered, what did they not get, what were the actual outcomes, what were the successes, what has been the impact, and were they satisfied with roadmapping (Kerr & Phaal, 2019). They agreed that this iteration satisfied all the objectives determined at kickoff. The actual outcomes were newsletters, work packages, schedules, and consensus around the data sources. When asked if they were satisfied with this initiative, the participants answered:

- It has been good seeing which steps would bring the outcomes and when. The roadmap and proof-of-concept newsletters are useful in the end.
- The DSR has been adequate for this particular project. We will see how it performs when there is a portfolio of projects.
- This initiative will carry data utilization to the next level. We had known the data sources on the roadmap all along, but the roadmapping helped our group and cleaned them up. Now, we have an idea about what to do with them.
- During the roadmapping initiative, we have seen new possibilities emerge as we combine information that originates from diverse data sources. We could not see these actionable possibilities before this initiative.

The rationale for deploying DSR was to see if it could deliver a quick win for the research group. The executive committee kept the scope small for the first iteration, the recommended mode of deployment. After the review meeting, we organized a retrospective session with the research group and applied changes to the DSR framework, particularly the process models. They agreed to move on to the second iteration, which

[8] http://www.wikicfp.com/cfp/

would add the strategy and organization layers. Moreover, the outputs of the first iteration helped determine the high-level trends for the second iteration.

6.4. Second Iteration: Building a Comprehensive Data Science Roadmap

6.4.1. Kickoff and Planning

The second iteration's motivation was to create a strategy considering three verticals: academic research, education, and industry collaboration. The three goals applicable to each vertical are creating value, managing resources, and staying up-to-date. The expectation is to produce high-quality research while creating an industrial impact. Table 25 explains the objectives and outcomes of this iteration. After the kickoff team determined the scope using the scoping template (Table 21), they assigned the roadmap champion and settled the tentative schedule.

Table 25: Objectives and expected outcomes for the second iteration

Objectives	**Expected outcomes**
• To plan the data science projects that will have an academic impact and industrial value • To adjust course offerings according to the research plan • To follow the high-level trends, and also lead the change if there is an opportunity • To plan data-related, technological, and organizational resources • To lead the industry partners and students • To create a shared understanding	• A system for staying up-to-date • A roadmap where puzzle pieces fit together • Data-driven decision making for ad-hoc opportunities • New methods and data science projects that have practical implications for the industry • New research ideas • Understanding of how to involve students in research projects

The planning team agreed that the roadmap should plan up to 2023 since they could not see the high-level trends beyond 3-years. After determining the roadmap architecture layers (Figures 20-22), they discussed the data objects and the internal and external data sources they could use in the pre-workshop. They decided on all stakeholders of the roadmapping board and then clarified the tentative schedule.

6.4.2. Pre-Workshop and Workshops

The research group had already identified the data sources they could analyze to determine high-level trends in the pilot study. They analyzed and synthesized these trends by pre-populating the landscape. Furthermore, they had ongoing data science activities from active industry collaborations. They met with data owners to pre-populate the following roadmap landscape layers regarding industry collaborations: data products, data science processes, data and metadata objects, and data sources. The current sections of the remaining layers were pre-populated based on discussions at the planning activity.

The workshop teams gathered four workshops: strategy, data, technology, and organization. All the workshops followed the same structure. Initially, all participants reviewed the landscape, adding, arranging, and adjusting the landmarks. They judged whether small group focus activities were necessary, and if not, agreed on a way forward for the next workshop. For example, the strategy workshop team decided that no focus activity was necessary. However, they put some research items in the academic research's vision section and requested the following workshop teams to figure out how to bridge the gaps in the layers below.

Meanwhile, small group focus activities were necessary for data workshops. Working with the data sources, small groups iteratively determined the viable data products and data science processes and when possible. Some of the processes mapped the strategic trends, such as rising data privacy (IEEE Computer Society, 2021). The research group hypothesized the new data objects and data sources to bridge these gaps. The requirements from the first two workshops, together with high-level strategic trends, set the agenda for the technology workshop. The group forecasted open-source data science technologies and updated the current infrastructure and platforms to prepare for the strategic trends and data layer requirements such as High-Performance Computing (HPC) and AI convergence (Georgiou et al., 2020), GPU-accelerated computing paradigm, and updated data science processes. In the next organization workshop, the research group figured out how to bridge the gaps to satisfy the plans in the respective data and technology layers.

The final roadmap is shown in Figures 20-22. Some details are aggregated and hidden, respecting the group's privacy. However, it is possible to see specific open-source technologies that map high-level strategic trends. Moreover, Figure 19 presents a deep dive into a time-series analytics study. The group planned the respective data and technology layers below for this study to obtain more accurate forecasts and create a reproducible production workflow. The time-series analytics processes in Figure 19 are part of the data science workflows for partner projects in the final roadmap (Figure 21).

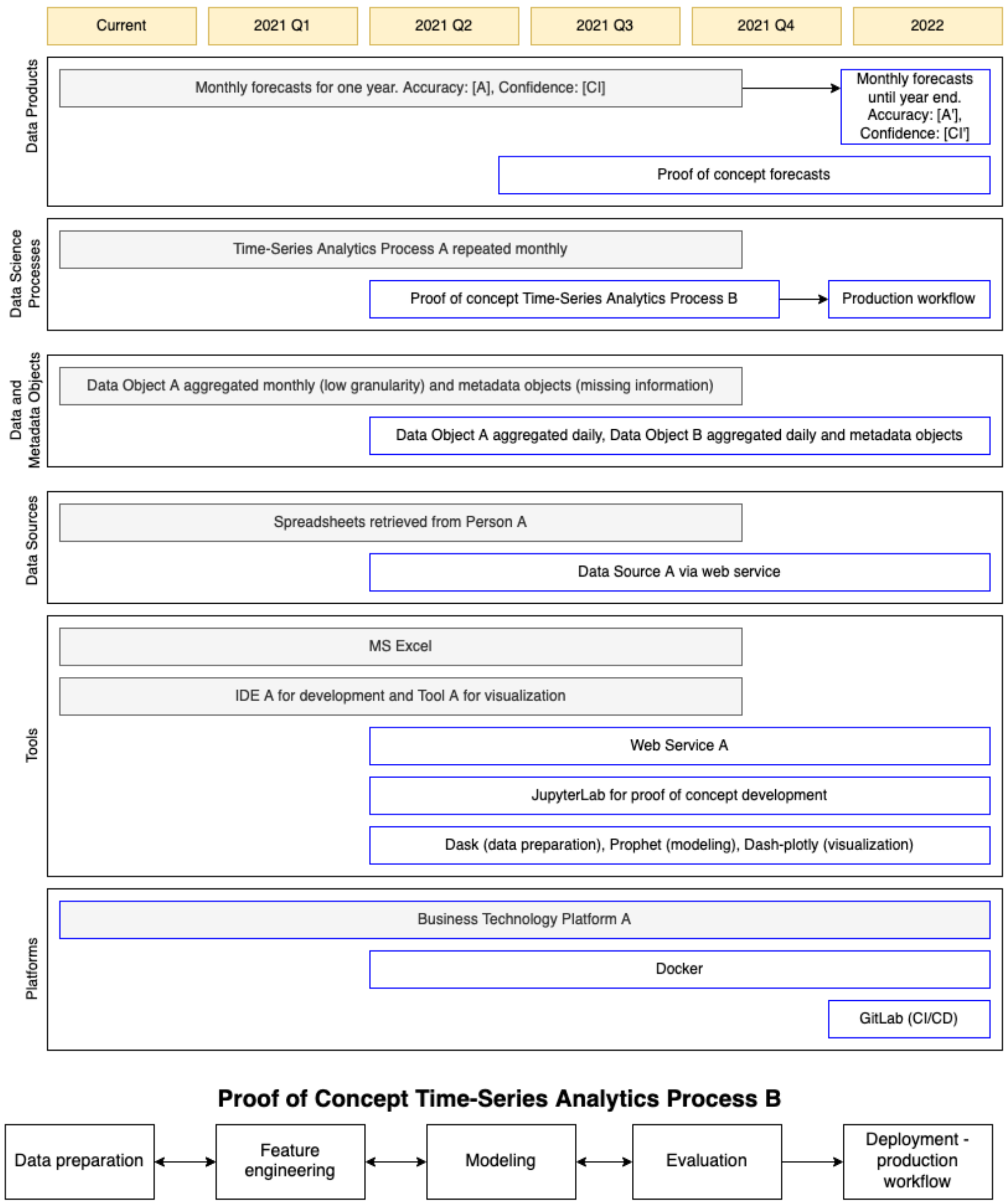

Figure 19: Deep-dive into a time-series analytics study. Specific details are hidden, respecting the group's privacy.

6.4.3. *Results of the Second Iteration*

The workshops produced many outputs, such as meeting records, journals, and Kanban boards, but not the roadmap. The post-workshop team carefully analyzed the outputs, synthesizing the roadmap for the review meeting. Several iterations with stakeholders were necessary before the final review. Since the roadmap was information-heavy, they also discussed how to present the roadmap to different stakeholders. Finally, they color-coded three strategic verticals and agreed on stories to tell at the review meeting. Figures 20-22 illustrate the final roadmap.

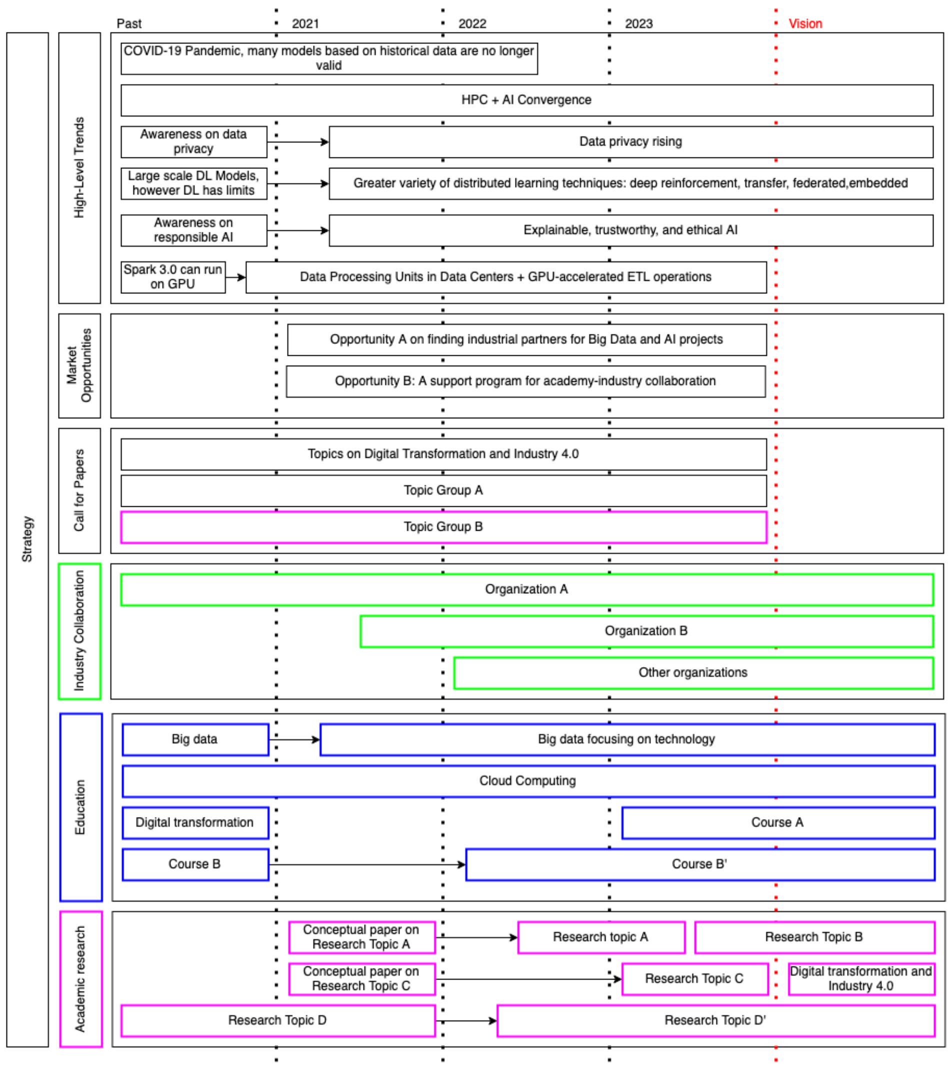

Figure 20: The final roadmap developed by the research group. Some details are aggregated and hidden, respecting the group's privacy.

The review committee met to decide whether the roadmap was ready for publication. We asked the committee to reflect on the objectives, actual outcomes, successes, and impact like the pilot study. The participants answered:

- Even though the academic research layer's landmarks may be subject to frequent changes, this is a good picture of what is possible.
- Until now, we planned all the layers in an ad hoc manner. In this initiative, we could collect all of them together and see the relationships.
- Roadmapping enabled group mind-mapping. The group members could align their perspectives together and reach a consensus. From now on, we can communicate in a more structured manner.
- We already achieved some of the objectives discussed at the kickoff. We also addressed the remaining objectives of this initiative, and they look promising. Nevertheless, these objectives are possible if we can keep the roadmap alive.

We also asked them what they did not get. In this initiative, the group could not explore all the research paths corresponding to the high-level trends. They had put too many targets ambitiously, and now they could see more realistically. Some of the items they put on the roadmap require more research, time, and effort. They could not involve all the students and partners, so there could be misalignments on that front. In summary, the group agreed to keep the roadmap alive and assigned the roadmap champion responsible for keeping the roadmap alive. Only successive iterations can enable the group to reach some objectives and share the outcomes with partners and students. They determined the sprint length to be at most six months or upon request by the roadmap champion, and some open issues were settled for the next iteration.

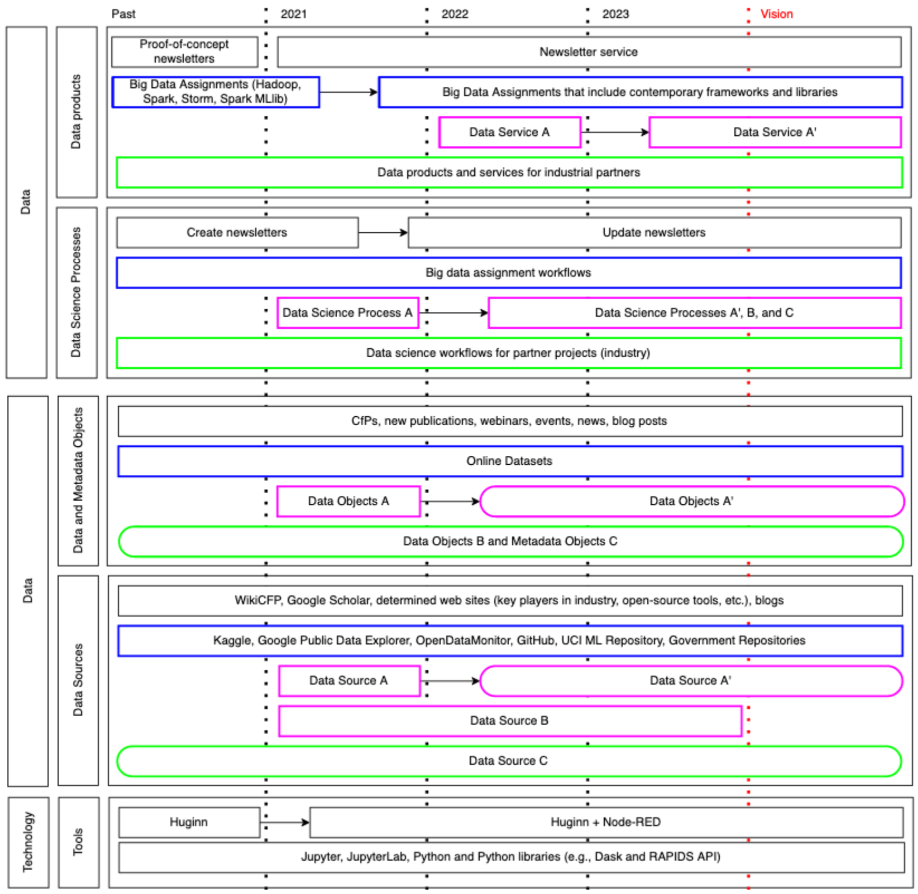

Figure 21: The final roadmap developed by the research group (cont.). Some details are aggregated and hidden, respecting the group's privacy.

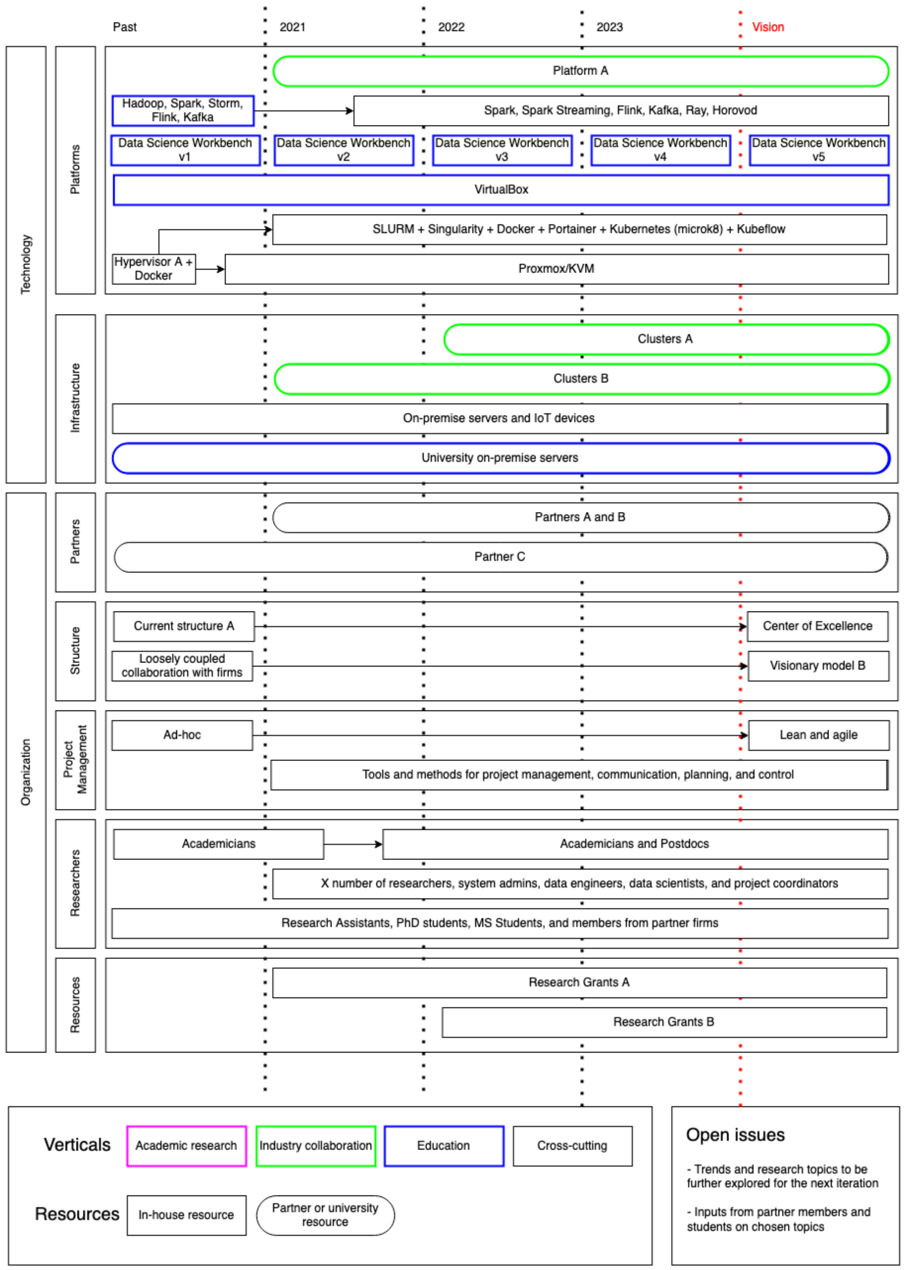

Figure 22: The final roadmap developed by the research group (cont.). Some details are aggregated and hidden, respecting the group's privacy.

6.5. Third iteration: Producing a First-Pass Roadmap for the Oil and Gas Company

6.5.1. Kickoff and Planning

The data analytics team at the oil and gas company wanted to employ DSR to create a first-pass roadmap to see what it could do. Therefore, they agreed on adopting an agile approach to plan a small number of projects. This planning would evaluate data science processes, identify gaps and improvement potentials, and align strategy, data, technology, and organization perspectives accordingly. Furthermore, the team could customize and experience the DSR process to prepare for a more extensive iteration.

There has been a negative trend at the sectoral level since the beginning of the COVID-19 pandemic. Digital transformation is happening faster, and technology trends steer organizations to use internal and external data more effectively. The sector players know they need to get more agile, lean, and data-driven to stay competitive, making data- and technology-oriented investments. Sustainability is a significant concern with regulations such as net-zero carbon on the horizon.

The oil and gas company is well aware of the trends and transformations at the sectoral level. It has strategies for reaching net-zero carbon, improving process efficiencies, and minimizing waste. However, the analytics team has experienced challenges due to communication difficulties and lack of consensus between siloed business units. The two data analytics centers in the university technoparks are relatively new with a narrow area of influence. Table 26 shows the agreed objectives and expected outcomes at the kickoff considering the motivation to deploy DSR and the market-business trends and drivers. The kickoff team also determined the scope (Table 27) using the corresponding template (Table 21).

Table 26: Objectives and expected outcomes for the third iteration

Objectives	**Expected Outcomes**
• To determine short-, mid-, and long-term project portfolios strategically. • To connect data science processes and outcomes to business strategy. • To understand the technological and organizational resource requirements and plan how to bridge the gaps.	• The first-pass data science roadmap • Systematic approaches for collaborating with partner organizations • Elimination of information asymmetries and consensus between stakeholders

• To enable sustainability and monitoring of data-related managerial and operational projects. • To design systems to motivate people in the organization and lead cultural change. • To decrease the turnover rate. • To enhance relationships with partner organizations and better understand how they bridge the gaps. • To move data science and related technology projects from pilot to production.	• Better communication and collaboration between business units

Table 27: Scope of the third iteration

Data	**Organization**	**Technology**	**Strategy**
Data science processes Data and metadata objects Data sources Data governance	Skills and HR Organizational structure and culture Finance Project management	Technology architecture Infrastructure management Technology governance	Business trends and drivers Market trends and drivers

The planning meeting started with assigning the roadmapping coordinators who will facilitate the workshops in this iteration. The participants decided on the timeframes and market-data-technology-organization layers that constitute the roadmap architecture. Next, the planning team discussed internal and external data sources for the pre-workshop and data workshop. After assigning the roles for the roadmapping board, they added the following to the tools to support this initiative (in addition to the tools depicted in Section 6.3.1):

- **Jupyter Notebook**[9]: A web-based interactive, flexible notebook-like development environment for data science. The team utilized Jupyter Notebook for data analysis.

- **Figjam**[10]: An online whiteboard for teams to brainstorm together. Figjam provided the sticky-note-like capability during the workshops in this iteration.

6.5.2. Pre-Workshop and Workshops

The objective of the pre-workshop was pre-populating the roadmap landscape to make the workshops effective and to discover unanticipated opportunities. In the beginning, the roadmapping board agreed on planning two analytics projects in the scope of this iteration to demonstrate the agile capabilities of the roadmapping process: Projects A and B. However, during the pre-workshop, the corresponding team faced difficulties transferring data related to Project B. They also discovered short-term plans and implementations related to the big data system driven by Project A. The pre-workshop team decided to keep the scope limited to Project A, the big data system, and Project A-related integrations from data sources to the big data system. Then they pre-populated the roadmap landscape relying on iterative data analysis and strategic, data-related, technological, and organizational discussions with corresponding stakeholders.

The teams gathered for four workshops as planned. In each workshop, the participants reviewed the landscape and landmarks and put new landmarks as necessary. If deep-dive was necessary for the landmarks, small teams did so until the next workshop and presented their cases before the next workshop began. The online nature of the workshops enabled this approach. Sectoral and business trends and drivers were discussed at the strategy workshop. A deep-dive was particularly necessary for the data workshop for creating the data science process models and a high-level data management plan to determine requirements for the next technology workshop. The technology workshop team debated the current test big data system environment, the required technology categories, and how to bridge the gaps according to data-level requirements, considering the data science trends at the strategy (market) level. The last organization workshop discussions comprised the organizational structure and culture, related business units and partners, and the distinct methodologies for working with different business units.

6.5.3. Results of the Third Iteration

[9] https://jupyter.org/

[10] https://www.figma.com/figjam/

As was the case with the first two iterations, the workshops produced lots of data in terms of logs, sticky notes, and recordings, but not the roadmap. Several iterations were necessary at the post-workshop stage with related stakeholders to produce the final roadmap. The four main paths or stories aligned strategy, data, technology, and organization layers: HPC-High-Performance Data Analytics (HPDA)-AI convergence, becoming lean, infrastructure, and analytics (Figure 25). These paths were not separate from each other but rather complementary. They facilitated the build-measure-learn cycles in the post-workshop stage with distributed stakeholders with limited time. We used these stories as assumptions and tested them quickly without long meetings. Several open issues that required more research, effort, and stakeholder involvement, were agreed on and included in the final version of the roadmap. Figures 23-25 illustrate the final roadmap.

The participants discussed whether to publish the roadmap with minor revisions in the final review meeting. We asked the company about the benefits, outcomes, and impact following the roadmap presentation. They answered:

- This roadmap helps us understand the current situation, long-term goals, and team directions (as a newcomer to the team).
- The roadmap illustrates the whole picture by considering the transformation end-to-end. A complex problem becomes manageable and monitorable. It is like opening the headlights while driving in the dark.
- It adequately shows what we need to do to bridge the gaps. We want to paint the whole picture by putting the rest of the projects in the roadmap.
- We believe roadmapping would boost other business units as they see the complete image and which puzzle piece they own.
- Up until now, we had projects started with specific goals in mind. We had not thought about how these goals, resources, and data services connect strategically. We will get buy-in from the digital transformation leader in the next iteration.

We also wanted the participants to debate how to communicate, transfer, and sustain the benefits of the roadmapping process.

- The roadmap is highly dynamic, which would increase the amount of effort required. The roadmapping methodology we followed can cope with this level of dynamism. Still, we need to consider the subsequent iterations carefully.
- The top management should not consider the items in this roadmap as strict deadlines and goals. If the plans do not meet reality, instead of focusing on the problem, we should do a retrospective and evaluate the lessons learned, what went wrong, and how to best update the roadmap.
- We may need to divide some of the packages into manageable chunks. We discussed many stakeholders connected to items in multiple layers. Instead of involving everyone at once, we can take the baby steps, pick the critical paths,

> and start working with one or two business units. We can use the roadmap to justify why we currently work with a particular business unit to the top management.

6.6. Discussion

We collaborated with two organizations in three iterations to refine the process models and validate DSR's applicability and usefulness. Although the complexity and scope of the final roadmap increased through iterations, all three DSR initiatives mostly met the expectations in terms of objectives and outcomes. We reflected revisions to the process models according to feedback during applications following the action research design (Table 28). Most revisions correspond to the first two iterations.

The two goals of DSR are to generate consensus among all stakeholders, including top management, data experts, domain experts, and IT experts, and back up practitioners with quantitative evidence, providing more confidence around decisions. Therefore, the roadmapping process customizes the workshop-based T-Plan and integrates the CRISP-DM, which results in a hybrid roadmap development methodology. The framework enables organizations to strategically plan their organization-wide data science initiatives and data assets to become data-driven. The roadmap development method is also data-driven. There are also opportunities to support the framework further, taking advantage of large textual databases (Geum et al., 2015; Jeong & Yoon, 2015), decision-making approaches (Daim et al., 2018; Daim & Oliver, 2008), and other management tools (J. H. Lee et al., 2013). We intend to explore integrating these supporting mechanisms with DSR and other data-integrated roadmap types (Han & Geum, 2020) as part of future work.

This study has implications for practice and research. It portrays that becoming a data-driven organization requires overcoming DOTS challenges and connecting business strategies with data-related, technological, and organizational resources. Our prior experience has shown a challenging and long journey, especially for data-disadvantaged and underachiever organizations. The technology architecture and organizational structure must evolve to satisfy the data science and data management plan requirements. Technology forecasting becomes particularly important in this context since many data science tools and application frameworks continue to emerge. It is costly to change the underlying technologies while operations persist. High-level strategic trends are also subject to rapid change.

The research group identified these trends during DSR application (Figure 20) and planned the data layers accordingly. The technologies in Figures 21 and 22 correspond to the high-level strategic trends and data science process requirements. Similarly, the oil and gas company planned the technology layers (Figures 24 and 25) according to strategy-level data science trends and data-level requirements (Figure 23). Nonetheless, the development of DSR also reveals new research directions. The framework can be supported further, particularly for technological and organizational planning, with quantitative assessment and forecasting tools integrated into the methodology. These quantitative approaches need to be evaluated and customized to plan technology and organizational layers for many IT and organizational contexts.

Furthermore, organizational restructuring is necessary but insufficient to become a data-driven organization. Such a transformation also involves significant social change for organizations to create a data-driven culture. The success of the transformation depends on data literacy, leadership skills, trust-oriented decisions, openness, and the continuous learning capacity of stakeholders. The data-driven transformation phenomenon affects the entire organization and necessitates data-driven and trust-oriented leadership. By applying DSR, the research group recognized the academic and industrial high-level trends (explainable, trustworthy, and ethical AI) on the roadmap and planned the layers below accordingly (Figures 20-25). DSR adopts a workshop-based process to catalyze the required social change by establishing communication channels between business functions and mitigating information asymmetries. It should also be emphasized that during the application, both the research group and the oil and gas company realized that they could not explore all the paths that correspond to the high-level trends even though the workshop processes were agile. Achieving some objectives and sharing outcomes with students and partners required the roadmap to be kept alive. Therefore, both organizations published open issues together with the roadmap (Figures 22 and 25) to be resolved until the next iteration, agreed on the schedule to review the roadmap, and assigned the roadmap champion responsible for keeping the roadmap alive. All three iterations recognized that the first developed roadmap is like a minimum viable product (Ries, 2011), promoting a lean approach and minimizing waste during workshops. We argue that becoming a data-driven organization similarly necessitates lean and agile approaches incorporated into the organizational processes, as the research group and the oil and gas company recognized in their final roadmaps.

Table 28: Feedback on the processes and agreed major revisions to the process models through iterations

Process – Step	**Feedback on the process**	**Agreed major revisions to the process model**
Pre-workshop – Request data for workshops	"At the beginning of the pre-workshop, the data experts and the domain experts need to determine data sources. That would allow the data experts to	Data experts from the pre-workshop team and data owners determine data sources at the

	define requirements concerning the IT environment for the workshops."	beginning of the pre-workshop in a separate pool (Figure 13).
Strategy workshop – Review landscape	"It would be possible to prepare beforehand if we saw the pre-populated roadmap landscape before this workshop began."	Add the last step for the pre-workshop: Mail pre-populated landscape to the strategy workshop team (Figure 13).
Strategy and data workshops – Synthesize cases on landscape	"We can leverage an agile methodology adding iterations to the workshops after synthesizing cases on the landscape."	After synthesizing cases on the landscape, add a possible iteration path to small group deep-dives. Turn workshops into two short big group meetings with small group deep-dives in between (Figures 14 and 15).
Data workshop –Review landscape	"The pre-workshop process would be more thorough if we discussed some data sources earlier."	In the planning process, add the following step before determining stakeholders: discuss internal and external data sources.
Data workshop - Select landmarks to deep-dive	"We need to evaluate and plan data and metadata objects independent of the data source since the data source is subject to change."	In pre-workshop and data workshop, treat data sources and data and metadata objects differently (Figures 13 and 15).
Technology workshop – Review landscape	"It is hard to understand technology requirements without a data management plan, which could be discussed earlier."	Add a pathway to data workshop deep-dive activity after making sense of data: Outline high-level data management plan (Figure 15).
Review the roadmap – Document comments	"It was not possible to explore all the high-level strategic trends. Some of the items on the roadmap require more research and effort. We can do so until the next update on the roadmap."	In the review meeting, add a last activity for the review committee: determine when to review and update the roadmap (the sprint length). The review committee should also document the open issues published together with the roadmap (Figures 22 and 25).

Review the roadmap – Document lessons learned	"It would be nice to run a retrospective session as part of the review process after publishing the final roadmap."	After the review committee meeting, the planning team publishes the roadmap, runs a retrospective session, documents lessons learned, and updates the lessons learned database.

6.7. Evaluation of the Action Research's Quality

Compared with other research approaches, action research is a more imprecise, uncertain, and unstable activity (Coughlan & Coghlan, 2002). Coghlan and Brannick (2009) discuss, the main characteristics of good action research are; the researcher intends to change the organization, the project has implications beyond those involved in it, and the project aims to develop a theory as well as it is beneficial for the organization. Zuber-Skerritt and Fletcher (2007) add the following to quality and rigor criteria:

- The design, explanation, and justification of the methodology
- The individuality and originality of contributions
- The use of relevant literature that justifies the candidate's choice
- The clearness and soundness of writing style

This study starts with an actual need from the industry. A conceptual framework is developed synthesizing the systematic literature review results (reported in Chapter 4), well-established TRM, big data, data science, and data-driven organization literature. The resulting DSR framework comprises a complete methodology with macro and micro-process models. The graphical representations of business processes with BPMN make it easier to iterate on the process models. The action research approach refines these process models, validating DSR's applicability and usefulness through three iterations with two organizations. Nevertheless, the action research projects have significant practical implications beyond the two organizations since they aim to transform both organizations to become data-driven while developing theory.

6.8. Limitations and Future Research

Roadmapping applications require customization for a specific organizational context and purpose. Therefore, one limitation of this study is that the methodology and application context may not be directly applicable to other data-disadvantaged or underachiever organizational settings. For example, for public organizations, data might be more sensitive, and organization structure and culture might be more rigid. Accordingly, the architecture and process models may require minor customizations regarding the organizational context. While both organizations want to keep the developed data science roadmap alive, this study does not explore the integration of DSR into business processes. Therefore, this study sets forth the need for roadmapping to facilitate data-driven transformations in organizations. However, we will explore the generalizability and adoption of DSR. All things considered, further research directions are as follows:

1. Follow up on the next iteration with the oil and gas company and see how open issues are resolved.

2. Apply DSR in businesses from other data-disadvantaged and underachiever sectors.

3. Explore the planning of data and respective technology layers in other data-integrated roadmap types.

4. Research technology assessment and forecasting tools that complement data-integrated roadmap development.

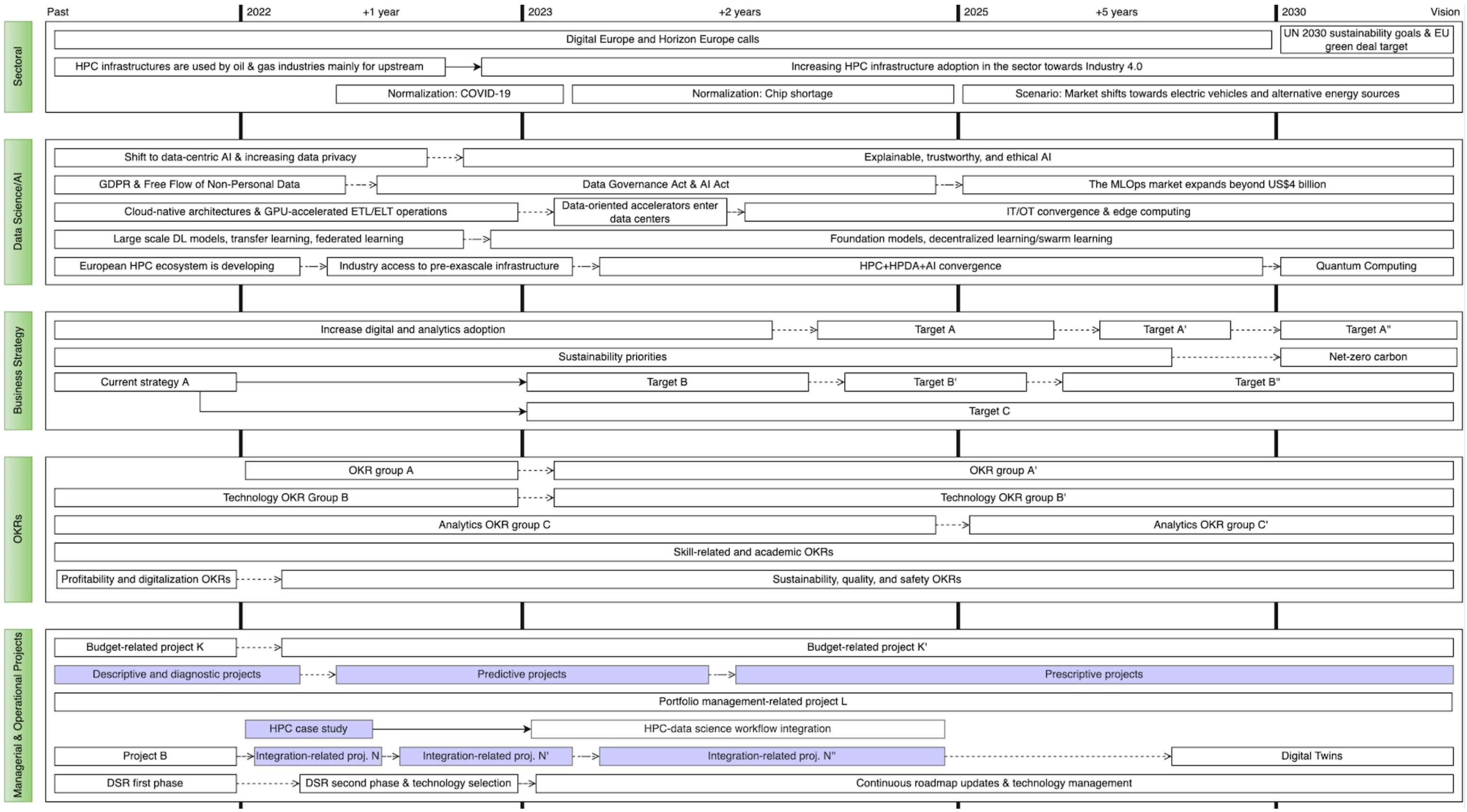

Figure 23: The final roadmap developed by the oil and gas company. Some details are aggregated and hidden, respecting company privacy.

Figure 24: The final roadmap developed by the oil and gas company (cont.). Some details are aggregated and hidden, respecting company privacy.

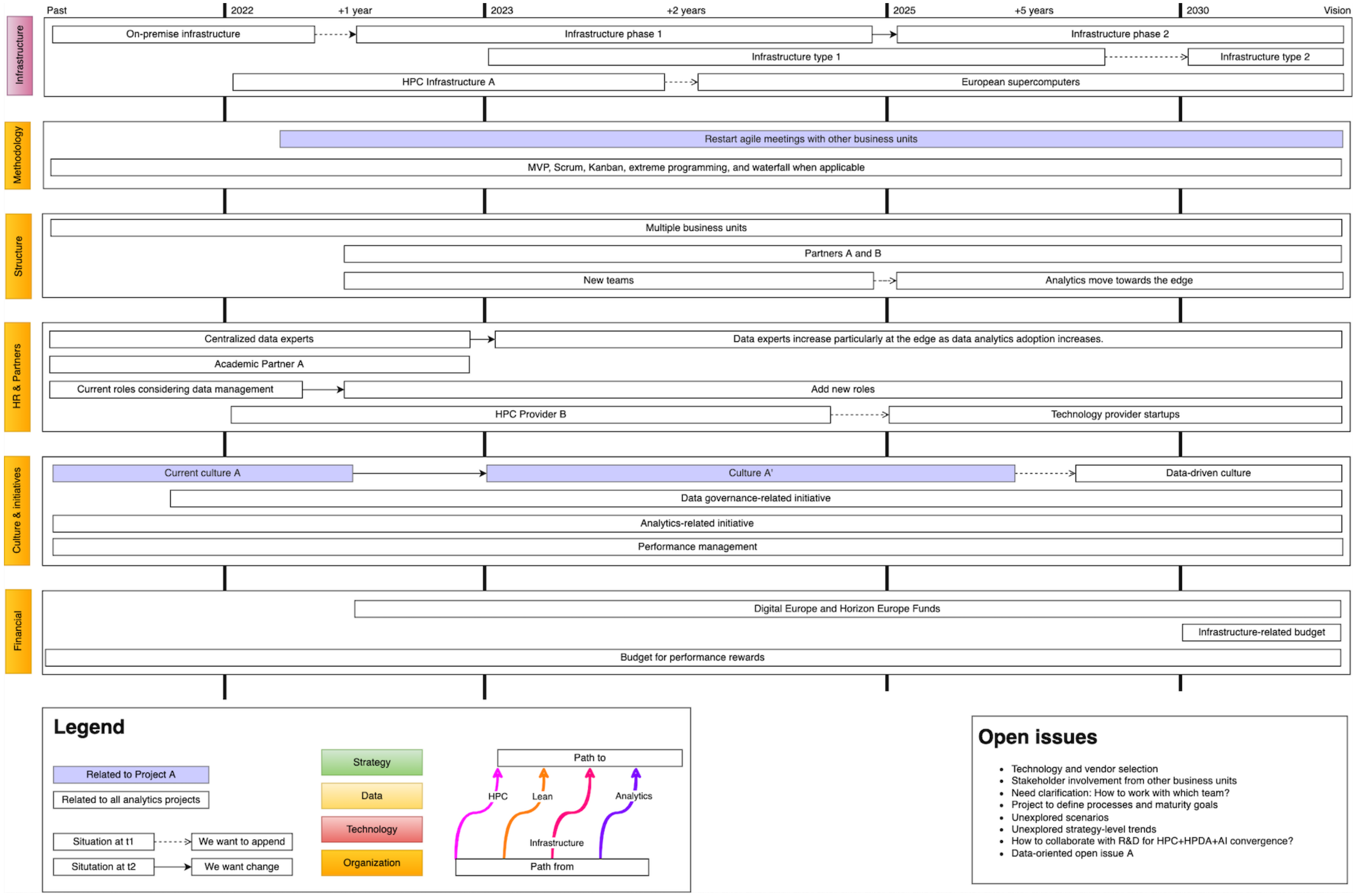

Figure 25: The final roadmap developed by the oil and gas company (cont.). Some details are aggregated and hidden, respecting company privacy.

CHAPTER 7

CONCLUSION

7.1. Summary

While organizations had traditionally produced data with careful planning, data has become a strategic asset for contemporary organizations because of the paradigm shifts, such as digital transformation, that produce big data without planning. Businesses have leveraged data science to gain a competitive edge in the marketplace by extracting knowledge from big data. However, not all sectors are well-suited to big data, and many industries, such as manufacturing and healthcare, have dragged behind. They still strive to become data-driven by overcoming data-related, organizational, technological, and strategic challenges.

Organizations from the not-so-well-suited sectors need to plan organization-wide data assets and data science projects to leverage data science in their processes effectively. They need a comprehensive approach to align their business strategy with data-related, technological, and organizational resources. This study begins after recognizing the industrial necessity for such planning. Literature review reveals the lack of studies that comprehensively identify the challenges for data-driven transforming organizations from the data science perspective. Furthermore, while prebuilt roadmaps in the literature acknowledge that reaching analytics goals depends on strategy, skills, and culture, these roadmaps disregard the iterative nature of roadmaps. In successive iterations, a successful approach should build trust and consensus between stakeholders and business functions.

Accordingly, this study develops the DOTS research framework using a systematic approach to identify the factors related to data science usage in organizations and the challenges that organizations face on their journey to become data-driven. Then it puts forth the need for roadmapping to facilitate data-driven organizational transformation. With contributions from industry and academia, we have developed the DSR framework by customizing TRM to help organizations connect business strategies with data-related, technological, and organizational resources. The proposed roadmapping framework provides a methodology that synthesizes the DOTS research framework with two industry best practices: CRISP-DM and the workshop-based T-Plan.

The framework has been refined and validated in three consecutive applications in real-world organizational settings, collaborating with a cross-disciplinary research group and an oil and gas company. The results show that the proposed framework facilitates DSR initiatives by creating a comprehensive roadmap capturing strategy, data, technology, and organizational perspectives. However, becoming data-driven also necessitates significant social change toward openness and trust. The DSR initiative facilitates this social change by opening communication channels, aligning perspectives, and generating consensus among stakeholders.

7.2. Contributions

This study contributes to the literature by bridging multiple research gaps. First, it systematically identifies the challenges that data-driven transforming organizations face and organizes them into the DOTS research framework. Second, the development of DSR is the first study that focuses on digitally transforming data-disadvantaged or underachiever organizations to plan organization-wide data science endeavors and data assets, aligning business strategy with data-related, technological, and organizational resources. Third, we provide extensible and detailed process models (Figures 10-17) for agile planning of the data layer, segmented to account for data sources, data and metadata objects, data science processes, and data products and services (Figure 8). We provide practitioners with specifications and templates (Tables 21, 22, 23, and 24) to make it easier to initiate roadmapping initiatives. Fourth, the development of DSR presents novel applications for the hybrid agile roadmapping methodology that output data-integrated roadmaps. Finally, to the best of our knowledge, this is the first roadmapping study that uses BPMN to delineate micro-level roadmapping activities via visual process models. BPMN has made getting feedback during the development process and applying the agreed revisions (Table 28) more manageable. Additionally, more systematic method development has been enabled, as researchers and practitioners can see all the comments and revisions of the process models. We also believe that the graphical representation of process models can aid practitioners and researchers in customizing and executing roadmapping initiatives as a valuable addition to checklist-based templates (Kerr & Phaal, 2019)

7.3. Limitations and Future Work

The following limitations are identified regarding this study: (1) The DOTS research framework may lack some domain-specific factors, not published in the literature, or cannot be detected by the systematic method. (2) Any roadmapping application requires customization to a specific organizational context and purpose. Consequently, the methodology and the presented application contexts may not immediately apply to other settings, such as the public sector, where the business builds data science capabilities. (3) The follow-up on the successive iterations with the research group and the oil and gas company on roadmapping is not observed. We do not investigate mechanisms that drive and sustain the adoption of DSR. (4) This research does not explore the technology assessment and forecasting tools that complement DSR.

This study sets forth the need for roadmapping to help organizations become data-driven and provides practitioners a framework to initiate roadmapping for data science projects. The following opportunities are identified corresponding to the limitations of the study:

1. Adoption of DSR after the production of the first-pass roadmap

2. Generalizability of DSR by application in organizations from different sectors

3. Development of data and technology layers in various data-integrated roadmaps

4. Development of technology assessment and forecasting tools that complement data-integrated roadmapping

REFERENCES

Acharya, A., Singh, S. K., Pereira, V., & Singh, P. (2018). Big data, knowledge co-creation and decision making in fashion industry. *International Journal of Information Management*, *42*, 90–101. https://doi.org/10.1016/j.ijinfomgt.2018.06.008

Adrian, C., Abdullah, R., Atan, R., & Jusoh, Y. Y. (2017). Factors influencing to the implementation success of big data analytics: A systematic literature review. *International Conference on Research and Innovation in Information Systems, ICRIIS*, 1–6. https://doi.org/10.1109/ICRIIS.2017.8002536

Ahmad, A., Ahmad, R., & Hashim, K. F. (2016). Innovation traits for business intelligence successful deployment. *Journal of Theoretical and Applied Information Technology*, *89*(1), 78–89.

Aho, A. M., & Uden, L. (2014). Developer Toolchains for Large-Scale Analytics: Two Case Studies. *Lecture Notes in Business Information Processing*, *185 LNBIP*, 99–107. https://doi.org/10.1007/978-3-319-08618-7_10

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. https://doi.org/10.1016/j.ijpe.2016.08.018

Alade, A. T. (2017). Business intelligence tools for informed decision-making: An overview. *Strategic Engineering for Cloud Computing and Big Data Analytics*, 207–223. https://doi.org/10.1007/978-3-319-52491-7_12

Al-Ali, A. G., & Phaal, R. (2019). Design Sprints for Roadmapping an Agile Digital Transformation. *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 1–6. https://doi.org/10.1109/ICE.2019.8792597

Aldea, A., Iacob, M.-E., Wombacher, A., Hiralal, M., & Franck, T. (2018). Enterprise Architecture 4.0 – A Vision, an Approach and Software Tool Support. *2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC)*, 1–10. https://doi.org/10.1109/EDOC.2018.00011

Alexander, D., & Lyytinen, K. (2017). Organizing successfully for big data to transform organizations. *Americas Conference on Information Systems (AMCIS)*.

Al-Hakimi, A. A. A. (2017). Big data skills required for successful application implementation in the banking sector. *Communications in Computer and Information Science*, *752*, 381–392. https://doi.org/10.1007/978-981-10-6502-6_34

Al-Hitmi, M., & Sherif, K. (2018). Employee Perceptions of Fairness towards IoT Monitoring. *VINE Journal of Information and Knowledge Management Systems*, 00–00. https://doi.org/10.1108/VJIKMS-01-2018-0007

Ali, O., Crvenkovski, P., & Johnson, H. (2016a). Using a business intelligence data analytics solution in healthcare. *7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEEE IEMCON 2016*. https://doi.org/10.1109/IEMCON.2016.7746328

Ali, O., Crvenkovski, P., & Johnson, H. (2016b). Using a business intelligence data analytics solution in healthcare. *7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEEE IEMCON 2016*. https://doi.org/10.1109/IEMCON.2016.7746328

Altarturi, H. H., Ng, K. Y., Ninggal, M. I. H., Nazri, A. S. A., & Ghani, A. A. A. (2018). A requirement engineering model for big data software. *2017 IEEE Conference on Big Data and Analytics, ICBDA 2017*, *2018-Janua*, 111–117. https://doi.org/10.1109/ICBDAA.2017.8284116

Altrichter, H., Kemmis, S., McTaggart, R., & Zuber-Skerritt, O. (2002). The concept of action research. *The Learning Organization*, *9*(3), 125–131. https://doi.org/10.1108/09696470210428840

Amankwah-Amoah, J., & Adomako, S. (2019). Big data analytics and business failures in data-Rich environments: An organizing framework. *Computers in Industry*, *105*, 204–212. https://doi.org/10.1016/j.compind.2018.12.015

Anderson, C. (2015). *Creating a data-driven organization: Practical advice from the trenches*. O'Reilly Media, Inc.

Angelopoulos, S., McAuley, D., Merali, Y., Mortier, R., & Price, D. (2016). Small Fish in a Big Pond: An Architectural Approach to Users Privacy, Rights and Security in the Age of Big Data. *ICIS 2016 Proceedings*, 1–18.

Ardagna, C. A., Ceravolo, P., & Damiani, E. (2016a). Big data analytics as-a-service: Issues and challenges. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 3638–3644. https://doi.org/10.1109/BigData.2016.7841029

Ardagna, C. A., Ceravolo, P., & Damiani, E. (2016b). Big data analytics as-a-service: Issues and challenges. *2016 IEEE International Conference on Big Data (Big Data)*, 3638–3644. https://doi.org/10.1109/BigData.2016.7841029

Arora, D., & Malik, P. (2015). Analytics: Key to go from generating big data to deriving business value. *Proceedings - 2015 IEEE 1st International Conference on Big Data Computing Service and Applications, BigDataService 2015*, 446–452. https://doi.org/10.1109/BigDataService.2015.62

Ashraf, S., & Khan, S. A. (2015). Visualizations-based analysis of Telco data for business intelligence. *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, *2015-Novem*, 242–246. https://doi.org/10.1109/ICSESS.2015.7339046

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79–80*, 3–15. https://doi.org/10.1016/j.jpdc.2014.08.003

Atyeh, A. J., Jaradat, M.-I. R. M., & Arabeyyat, O. S. (2017). Big Data Analytics Evaluation, Selection and Adoption: A Developing Country Perspective. *International Journal of Computer Science and Network Security*, *17*(9), 159–171.

Aydiner, A. S., Tatoglu, E., Bayraktar, E., Zaim, S., & Delen, D. (2019). Business analytics and firm performance: The mediating role of business process performance. *Journal of Business Research*, *96*, 228–237. https://doi.org/10.1016/j.jbusres.2018.11.028

Azevedo, A. I. R. L., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: A parallel overview. *IADS-DM*.

Baars, H., & Ereth, J. (2016). *From data warehouses to analytical atoms—The internet of things as a centrifugal force in business intelligence and analytics*.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational Issues of Big Data and Analytics in Networked Business. *MIS Quarterly*, *40*(4), 807–818. https://doi.org/10.25300/MISQ/2016/40:4.03

Banda, M., & Ngassam, E. K. (2017). A data management and analytic model for business intelligence applications. *2017 IST-Africa Week Conference, IST-Africa 2017*. https://doi.org/10.23919/ISTAFRICA.2017.8102350

Barahona, I., Lopez-Corona, O., & Riba, A. (2016). The level of adoption of statistical tools. A case of statistical engineering. *Proceedings of the 12th International Conference on Industrial Engineering, ICIE 2016*, 27–35. https://doi.org/10.1109/INDUSENG.2016.7519345

Baškarada, S., & Koronios, A. (2017). Unicorn data scientist: The rarest of breeds. *Program*, *51*(1), 65–74. https://doi.org/10.1108/PROG-07-2016-0053

Batra, D. (2017). Adapting Agile Practices for Data Warehousing, Business Intelligence, and Analytics. *Journal of Database Management (JDM)*, *28*(4), 1–23. https://doi.org/10.4018/JDM.2017100101

Batra, D. (2018). Agile values or plan-driven aspects: Which factor contributes more toward the success of data warehousing, business intelligence, and analytics project development? *Journal of Systems and Software*, *146*, 249–262. https://doi.org/10.1016/j.jss.2018.09.081

Bayamlıoğlu, E., & Leenes, R. (2018). The 'rule of law' implications of data-driven decision-making: A techno-regulatory perspective. *Law, Innovation and Technology*, *10*(2), 295–313. https://doi.org/10.1080/17579961.2018.1527475

Bedeley, R. T., Ghoshal, T., Iyer, L. S., & Bhadury, J. (2018). Business Analytics and Organizational Value Chains: A Relational Mapping. *Journal of Computer Information Systems*, *58*(2), 151–161. https://doi.org/10.1080/08874417.2016.1220238

Beheshti, A., Benatallah, B., & Motahari-Nezhad, H. R. (2018). ProcessAtlas: A scalable and extensible platform for business process analytics. *Software - Practice and Experience*, *48*(4), 842–866. https://doi.org/10.1002/spe.2558

Bekmamedova, N., & Shanks, G. (2014). Social media analytics and business value: A theoretical framework and case study. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3728–3737. https://doi.org/10.1109/HICSS.2014.464

Belfo, F. P., & Andreica, A. B. (2018). *A Comprehensive Methodology to Implement Business Intelligence and Analytics Through Knowledge Discovery in Databases*. 102–111. https://doi.org/10.1007/978-3-030-05918-7_10

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial Intelligence in Organizations: Current State and Future Opportunities. *MIS Quarterly Executive*, *19*(4).

Berntsson Svensson, R., & Taghavianfar, M. (2020). Toward Becoming a Data-Driven Organization: Challenges and Benefits. In F. Dalpiaz, J. Zdravkovic, & P. Loucopoulos (Eds.), *Research Challenges in Information Science* (pp. 3–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-50316-1_1

Bhattacharya, I., Godbole, S., Gupta, A., Verma, A., Achtermann, J., & English, K. (2009). Enabling analysts in managed services for CRM analytics. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '09*, 1077. https://doi.org/10.1145/1557019.1557136

Bhimani, A., & Willcocks, L. (2014). Digitisation, Big Data and the transformation of accounting information. *Accounting and Business Research*, *44*(4), 469–490. https://doi.org/10.1080/00014788.2014.910051

Bi, W., Cai, M., Liu, M., & Li, G. (2016). A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn. *IEEE Transactions on Industrial Informatics*, *12*(3), 1270–1281. https://doi.org/10.1109/TII.2016.2547584

Bihl, T. J., Young II, W. A., & Weckman, G. R. (2016). Defining, Understanding, and Addressing Big Data. *International Journal of Business Analytics*, *3*(2), 1–32. https://doi.org/10.4018/IJBAN.2016040101

Blake, C., & Gabb, H. A. (2014). Parameter tuning: Exposing the gap between data curation and effective data analytics. *Proceedings of the ASIST Annual Meeting*, *51*(1). https://doi.org/10.1002/meet.2014.14505101138

Božič, K., & Dimovski, V. (2019). Business intelligence and analytics for value creation: The role of absorptive capacity. *International Journal of Information Management*, *46*, 93–103. https://doi.org/10.1016/j.ijinfomgt.2018.11.020

Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, *70*, 328–337. https://doi.org/10.1016/j.jbusres.2016.08.006

Brennan, R., Attard, J., & Helfert, M. (2018). Management of data value chains, a value monitoring capability maturity model. *ICEIS 2018 - Proceedings of the 20th International Conference on Enterprise Information Systems*, *2*, 573–584. https://doi.org/10.1016/S0141-1136(00)00089-1

Brichni, M., & Guedria, W. (2018a). *Data Analytics Challenges in Industry 4.0: A Case-Based Approach*. 209–221. https://doi.org/10.1007/978-3-030-02671-4_12

Brichni, M., & Guedria, W. (2018b). *Data Analytics Challenges in Industry 4.0: A Case-Based Approach*. 209–221. https://doi.org/10.1007/978-3-030-02671-4_12

Brock, V., & Khan, H. U. (2017). Big data analytics: Does organizational factor matters impact technology acceptance? *Journal of Big Data*, *4*(1). https://doi.org/10.1186/s40537-017-0081-8

Bumblauskas, D., Nold, H., Bumblauskas, P., & Igou, A. (2017). Big data analytics: Transforming data to action. *Business Process Management Journal*, *23*(3), 703–720. https://doi.org/10.1108/BPMJ-03-2016-0056

Bygstad, B., Øvrelid, E., & Lie, T. (2019). Establishing an analytics capability in a hospital. *IFIP Advances in Information and Communication Technology*, *533*, 3–14. https://doi.org/10.1007/978-3-030-04315-5_1

Caesarius, L. M., & Hohenthal, J. (2018). Searching for big data: How incumbents explore a possible adoption of big data technologies. *Scandinavian Journal of Management*, *34*(2), 129–140. https://doi.org/10.1016/j.scaman.2017.12.002

Cao, G., & Duan, Y. (2017). How do top- and bottom-performing companies differ in using business analytics? *Journal of Enterprise Information Management*, *30*(6), 874–892. https://doi.org/10.1108/JEIM-04-2016-0080

Cao, G., Duan, Y., & Cadden, T. (2019). The link between information processing capability and competitive advantage mediated through decision-making effectiveness. *International Journal of Information Management*, *44*, 121–131. https://doi.org/10.1016/j.ijinfomgt.2018.10.003

Carillo, K. D. A. (2017). Let's stop trying to be "sexy" – preparing managers for the (big) data-driven business era. *Business Process Management Journal*, *23*(3), 598–622. https://doi.org/10.1108/BPMJ-09-2016-0188

Carter, D., & Sholler, D. (2015). Data science on the ground: Hype, criticism, and everyday work. *Journal of the Association for Information Science and Technology*, *67*(10), 2309–2319. https://doi.org/10.1002/asi.23563

Castellanos, A., Castillo, A., Lukyanenko, R., & Tremblay, M. C. (2017). Understanding benefits and limitations of unstructured data collection for repurposing organizational data. *Lecture Notes in Business Information Processing*, *300*, 13–24. https://doi.org/10.1007/978-3-319-66996-0_2

Cegielski, C. G., & Jones-Farmer, L. A. (2016). Knowledge, Skills, and Abilities for Entry-Level Business Analytics Positions: A Multi-Method Study. *Decision Sciences Journal of Innovative Education*, *14*(1), 91–118. https://doi.org/10.1111/dsji.12086

Chaoui, H., & Makdoun, I. (2017). A new secure model for the use of cloud computing in big data analytics. *Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing*, 1–11. https://doi.org/10.1145/3018896.3018913

Chatfield, A. T., Shlemoon, V. N., Redublado, W., Rahman, F., Qiu, R., Cantwell, J., & Laney, D. (2014). Data Scientists as Game Changers in Big Data Environments. *Application Delivery Strategies*, *24*(Trounson), 4. https://doi.org/10.1016/j.infsof.2008.09.005

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *Journal of Management Information Systems*, *32*(4), 4–39. https://doi.org/10.1080/07421222.2015.1138364

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165–1188. JSTOR. https://doi.org/10.2307/41703503

Chen, H. M., Kazman, R., Haziyev, S., Kropov, V., & Chtchourov, D. (2016). Big Data as a Service: A Neo-Metropolis Model Approach for Innovation. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2016-March*, 5458–5467. https://doi.org/10.1109/HICSS.2016.674

Chen, H. M., Schutz, R., Kazman, R., & Matthes, F. (2016). Amazon in the air: Innovating with big data at Lufthansa. *Proceedings of the Annual Hawaii International*

Conference on System Sciences, *2016-March*, 5096–5105. https://doi.org/10.1109/HICSS.2016.631

Chen, H.-M., Kazman, R., & Haziyev, S. (2016). Agile Big Data Analytics for Web-Based Systems: An Architecture-Centric Approach. *IEEE Transactions on Big Data*, *2*(3), 234–248. https://doi.org/10.1109/tbdata.2016.2564982

Chen, Y., Lelescu, A., & Spohrer, J. (2008). Three factors to sustainable service system excellence: A case study of service systems. *Proceedings - 2008 IEEE International Conference on Services Computing, SCC 2008*, *2*, 119–126. https://doi.org/10.1109/SCC.2008.28

Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big Data Analytics in Operations Management. *Production and Operations Management*, *27*(10), 1868–1883. https://doi.org/10.1111/poms.12838

Chong, D., & Shi, H. (2015). Big data analytics: A literature review. *Journal of Management Analytics*, *2*(3), 175–201. https://doi.org/10.1080/23270012.2015.1082449

Chung, P. T., & Chung, S. H. (2013). On data integration and data mining for developing business intelligence. *9th Annual Conference on Long Island Systems, Applications and Technology, LISAT 2013*. https://doi.org/10.1109/LISAT.2013.6578235

Coghlan, D., & Brannick, T. (2009). *Doing Action Research in Your Own Organization* (Third Edition). SAGE Publications Ltd.

Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., & Reis, M. S. (2016). How Can SMEs Benefit from Big Data? Challenges and a Path Forward. *Quality and Reliability Engineering International*, *32*(6), 2151–2164. https://doi.org/10.1002/qre.2008

Comuzzi, M., & Patel, A. (2016). How organisations leverage Big Data: A maturity model. *Industrial Management and Data Systems*, *116*(8), 1468–1492. https://doi.org/10.1108/IMDS-12-2015-0495

Convertino, G., & Echenique, A. (2017). Self-service data preparation and analysis by business users: New needs, skills, and tools. *Sigchi Ea 2017*, *Part F1276*, 1075–1083. https://doi.org/10.1145/3027063.3053359

Coughlan, P., & Coghlan, D. (2002). Action research for operations management. *International Journal of Operations & Production Management*, *22*(2), 220–240. https://doi.org/10.1108/01443570210417515

Daily, J., & Peterson, J. (2017). Predictive Maintenance: How Big Data Analysis Can Improve Maintenance. *Supply Chain Integration Challenges in Commercial Aerospace*, 267–278. https://doi.org/10.1007/978-3-319-46155-7_18

Daim, T. U., & Oliver, T. (2008). Implementing technology roadmap process in the energy services sector: A case study of a government agency. *Technological Forecasting and Social Change*, *75*(5), 687–720. https://doi.org/10.1016/j.techfore.2007.04.006

Daim, T. U., Yoon, B.-S., Lindenberg, J., Grizzi, R., Estep, J., & Oliver, T. (2018). Strategic roadmapping of robotics technologies for the power industry: A multicriteria technology assessment. *Technological Forecasting and Social Change*, *131*, 49–66. https://doi.org/10.1016/j.techfore.2017.06.006

Daradkeh, M. (2019). Visual Analytics Adoption in Business Enterprises: An Integrated Model of Technology Acceptance and Task-Technology Fit. *International Journal of Information Systems in the Service Sector (IJISSS)*, *11*(1), 68–89. https://doi.org/10.4018/IJISSS.2019010105

Davenport, T. H. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Harvard Business Review Press.

Davenport, T. H., & Dyché, J. (2013). *Big data in big companies* (pp. 1–31).

Davenport, T. H., Harris, J., & Abney, D. (2017). *Competing on Analytics: The New Science of Winning; With a New Introduction* (Revised edition). Harvard Business Review Press.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at Work: Smarter Decisions, Better Results*. Harvard Business Review Press.

De Oliveira Cordeiro, G., Deschamps, F., & De Lima, E. P. (2017). Developing a big data/analytics project: A case study in the auto industry. *67th Annual Conference and Expo of the Institute of Industrial Engineers 2017*, 1753–1758.

de Souza, M. L. P., de Souza, W. C., Freitas, J. S., Del Rey de Melo Filho, L., & Bagno, R. B. (2020). Agile Roadmapping: A management Tool for Digital

Entrepreneurship. *IEEE Transactions on Engineering Management*, 1–15. https://doi.org/10.1109/TEM.2020.3027918

Debortoli, S., Sc, M., Müller, O., & Jan, P. (2010). *Comparing Business Intelligence and Big Data Skills A Text Mining Study Using Job Advertisements*. 289–300. https://doi.org/10.1007/s12599-014-0344-2

Del Vecchio, P., Secundo, G., & Passiante, G. (2018). Analyzing Big Data through the lens of customer knowledge management: Evidence from a set of regional tourism experiences. *Kybernetes*, *47*(7), 1348–1362. https://doi.org/10.1108/K-07-2017-0273

Depeige, A., & Doyencourt, D. (2015). Actionable Knowledge As A Service (AKAAS): Leveraging big data analytics in cloud computing environments. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-015-0023-2

Derguech, W., Bruke, E., & Curry, E. (2014). An Autonomic Approach to Real-Time Predictive Analytics Using Open Data and Internet of Things. *2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops*, 204–211. https://doi.org/10.1109/UIC-ATC-ScalCom.2014.137

Dittert, M., Härting, R. C., Reichstein, C., & Bayer, C. (2018). A data analytics framework for business in small and medium-sized organizations. *Smart Innovation, Systems and Technologies*, *73*, 169–181. https://doi.org/10.1007/978-3-319-59424-8_16

Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in big data and business analytics. *Industrial and Commercial Training*, *47*(4), 174–181. https://doi.org/10.1108/ICT-08-2014-0059

Duke, L., & Ashraf, A. (2019). *Aligning Analytics with Marketing Strategy: Using Analytics to Drive Marketing Strategy with New Media Applications*. *1*. https://doi.org/10.1007/978-3-319-93299-6

Dutta, D., & Bose, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. *International Journal of Production Economics*, *165*, 293–306. https://doi.org/10.1016/j.ijpe.2014.12.032

Dwivedi, V., & Kulkarni, N. (2008). Information as a service in a data analytics scenario—A case study. *Proceedings of the IEEE International Conference on Web Services, ICWS 2008*, 615–620. https://doi.org/10.1109/ICWS.2008.119

Easterby-Smith, M., Thorpe, R., & Jackson, P. R. (2015). *Management and business research*. Sage.

Edge, D., Larson, J., & White, C. (2018). Bringing AI to BI: Enabling visual analytics of unstructured data in a modern business intelligence platform. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–9. https://doi.org/10.1145/3170427.3174367

Enders, T. (2018). Exploring the Value of Data – A Research Agenda. *Lecture Notes in Business Information Processing*, *331*, 274–286. https://doi.org/10.1007/978-3-030-00713-3_21

Endert, A., Hossain, M. S., Ramakrishnan, N., North, C., Fiaux, P., & Andrews, C. (2014). The human is the loop: New directions for visual analytics. *Journal of Intelligent Information Systems*, *43*(3), 411–435. https://doi.org/10.1007/s10844-014-0304-9

Engelseth, P., & Wang, H. (2018). Big data and connectivity in long-linked supply chains. *Journal of Business and Industrial Marketing*, *33*(8), 1201–1208. https://doi.org/10.1108/JBIM-07-2017-0168

Ereth, J. (2018). DataOps – Towards a definition. *CEUR Workshop Proceedings*, *2191*, 104–112.

Ereth, J., & Baars, H. (2015). Cloud-Based Business Intelligence and Analytics Applications – Business Value and Feasibility. *PACIS 2015 Proceedings*.

Esswein, M., & Chamoni, P. (2018). *Business Analytics in the Finance Department – A Literature Review*.

Fabijan, A., Dmitriev, P., Olsson, H. H., & Bosch, J. (2017). The Evolution of Continuous Experimentation in Software Product Development: From Data to a Data-Driven Organization at Scale. *Proceedings - 2017 IEEE/ACM 39th International Conference on Software Engineering, ICSE 2017*, 770–780. https://doi.org/10.1109/ICSE.2017.76

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27–34.

Fernando, F., & Engel, T. (2018). Big Data and Business Analytic Concepts: A Literature Review. *AMCIS 2018 Proceedings*, *2011*, 1–10.

Fernando, L. K. W., & Haddela, P. S. (2017). Hybrid framework for master data management. *2017 Seventeenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, 1–7. https://doi.org/10.1109/ICTER.2017.8257785

Flath, C. M., & Stein, N. (2018). Towards a data science toolbox for industrial analytics applications. *Computers in Industry*, *94*, 16–25. https://doi.org/10.1016/j.compind.2017.09.003

Foshay, N., Yeoh, W., Boo, Y. L., Ong, K. L., & Mattie, D. (2015). A comprehensive diagnostic framework for evaluating business intelligence and analytics effectiveness. *Australasian Journal of Information Systems*, *19*, S37–S54. https://doi.org/10.3127/ajis.v19i0.1178

Fosso Wamba, S., 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. https://doi.org/10.1016/j.ijpe.2014.12.031

Fraser, N. (2017). Strategy development and big data analytics. *Internet of Things and Data Analytics Handbook*, 347–364. https://doi.org/10.1002/9781119173601.ch20

Fromm, H., Habryn, F., & Satzger, G. (2012). Service analytics: Leveraging data across enterprise boundaries for competitive advantage. *Globalization of Professional Services: Innovative Strategies, Successful Processes, Inspired Talent Management, and First-Hand Experiences*, 139–149. https://doi.org/10.1007/978-3-642-29181-4_13

Furtado, L., Dutra, M., & MacEdo, D. (2016). Characterizing the value creation in organizations that implement big data environments. *Advances in Transdisciplinary Engineering*, *4*, 915–924. https://doi.org/10.3233/978-1-61499-703-0-915

Garmaki, M., Boughzala, I., Fosso Wamba, S., & Fosso, S. (2016). The Effect of Big Data Analytics Capability on Firm Performance. *Pacific Asia Conference on Information Systems*. https://doi.org/10.1016/j.lrp.2017.02.005

Gear, T. E., Lockett, A. G., & Muhlemann, A. P. (1982). A Unified Approach To the Acquisition of Subjective Data in R&D. *IEEE Transactions on Engineering Management*, *EM-29*(1), 11–19. https://doi.org/10.1109/TEM.1982.6447461

Georgiou, Y., Zhou, N., Zhong, L., Hoppe, D., Pospieszny, M., Papadopoulou, N., Nikas, K., Nikolos, O. L., Kranas, P., Karagiorgou, S., Pascolo, E., Mercier, M., & Velho, P. (2020). Converging HPC, Big Data and Cloud Technologies for Precision Agriculture Data Analytics on Supercomputers. In H. Jagode, H. Anzt, G. Juckeland, & H. Ltaief (Eds.), *High Performance Computing* (pp. 368–379). Springer International Publishing. https://doi.org/10.1007/978-3-030-59851-8_25

Geum, Y., Lee, H., Lee, Y., & Park, Y. (2015). Development of data-driven technology roadmap considering dependency: An ARM-based technology roadmapping. *Technological Forecasting and Social Change*, *91*, 264–279. https://doi.org/10.1016/j.techfore.2014.03.003

Ghabri, R., Hirmer, P., & Mitschang, B. (2018). A hybrid approach to implement data driven optimization into production environments. *Lecture Notes in Business Information Processing*, *320*, 3–14. https://doi.org/10.1007/978-3-319-93931-5_1

Ghose, A. K., & Dam, H. (2014). The post-theoretic enterprise: A service-oriented view. *Proceedings - IEEE 7th International Conference on Service-Oriented Computing and Applications, SOCA 2014*, 175–182. https://doi.org/10.1109/SOCA.2014.59

Giacumo, L. A., Villachica, S. W., & Breman, J. (2018). Workplace Learning, Big Data, and Organizational Readiness: Where to Start? In D. Ifenthaler (Ed.), *Digital Workplace Learning: Bridging Formal and Informal Learning with Digital Technologies* (pp. 107–127). Springer International Publishing. https://doi.org/10.1007/978-3-319-46215-8_7

Glissmann, S., Sanz, J., Liu, R., & Becker, V. (2012). Disentangling the value of information and analytics through componentized business architecture. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 4250–4259. https://doi.org/10.1109/HICSS.2012.686

Goben, A., & Raszewski, R. (2015). The data life cycle applied to our own data. *Journal of the Medical Library Association : JMLA*, *103*(1), 40–44. https://doi.org/10.3163/1536-5050.103.1.008

Gokalp, M. O., Kayabay, K., Akyol, M. A., Eren, P. E., & Koçyiğit, A. (2016). Big Data for Industry 4.0: A Conceptual Framework. *2016 International Conference on*

Computational Science and Computational Intelligence (CSCI), 431–434. https://doi.org/10.1109/CSCI.2016.0088

Gökalp, M. O., Kayabay, K., Gökalp, E., Koçyiğit, A., & Eren, P. E. (2020). Towards a Model Based Process Assessment for Data Analytics: An Exploratory Case Study. In M. Yilmaz, J. Niemann, P. Clarke, & R. Messnarz (Eds.), *Systems, Software and Services Process Improvement* (pp. 617–628). Springer International Publishing. https://doi.org/10.1007/978-3-030-56441-4_46

Gökalp, M. O., Kayabay, K., Zaki, M., Koçyiğit, A., Eren, P. E., & Neely, A. (2019). Open-Source Big Data Analytics Architecture for Businesses. *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 1–6. https://doi.org/10.1109/UBMYK48245.2019.8965572

Golightly, D., Kefalidou, G., & Sharples, S. (2018). A cross-sector analysis of human and organisational factors in the deployment of data-driven predictive maintenance. *Information Systems and E-Business Management*, *16*(3), 627–648. https://doi.org/10.1007/s10257-017-0343-1

287Gončarovs, P., & Grabis, J. (2017). Using Data Analytics for Continuous Improvement of CRM Processes: Case of Financial Institution. In M. Kirikova, K. Nørvåg, G. A. Papadopoulos, J. Gamper, R. Wrembel, J. Darmont, & S. Rizzi (Eds.), *New Trends in Databases and Information Systems* (pp. 313–323). Springer International Publishing. https://doi.org/10.1007/978-3-319-67162-8_31

Google. (2020). *MLOps: Continuous delivery and automation pipelines in machine learning*. Google Cloud. https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning

Gopalkrishnan, V., Steier, D., Lewis, H., & Guszcza, J. (2012). Big data, big business: Bridging the gap. *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, 7–11. https://doi.org/10.1145/2351316.2351318

Goul, M. (2010). Business Intelligence at the crossroads: Convergence or confusion ahead? *Information Technology Interfaces ITI 2010 32nd International Conference On*, 21–28.

Grover, P., & Kar, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. *Global Journal of Flexible Systems Management*, *18*(3), 203–229. https://doi.org/10.1007/s40171-017-0159-3

Grover, V., Chiang, R. H. L., Liang, T. P., & Zhang, D. (2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*, *35*(2), 388–423. https://doi.org/10.1080/07421222.2018.1451951

Grund, C. K., & Meier, M. C. (2016). Towards game-based management decision support: Using serious games to improve the decision process. *Multikonferenz Wirtschaftsinformatik, MKWI 2016 - Multi-Conference on Information Systems, MKWI 2016*, *1*, 155–166. https://doi.org/10.1139/F05-122

Gudfinnsson, K., & Strand, M. (2018). Challenges with BI adoption in SMEs. *2017 8th International Conference on Information, Intelligence, Systems and Applications, IISA 2017*, *2018-Janua*, 1–6. https://doi.org/10.1109/IISA.2017.8316407

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, *70*, 308–317. https://doi.org/10.1016/j.jbusres.2016.08.004

Gunasekaran, A., Yusuf, Y. Y., Adeleye, E. O., & Papadopoulos, T. (2018). Agile manufacturing practices: The role of big data and business analytics with multiple case studies. *International Journal of Production Research*, *56*(1–2), 385–397. https://doi.org/10.1080/00207543.2017.1395488

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Gupta, R., Gupta, H., & Mohania, M. (2013). Efficient Data Analytics Over Cloud. *Advances in Computers*, *90*, 367–401. https://doi.org/10.1016/B978-0-12-408091-1.00006-3

Hamister, J. W., Magazine, M. J., & Polak, G. G. (2018). Integrating Analytics Through the Big Data Information Chain: A Case From Supply Chain Management. *Journal of Business Logistics*, *39*(3), 220–230. https://doi.org/10.1111/jbl.12192

Han, M., & Geum, Y. (2020). Roadmapping for Data: Concept and Typology of Data-Integrated Smart-Service Roadmaps. *IEEE Transactions on Engineering Management*, 1–13. https://doi.org/10.1109/TEM.2020.3013295

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management*, *36*(10), 1382–1406. https://doi.org/10.1108/IJOPM-02-2014-0098

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80. https://doi.org/10.1016/j.ijpe.2014.04.018

Hazen, B. T., Skipper, J. B., Boone, C. A., & Hill, R. R. (2018). Back in business: Operations research in support of big data analytics for operations and supply chain management. *Annals of Operations Research*, *270*(1), 201–211. https://doi.org/10.1007/s10479-016-2226-0

Hazen, B. T., Weigel, F. K., Ezell, J. D., Boehmke, B. C., & Bradley, V. R. (2017). Toward understanding outcomes associated with data quality improvement. *International Journal of Production Economics*, *193*, 737–747. https://doi.org/10.1016/j.ijpe.2017.08.027

Hee Yeong Kim, J.-S. C., Kim, H. Y., & Cho, J.-S. (2018a). Data governance framework for big data implementation with NPS Case Analysis in Korea. *Journal of Business and Retail Management Research*, *12*(3), 36–46.

Hee Yeong Kim, J.-S. C., Kim, H. Y., & Cho, J.-S. (2018b). Data governance framework for big data implementation with NPS Case Analysis in Korea. *Journal of Business and Retail Management Research*, *12*(3), 36–46.

Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, *268*(3), 836–851. https://doi.org/10.1016/j.ejor.2017.06.031

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, *64*, 130–141. https://doi.org/10.1016/j.dss.2014.05.013

Hosoya, R., & Kamioka, T. (2018). Understanding How the Ad Hoc use of Big Data Analytics Impacts Agility: A Sensemaking-Based Model. *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems, IcABCD 2018*. https://doi.org/10.1109/ICABCD.2018.8465446

Huang, H., Ruan, Y., Shaikh, A., Routray, R., Tan, C. H., & Gopisetty, S. (2009). Building end-to-end management analytics for enterprise data centers. *2009 IFIP/IEEE International Symposium on Integrated Network Management, IM 2009*, 661–675. https://doi.org/10.1109/INM.2009.5188875

IEEE Computer Society. (2021). *Call for Papers: Verifiable and Auditable Federated Learning*. Call for Papers: Special Issue on Trustable, Verifiable, and Auditable Federated Learning. https://www.computer.org/digital-library/journals/bd/call-for-papers-special-issue-on-trustable-verifiable-and-auditable-federated-learning/

Intezari, A., & Gressel, S. (2017). Information and reformation in KM systems: Big data and strategic decision-making. *Journal of Knowledge Management*, *21*(1), 71–91. https://doi.org/10.1108/JKM-07-2015-0293

Iqbal, M., Kazmi, S. H. A., Manzoor, A., Soomrani, A. R., Butt, S. H., & Shaikh, K. A. (2018). A study of big data for business growth in SMEs: Opportunities & challenges. *2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, ICoMET 2018 - Proceedings*, *2018-Janua*, 1–7. https://doi.org/10.1109/ICOMET.2018.8346368

Ismail, M. H., Khater, M., & Zaki, M. (2017). *Digital Business Transformation and Strategy: What Do We Know So Far? November 2017*, 36. https://doi.org/10.13140/RG.2.2.36492.62086

Jaklič, J., Grublješič, T., & Popovič, A. (2018). The role of compatibility in predicting business intelligence and analytics use intentions. *International Journal of Information Management*, *43*, 305–318. https://doi.org/10.1016/j.ijinfomgt.2018.08.017

Janković, S., Mladenović, S., Mladenović, D., Vesković, S., & Glavić, D. (2018). Schema on read modeling approach as a basis of big data analytics integration in EIS. *Enterprise Information Systems*, *12*(8–9), 1180–1201. https://doi.org/10.1080/17517575.2018.1462404

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Jelonek, D., Stępniak, C., & Ziora, L. (2019). The meaning of big data in the support of managerial decisions in contemporary organizations: Review of selected research. *Advances in Intelligent Systems and Computing*, *886*, 361–368. https://doi.org/10.1007/978-3-030-03402-3_24

Jeong, Y., & Yoon, B. (2015). Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. *Technovation*, *39–40*, 37–52. https://doi.org/10.1016/j.technovation.2014.03.001

Ji-fan Ren, S., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, *55*(17), 5011–5026. https://doi.org/10.1080/00207543.2016.1154209

Jobs, C. G., Gilfoil, D. M., & Aukers, S. M. (2016). How marketing organizations can benefit from big data advertising analytics. *Academy of Marketing Studies Journal*, *20*(1), 18–35. https://doi.org/1528-2678

Johnson, J. S., Friend, S. B., & Lee, H. S. (2017). Big Data Facilitation, Utilization, and Monetization: Exploring the 3Vs in a New Product Development Process. *Journal of Product Innovation Management*, *34*(5), 640–658. https://doi.org/10.1111/jpim.12397

Jonas, R. K. (2018). Digital Transformation in Alumina Refining. *Minerals, Metals and Materials Series*, *Part F4*, 79–87. https://doi.org/10.1007/978-3-319-72284-9_12

Jose, B., Ramanan, T. R., & Kumar, S. D. M. (2017). Big data provenance and analytics in telecom contact centers. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, *2017-Decem*, 1573–1578. https://doi.org/10.1109/TENCON.2017.8228107

Kademeteme, E., Kalema, B. M., & Pretorius, P. (2017). Managing and improving data quality through the adoption of data warehouse in the public sector. *African Journal of Science, Technology, Innovation and Development*, *9*(1), 31–41. https://doi.org/10.1080/20421338.2016.1258025

Kamoun-Chouk, S., Berger, H., & Sie, B. H. (2017). Towards integrated model of big data (BD), business intelligence (BI) and knowledge management (KM). *Communications in Computer and Information Science*, *731*, 482–493. https://doi.org/10.1007/978-3-319-62698-7_40

Kasim, H., Hung, T., & Li, X. (2012). Data Value Chain as a Service framework: For enabling data handling, data security and data analysis in the cloud. *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, 804–809. https://doi.org/10.1109/ICPADS.2012.131

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and Good practices. *2013 Sixth International Conference on Contemporary Computing (IC3)*, 404–409. https://doi.org/10.1109/IC3.2013.6612229

Kerr, C., Farrukh, C., Phaal, R., & Probert, D. (2013). Key principles for developing industrially relevant strategic technology management toolkits. *Technological Forecasting and Social Change*, *80*(6), 1050–1070. https://doi.org/10.1016/j.techfore.2012.09.006

Kerr, C., & Phaal, R. (2019). Defining the Scope of a Roadmapping Initiative: A Checklist-Based Template for Organizational Stakeholders. *2019 Portland International Conference on Management of Engineering and Technology (PICMET)*, 1–10. https://doi.org/10.23919/PICMET.2019.8893851

Kerr, C., Phaal, R., & Thams, K. (2019). Customising and deploying roadmapping in an organisational setting: The LEGO Group experience. *Journal of Engineering and Technology Management*, *52*, 48–60. https://doi.org/10.1016/j.jengtecman.2017.10.003

Khachatryan, D., & Karst, N. (2017). V for voice: Strategies for bolstering communication skills in statistics. *Journal of Statistics Education*, *25*(2), 68–78. https://doi.org/10.1080/10691898.2017.1305261

Khatri, V. (2016). Managerial work in the realm of the digital universe: The role of the data triad. *Business Horizons*, *59*(6), 673–688. https://doi.org/10.1016/j.bushor.2016.06.001

Kim, K., & Lee, S. (2018). How can big data complement expert analysis? A value chain case study. *Sustainability (Switzerland)*, *10*(3). https://doi.org/10.3390/su10030709

Kim, M.-K., & Park, J.-H. (2016). Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare. *Information Development*, *September*. https://doi.org/10.1177/0266666916652671

Kimble, C., & Milolidakis, G. (2015). Big Data and Business Intelligence: Debunking the Myths. *Global Business and Organizational Excellence*, *35*(1), 23–34. https://doi.org/10.1002/joe.21642

King, K. G. (2016). Data Analytics in Human Resources: A Case Study and Critical Review. *Human Resource Development Review*, *15*(4), 487–495. https://doi.org/10.1177/1534484316675818

Kisielnicki, J., & Misiak, A. M. (2016). Effectiveness of agile implementation methods in business intelligence projects from an end-user perspective. *Informing Science*, *19*(1), 161–172. https://doi.org/10.1515/fman-2017-0021

Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals* (1 edition). Wiley.

Koch, H., & Peters, U. (2017). Shadow Analytics. *AMCIS 2017 Proceedings*, 1–10.

Koelbl, M., Lehner, F., & Mathew, S. K. (2018). *The IT department's role in enabling business value from business analytics*.

Koronios, A., Gao, J., & Selle, S. (2014). Big Data Project Success—A Meta Analysis. *18th Pacific Asia Conference on Information Systems, PACIS*, 376–390.

Kowalczyk, M., & Buxmann, P. (2014). Big data and information processing in organizational decision processes: A multiple case study. *Business and Information Systems Engineering*, *6*(5), 267–278. https://doi.org/10.1007/s12599-014-0341-5

Kowalczyk, M., & Buxmann, P. (2015a). An ambidextrous perspective on business intelligence and analytics support in decision processes: Insights from a multiple case study. *Decision Support Systems*, *80*, 1–13. https://doi.org/10.1016/j.dss.2015.08.010

Kowalczyk, M., & Buxmann, P. (2015b). Perspectives on Collaboration Procedures and Politics During the Support of Decision Pro- Cesses With Business Intelligence & Analytics. *Twenty-Third European Conference on Information Systems (ECIS), Münster, Germany, 2015*, 1–16.

Kretzer, M., Maedche, A., & Gass, O. (2014a). Barriers to BI & A Generativity: Which Factors impede Stable BI & A Platforms from Enabling Organizational Agility ? *AMCIS 2014 Proceedings*.

Kretzer, M., Maedche, A., & Gass, O. (2014b). Barriers to BI & A Generativity: Which Factors impede Stable BI & A Platforms from Enabling Organizational Agility ? *AMCIS 2014 Proceedings*.

Kridel, D., & Dolk, D. (2013). Automated self-service modeling: Predictive analytics as a service. *Information Systems and E-Business Management*, *11*(1), 119–140. https://doi.org/10.1007/s10257-011-0185-1

Krishnamoorthi, S., & Mathew, S. K. (2018). Business analytics and business value: A comparative case study. *Information and Management*, *55*(5), 643–666. https://doi.org/10.1016/j.im.2018.01.005

Kurniawati, K., Shanks, G., & Bekmamedova, N. (2013). The Business Impact Of Social Media Analytics. *ECIS 2013 Proceedings*, 1–12. https://doi.org/10.1016/j.ejogrb.2015.06.034

Laha, A. (2016). Statistical challenges with big data in management science. *Big Data Analytics: Methods and Applications*, 41–55. https://doi.org/10.1007/978-81-322-3628-3_3

Lai, Y., Sun, H., & Ren, J. (2018). Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: An empirical investigation. *International Journal of Logistics Management*, *29*(2), 676–703. https://doi.org/10.1108/IJLM-06-2017-0153

Lakoju, M., & Serrano, A. (2018). Saving costs with a big data strategy framework. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, *2018-Janua*, 2340–2347. https://doi.org/10.1109/BigData.2017.8258188

Lam, S. K., Sleep, S., Hennig-Thurau, T., Sridhar, S., & Saboo, A. R. (2017). Leveraging Frontline Employees' Small Data and Firm-Level Big Data in Frontline Management: An Absorptive Capacity Perspective. *Journal of Service Research*, *20*(1), 12–28. https://doi.org/10.1177/1094670516679271

Lamba, K., & Singh, S. P. (2018). Modeling big data enablers for operations and supply chain management. *The International Journal of Logistics Management*, *29*(2), 629–658. https://doi.org/10.1108/IJLM-07-2017-0183

Laudon, K. C., & Laudon, J. P. (2019). *Management Information Systems: Managing the Digital Firm* (16 edition). Pearson.

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, *52*(2). https://doi.org/10.0000/PMID57750728

Lee, C. K. M., Yeung, C. L., & Cheng, M. N. (2015). Research on IoT based Cyber Physical System for Industrial big data Analytics. *2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 1855–1859. https://doi.org/10.1109/IEEM.2015.7385969

Lee, J. H., Phaal, R., & Lee, S.-H. (2013). An integrated service-device-technology roadmap for smart city development. *Technological Forecasting and Social Change*, *80*(2), 286–306. https://doi.org/10.1016/j.techfore.2012.09.020

Lee, J. Y., Yoon, J. S., & Kim, B. H. (2017). A big data analytics platform for smart factories in small and medium-sized manufacturing enterprises: An empirical case study of a die casting factory. *International Journal of Precision Engineering and Manufacturing*, *18*(10), 1353–1361. https://doi.org/10.1007/s12541-017-0161-x

Lee, S., & Park, Y. (2005). Customization of technology roadmaps according to roadmapping purposes: Overall process and detailed modules. *Technological Forecasting and Social Change*, *72*(5), 567–583. https://doi.org/10.1016/j.techfore.2004.11.006

Lehrer, C., Wieneke, A., vom Brocke, J., Jung, R., & Seidel, S. (2018). How Big Data Analytics Enables Service Innovation: Materiality, Affordance, and the Individualization of Service. *Journal of Management Information Systems*, *35*(2), 424–460. https://doi.org/10.1080/07421222.2018.1451953

Li, Q., Lan, L., Zeng, N., You, L., Yin, J., Zhou, X., & Meng, Q. (2019). A Framework for Big Data Governance to Advance RHINs: A Case Study of China. *IEEE Access*, *7*, 50330–50338. https://doi.org/10.1109/ACCESS.2019.2910838

Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, *37*(3), 114–124. https://doi.org/10.1016/j.ijinfomgt.2016.12.003

Liu, K., & Shi, J. (2015). A Systematic Approach for Business Data Analytics with a Real Case Study. *International Journal of Business Analytics (IJBAN)*, *2*(4), 23–44. https://doi.org/10.4018/IJBAN.2015100102

Lizotte-Latendresse, S., & Beauregard, Y. (2018). Implementing self-service business analytics supporting lean manufacturing: A state-of-the-art review. *IFAC-PapersOnLine*, *51*(11), 1143–1148. https://doi.org/10.1016/j.ifacol.2018.08.436

Llave, M. R. (2017). Business Intelligence and Analytics in Small and Medium-sized Enterprises: A Systematic Literature Review. *Procedia Computer Science*, *121*, 194–205. https://doi.org/10.1016/j.procs.2017.11.027

Lu, J. (2018). A data-driven framework for business analytics in the context of big data. *Communications in Computer and Information Science*, *909*, 339–351. https://doi.org/10.1007/978-3-030-00063-9_32

Magee, B., Sammon, D., Nagle, T., & O'Raghallaigh, P. (2016). Introducing data driven practices into sales environments: Examining the impact of data visualisation on user engagement and sales results. *Journal of Decision Systems*, *25*, 313–328. https://doi.org/10.1080/12460125.2016.1187403

Malaka, I., & Brown, I. (2015). Challenges to the Organisational Adoption of Big Data Analytics: A Case Study in the South African Telecommunications Industry. *Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists*, 1–9. https://doi.org/10.1145/2815782.2815793

Maleki, D., Arianyan, E., & Ahmadi, M. R. (2016). Big data: A Coevolutionary strategic roadmap. *2016 8th International Symposium on Telecommunications (IST)*, 202–207. https://doi.org/10.1109/ISTEL.2016.7881810

Mason, H., & Wiggins, C. (2010, September 25). A Taxonomy of Data Science. *Dataists*. http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

Maturana, F. P., & Asenjo, J. L. (2015). Chapter 11—Cloud Based Agent Framework for the Industrial Automation Sector. In P. Leitão & S. Karnouskos (Eds.), *Industrial Agents* (pp. 193–206). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-800341-1.00011-5

Mayer-Schönberger, V., & Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Reprint edition). Eamon Dolan/Mariner Books.

McAfee, A., & Brynjolfsson, E. (2012, October 1). Big Data: The Management Revolution. *Harvard Business Review*, *October 2012*. https://hbr.org/2012/10/big-data-the-management-revolution

Melville, N. P. (2015). Digital fitness: Four principles for successful development of digital initiatives. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2015-March*, 4387–4396. https://doi.org/10.1109/HICSS.2015.526

Microsoft. (2020, January 10). *What is the Team Data Science Process?* https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview

Mikalef, P., Giannakos, M. N., Pappas, I. O., & Krogstie, J. (2018). The human side of big data: Understanding the skills of the data scientist in education and industry. *IEEE Global Engineering Education Conference, EDUCON*, *2018-April*, 503–512. https://doi.org/10.1109/EDUCON.2018.8363273

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

Mikalef, P., van de Wetering, R., & Krogstie, J. (2018). Big Data Enabled Organizational Transformation: The Effect of Inertia in Adoption and Diffusion. In W. Abramowicz & A. Paschke (Eds.), *Business Information Systems* (pp. 135–147). Springer International Publishing. https://doi.org/10.1007/978-3-319-93931-5_10

Miller, G. J. (2018). *Comparative Analysis of Big Data and BI Projects*. 701–705. https://doi.org/10.15439/2018F125

Milosevic, Z., Chen, W., Berry, A., & Rabhi, F. A. (2016). Real-Time Analytics. *Big Data: Principles and Paradigms*, 39–61. https://doi.org/10.1016/B978-0-12-805394-2.00002-7

Mishra, R., & Saini, A. K. (2016). Business intelligence and analytics: Paving way for operational excellence in indian banks. *Proceedings of the International*

Conference on Industrial Engineering and Operations Management, *8-10 March*, 53–58.

Mohapatra, S., & Ghosh, M. S. (2016). Enterprise IT strategy framework for adopting big data analytics and cloud a case on media sector. *International Journal of Enterprise Network Management*, *7*(4), 334. https://doi.org/10.1504/IJENM.2016.080461

Moreira, J., Carvalho, A., & Horvath, T. (2018). *A General Introduction to Data Analytics* (1 edition). Wiley-Interscience.

Mounir, N., Guo, Y., Panchal, Y., Mohamed, I. M., Abou-Sayed, A., & Abou-Sayed, O. (2018). Integrating Big Data: Simulation, Predictive Analytics, Real Time Monitoring, and Data Warehousing in a Single Cloud Application. *Offshore Technology Conference*. https://doi.org/10.4043/28910-MS

Mousannif, H., Sabah, H., Douiji, Y., & Oulad Sayad, Y. (2016). Big data projects: Just jump right in! *International Journal of Pervasive Computing and Communications*, *12*(2), 260–288. https://doi.org/10.1108/IJPCC-04-2016-0023

Mowrer, M., Roberts, B., & Paula, H. (2017). Tapping into your current data reserves. *Proceedings of the Annual Offshore Technology Conference*, *5*, 3748–3762.

Muhammad, A., Khan, A., Amin, N., & Lambrou, N. (2010). Drivers and barriers to business intelligence adoption: A case of Pakistan. *European and Mediterranean Conference on Information Systems 2010*, *2010*, 1–23. http://dx.doi.org/10.1016/S0140-6736%2810%2962036-3

Muhammad Habib ur, R., Victo, C., Aish, B., & Teh Ying, W. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, *36*, 917–928.

Müller, O., Fay, M., & vom Brocke, J. (2018). The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems*, *35*(2), 488–509. https://doi.org/10.1080/07421222.2018.1451955

Mushore, R., & Kyobe, M. (2017). Leveraging predictive analytical for business value: Theoretical foundations. *Proceedings - 2016 3rd International Conference on Advances in Computing, Communication and Engineering, ICACCE 2016*, 395–400. https://doi.org/10.1109/ICACCE.2016.8073781

Nair, P. R. (2015). Tackling supply chain management through business analytics: Opportunities and challenges. *Advances in Intelligent Systems and Computing*, *337*, 569–576. https://doi.org/10.1007/978-3-319-13728-5_64

Nalchigar, S., & Yu, E. (2013). From business intelligence insights to actions: A methodology for closing the sense-and-respond loop in the adaptive enterprise. *Lecture Notes in Business Information Processing*, *165 LNBIP*, 114–128. https://doi.org/10.1007/978-3-642-41641-5_9

Nalchigar, S., & Yu, E. (2017). Conceptual modeling for business analytics: A framework and potential benefits. *Proceedings - 2017 IEEE 19th Conference on Business Informatics, CBI 2017*, *1*, 369–378. https://doi.org/10.1109/CBI.2017.63

Nalchigar, S., & Yu, E. (2018). Business-driven data analytics: A conceptual modeling framework. *Data and Knowledge Engineering*, *117*, 359–372. https://doi.org/10.1016/j.datak.2018.04.006

Nalchigar, S., Yu, E., & Easterbrook, S. (2014). Towards actionable business intelligence: Can system dynamics help? *Lecture Notes in Business Information Processing*, *197*, 246–260. https://doi.org/10.1007/978-3-662-45501-2_18

Nanjappan, V., Liang, H. N., Wang, W., & Man, K. L. (2017). Big Data: A Classification of Acquisition and Generation Methods. *Big Data Analytics for Sensor-Network Collected Intelligence*, 3–20. https://doi.org/10.1016/B978-0-12-809393-1.00001-5

Newman, R., Chang, V., Walters, R. J., & Wills, G. B. (2016). Model and experimental development for Business Data Science. *International Journal of Information Management*, *36*(4), 607–617. https://doi.org/10.1016/j.ijinfomgt.2016.04.004

Nino, M., Blanco, J. M., & Illarramendi, A. (2015). Business understanding, challenges and issues of Big Data Analytics for the servitization of a capital equipment manufacturer. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 1368–1377. https://doi.org/10.1109/BigData.2015.7363897

Niño, M., Sáenz, F., Blanco, J. M., & Illarramendi, A. (2016). Requirements for a big data capturing and integration architecture in a distributed manufacturing scenario. *Ieee Indin*. https://doi.org/10.1109/INDIN.2016.7819372

O 'donovan, P., Bruton, K., & O 'sullivan, D. T. J. (2016). Case study: The implementation of a data—Driven industrial analytics methodology and platform for smart manufacturing. *International Journal of Prognostics and Health Management*, 2153–2648.

O'Donovan, P., Bruton, K., & O'Sullivan, D. T. J. (2016). Case study: The implementation of a data-driven industrial analytics methodology and platform for smart manufacturing. *International Journal of Prognostics and Health Management*, *7*, 1–22.

O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. (2015). Big data in manufacturing: A systematic mapping study. *Journal of Big Data*, *2*(1), 20. https://doi.org/10.1186/s40537-015-0028-x

Olszak, C. M. (2016). Toward Better Understanding and Use of Business Intelligence in Organizations. *Information Systems Management*, *33*(2), 105–123. https://doi.org/10.1080/10580530.2016.1155946

Ordonez, C., & Garcia-Garcia, J. (2016). Managing Big Data Analytics Workflows with a Database System. *Proceedings - 2016 16th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, CCGrid 2016*, 649–655. https://doi.org/10.1109/CCGrid.2016.63

Ozkaya, I. (2020). What Is Really Different in Engineering AI-Enabled Systems? *IEEE Software*, *37*(4), 3–6. https://doi.org/10.1109/MS.2020.2993662

Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research*, *252*(2), 687–698. https://doi.org/10.1016/j.ejor.2016.01.052

Park, G., Park, S., Khan, L., & Chung, L. (2017). IRIS: A goal-oriented big data analytics framework on Spark for better Business decisions. *2017 IEEE International Conference on Big Data and Smart Computing, BigComp 2017*, 76–83. https://doi.org/10.1109/BIGCOMP.2017.7881719

Park, H., Phaal, R., Ho, J.-Y., & O'Sullivan, E. (2020). Twenty years of technology and strategic roadmapping research: A school of thought perspective. *Technological Forecasting and Social Change*, *154*, 119965. https://doi.org/10.1016/j.techfore.2020.119965

Paschke, A. (2016). Provalets: Component-based mobile agents as microservices for rule-based data access, processing and analytics. *Business and Information Systems Engineering*, *58*(5), 329–340. https://doi.org/10.1007/s12599-016-0447-z

Pauleen, D. J., Rooney, D., & Intezari, A. (2017). Big data, little wisdom: Trouble brewing? Ethical implications for the information systems discipline. *Social Epistemology*, *31*(4), 400–416. https://doi.org/10.1080/02691728.2016.1249436

Pearson, R. J., Costley, A. E., Phaal, R., & Nuttall, W. J. (2020). Technology Roadmapping for mission-led agile hardware development: A case study of a commercial fusion energy start-up. *Technological Forecasting and Social Change*, *158*, 120064. https://doi.org/10.1016/j.techfore.2020.120064

Pérez-González, C. J., Colebrook, M., Roda-García, J. L., & Rosa-Remedios, C. B. (2019). Developing a data analytics platform to support decision making in emergency and security management. *Expert Systems with Applications*, *120*, 167–184. https://doi.org/10.1016/j.eswa.2018.11.023

Phaal, R., Farrukh, C. J. P., & Probert, D. R. (2004a). Technology roadmapping—A planning framework for evolution and revolution. *Technological Forecasting and Social Change*, *71*(1), 5–26. https://doi.org/10.1016/S0040-1625(03)00072-6

Phaal, R., Farrukh, C. J. P., & Probert, D. R. (2007). Strategic Roadmapping: A Workshop-based Approach for Identifying and Exploring Strategic Issues and Opportunities. *Engineering Management Journal*, *19*(1), 3–12. https://doi.org/10.1080/10429247.2007.11431716

Phaal, R., Farrukh, C., & Probert, D. (2004b). Customizing roadmapping. *Research-Technology Management*, *47*(2), 26–37.

Phaal, R., Farrukh, C., & Probert, D. R. (2013). Fast-start roadmapping workshop approaches. In *Technology Roadmapping for Strategy and Innovation* (pp. 91–106). Springer.

Phaal, R., & Muller, G. (2009). An architectural framework for roadmapping: Towards visual strategy. *Technological Forecasting and Social Change*, *76*(1), 39–49. https://doi.org/10.1016/j.techfore.2008.03.018

Phillips-Wren, G., & Hoskisson, A. (2014). Decision support with big data: A case study in the hospitality industry. *Frontiers in Artificial Intelligence and Applications*, *261*, 401–413. https://doi.org/10.3233/978-1-61499-399-5-401

Phillips-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. *Journal of Decision Systems*, *24*(1), 87–102. https://doi.org/10.1080/12460125.2015.994333

Pickering, C. K. (2013). Synergizing people, process, and technology to motivate knowledge sharing and collaboration Industry case study. *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*, 35–39. https://doi.org/10.1109/CTS.2013.6567200

Plomp, M. G. A. (2017). *Data-driven Decision-making and the Role of Personality and Cognitive Style: A Systematic Literature Review Full Paper*. 1–10.

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. https://doi.org/10.1007/s10796-016-9720-4

Pora, U., Gerdsri, N., Thawesaengskulthai, N., & Triukose, S. (2020). Data-Driven Roadmapping (DDRM): Approach and Case Demonstration. *IEEE Transactions on Engineering Management*, 1–19. https://doi.org/10.1109/TEM.2020.3005341

Power, D. J. (2015). 'Big data' decision making use cases. *Lecture Notes in Business Information Processing*, *216*. https://doi.org/10.1007/978-3-319-18533-0_1

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking* (1 edition). O'Reilly Media.

Pusala, M. K., Amini Salehi, M., Katukuri, J. R., Xie, Y., & Raghavan, V. (2016). Massive data analysis: Tasks, tools, applications, and challenges. *Big Data Analytics: Methods and Applications*, 11–40. https://doi.org/10.1007/978-81-322-3628-3_2

Radha, K., Rao, B. T., Babu, S. M., Rao, K. T., Reddy, V. K., & Saikiran, P. (2015). Service Level Agreements in Cloud Computing and Big Data. *International Journal of Electrical and Acomputer Engineering*, *5*(1), 158–165.

Raffoni, A., Visani, F., Bartolini, M., & Silvi, R. (2018). Business Performance Analytics: Exploring the potential for Performance Management Systems. *Production Planning and Control*, *29*(1), 51–67. https://doi.org/10.1080/09537287.2017.1381887

Raguseo, E., & Vitari, C. (2018). Investments in big data analytics and firm performance: An empirical investigation of direct and mediating effects. *International Journal of Production Research*, *56*(15), 5206–5221. https://doi.org/10.1080/00207543.2018.1427900

Rahman, N., & Aldhaban, F. (2015). Assessing the effectiveness of big data initiatives. *Portland International Conference on Management of Engineering and*

Technology, *2015-Septe*, 478–484. https://doi.org/10.1109/PICMET.2015.7273189

Ramanathan, R., Philpott, E., Duan, Y., & Cao, G. (2017). Adoption of business analytics and impact on performance: A qualitative study in retail. *Production Planning and Control*, *28*(11–12), 985–998. https://doi.org/10.1080/09537287.2017.1336800

Ramesh, B., & Ramakrishna, A. (2018). Unified business intelligence ecosystem: A project management approach to address business intelligence challenges. *PICMET 2018 - Portland International Conference on Management of Engineering and Technology: Managing Technological Entrepreneurship: The Engine for Economic Growth, Proceedings*. https://doi.org/10.23919/PICMET.2018.8481744

Rane, S. B., & Mishra, N. (2018). Roadmap for business analytics implementation using DIPPS model for sustainable business excellence: Case studies from the multiple fields. *International Journal of Business Excellence*, *15*(3), 308–334. https://doi.org/10.1504/IJBEX.2018.092574

Rao, L., McNaughton, M., & Mansingh, G. (2018). An Agile Integrated Methodology for Strategic Business Intelligence (AimS-BI). *AMCIS 2018 Proceedings*. https://aisel.aisnet.org/amcis2018/DigitalAgility/Presentations/4

Rekha, J. H., & Parvathi, R. (2015). Survey on software project risks and big data analytics. *Procedia Computer Science*, *50*, 295–300. https://doi.org/10.1016/j.procs.2015.04.045

Ren, Z., Wan, J., Shi, W., Xu, X., & Zhou, M. (2014). Workload analysis, implications, and optimization on a production hadoop cluster: A case study on taobao. *IEEE Transactions on Services Computing*, *7*(2), 307–321. https://doi.org/10.1109/TSC.2013.40

Rialti, R., Marzi, G., Silic, M., & Ciappei, C. (2018). Ambidextrous organization and agility in big data era: The role of business process management systems. *Business Process Management Journal*, *24*(5), 1091–1109. https://doi.org/10.1108/BPMJ-07-2017-0210

Ricardo, J., Bernardino, J., & Almeida, A. (2008). Cloud Business Intelligence for Virtual Organizations. *Proceedings of the 2014 International C* Conference on Computer Science & Software Engineering - C3S2E '14*, 1–7. https://doi.org/10.1145/2641483.2641520

Ries, E. (2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses* (1 edition). Currency.

Rikhardsson, P., & Yigitbasioglu, O. (2018). Business intelligence & analytics in management accounting research: Status and future focus. *International Journal of Accounting Information Systems*, *29*, 37–58. https://doi.org/10.1016/j.accinf.2018.03.001

Riungu-Kalliosaari, L., Kauppinen, M., & Männistö, T. (2017). What can be learnt from experienced data scientists? A case study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10611 LNCS*, 55–70. https://doi.org/10.1007/978-3-319-69926-4_5

Roderick, O., Marko, N., Sanchez, D., & Aryasomajula, A. (2017). Data analysis and machine learning effort in healthcare: Organization, limitations, and development of an approach. *Internet of Things and Data Analytics Handbook*, 295–328. https://doi.org/10.1002/9781119173601.ch18

Rogers, D. (2016). *The Digital Transformation Playbook, Rethink Your Business for the Digital Age*. Columbia University Press. https://doi.org/10.7312/roge17544

Rohleder, C., Lin, J., Kusumah, I., & Özkan, G. (2014). Requirements engineering in business analytics for innovation and product lifecycle management. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8697*, 51–58.

Rosenthal, S., McMillan, S., & Gaston, M. E. (2015). Developer toolchains for large-scale analytics: Two case studies. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 1311–1316. https://doi.org/10.1109/BigData.2015.7363887

Roy, U., Zhu, B., Li, Y., Zhang, H., & Yaman, O. (2014). Mining big data in manufacturing: Requirement analysis, tools and techniques. *ASME 2014 International Mechanical Engineering Congress and Exposition, IMECE 2014*, *11*. https://doi.org/10.1115/IMECE2014-38822

Rubab, S., Taqvi, S. A., & Hassan, M. F. (2019). Realizing the value of big data in process monitoring and control: Current issues and opportunities. *Advances in Intelligent Systems and Computing*, *843*, 128–138. https://doi.org/10.1007/978-3-319-99007-1_13

Sá, J. O. e., Martins, C., & Simões, P. (2015). Big data in cloud: A data architecture. In *Advances in Intelligent Systems and Computing* (Vol. 353, pp. 723–732). https://doi.org/10.1007/978-3-319-16486-1_71

Sahu, S. K., Jacintha, M. M., & Singh, A. P. (2017). Comparative study of tools for big data analytics: An analytical study. *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2017*, *2017-Janua*, 37–41. https://doi.org/10.1109/CCAA.2017.8229827

Saltz, J. S., & Grady, N. W. (2017). The ambiguity of data science team roles and the need for a data science workforce framework. *2017 IEEE International Conference on Big Data (Big Data)*, 2355–2361. https://doi.org/10.1109/BigData.2017.8258190

Saltz, J. S., & Shamshurin, I. (2018). Does pair programming work in a data science context? An initial case study. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, *2018-Janua*, 2348–2354. https://doi.org/10.1109/BigData.2017.8258189

Saltz, J. S., Yilmazel, S., & Yilmazel, O. (2016). Not all software engineers can become good data engineers. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 2896–2901. https://doi.org/10.1109/BigData.2016.7840939

Saltz, J., Shamshurin, I., & Connors, C. (2017). A framework for describing big data projects. *Lecture Notes in Business Information Processing*, *263*, 183–195. https://doi.org/10.1007/978-3-319-52464-1_17

Sangupamba Mwilu, O., Comyn-Wattiau, I., & Prat, N. (2016). Design science research contribution to business intelligence in the cloud—A systematic literature review. *Future Generation Computer Systems*, *63*, 108–122. https://doi.org/10.1016/j.future.2015.11.014

Schuff, D., Corral, K., St. Louis, R. D., & Schymik, G. (2018). Enabling self-service BI: A methodology and a case study for a model management warehouse. *Information Systems Frontiers*, *20*(2), 275–288. https://doi.org/10.1007/s10796-016-9722-2

Schuh, G., Gudergan, G., Feige, B. A., Buschmeyer, A., & Krechting, D. (2015). Business transformation in the manufacturing industry—How information acquisition, analysis, usage and distribution affects the success of lifecycle-product-service-systems. *Procedia CIRP*, *30*, 335–340. https://doi.org/10.1016/j.procir.2015.02.133

Schüll, A., & Maslan, N. (2018). *On the Adoption of Big Data Analytics: Interdependencies of Contextual Factors*. 425–431. https://doi.org/10.5220/0006759904250431

Schüritz, R., Brand, E., Satzger, G., & Bischhoffshausen, J. (2017). How To Cultivate Analytics Capabilities Within an Organization? – Design and Types of Analytics Competency Centers. *Proceedings of the 25th European Conference on Information Systems (ECIS)*, *2017*, 1–15.

Schuritz, R., & Satzger, G. (2016). Patterns of Data-Infused Business Model Innovation. *Proceedings - CBI 2016: 18th IEEE Conference on Business Informatics*, *1*, 133–142. https://doi.org/10.1109/CBI.2016.23

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., & Young, M. (2014). Machine Learning: The High-Interest Credit Card of Technical Debt. *NIPS 2014 Workshop on Software Engineering for Machine Learning (SE4ML)*, 1–9. https://doi.org/10.1007/s13398-014-0173-7.2

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems*, 2503–2511.

Seddon, P. B., & Constantinidis, D. (2012). Through What Mechanisms Does Business Analytics Contribute To Business Value? *PACIS 2012 Proceedings*, 1–12.

Seddon, P. B., Constantinidis, D., Tamm, T., & Dod, H. (2017). How does business analytics contribute to business value? *Information Systems Journal*, *27*(3), 237–269. https://doi.org/10.1111/isj.12101

Segooa, M. A., & Kalema, B. M. (2018). Leveraging big data analytics to improve decision making in South African public universities. *2018 IEEE 3rd International Conference on Big Data Analysis, ICBDA 2018*, 9–13. https://doi.org/10.1109/ICBDA.2018.8367462

Sena, V., Bhaumik, S., Sengupta, A., & Demirbag, M. (2019). Big Data and Performance: What Can Management Research Tell us? *British Journal of Management*, *30*(2), 219–228. https://doi.org/10.1111/1467-8551.12362

Serrato, M., & Ramirez, J. (2016). The strategic business value of big data. *Big Data Management*, 47–70. https://doi.org/10.1007/978-3-319-45498-6_3

Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research*, *70*, 366–378. https://doi.org/10.1016/j.jbusres.2016.08.010

Shah, S., Soriano, C. B., & Coutroubis, A. D. (2018). Is big data for everyone? The challenges of big data adoption in SMEs. *IEEE International Conference on Industrial Engineering and Engineering Management*, *2017-Decem*, 803–807. https://doi.org/10.1109/IEEM.2017.8290002

Shanks, G., & Bekmamedova, N. (2012a). Achieving benefits with business analytics systems: An evolutionary process perspective. *Journal of Decision Systems*, *21*(3), 231–244. https://doi.org/10.1080/12460125.2012.729182

Shanks, G., & Bekmamedova, N. (2012b). The Impact of Strategy on Business Analytics Success. *Proceedings of the 23rd Australasian Conference on Information Systems 2012*, 1–11.

Shanks, G., Bekmamedova, N., & Willcocks, L. (2012a). Business Analytics: Enabling Strategic Alignment and Organizational Transformation. *ECIS 2012 Proceedings*, 18. https://doi.org/10.4018/ijbir.2013070101

Shanks, G., Bekmamedova, N., & Willcocks, L. (2012b). Business Analytics: Enabling Strategic Alignment and Organizational Transformation. *ECIS 2012 Proceedings*, 18. https://doi.org/10.4018/ijbir.2013070101

Shanks, G., & Sharma, R. (2011). Creating value from business analytics systems: The impact of strategy. *15th Pacific Asia Conference on Informaton Systems*, *2011*, 1–12.

Shanks, G., Sharma, R., Seddon, P., & Reynolds, P. (2010). The impact of strategy and maturity on business analytics and firm performance: A review and research agenda. *Australasian Conference on Information Systems*, 1–11.

Shanmuganathan, M. (2019). *The IBS as a Catalyst: Data Driven Insights*. 311–329. https://doi.org/10.1007/978-3-030-03402-3_21

Sharma, V., Stranieri, A., Ugon, J., Vamplew, P., & Martin, L. (2017). An agile group aware process beyond CRISP-DM: A hospital data mining case study. *ACM International Conference Proceeding Series*, *Part F1302*. https://doi.org/10.1145/3093241.3093273

Shcherbakov, M., Shcherbakova, N., Brebels, A., Janovsky, T., & Kamaev, V. (2014). Lean Data Science Research Life Cycle: A Concept for Data Analysis Software Development. *Communications in Computer and Information Science*, *466 CCIS*, 708–716. https://doi.org/10.1007/978-3-319-11854-3_61

Shollo, A. (2011). *Using Business Intelligence in IT Governance Decision Making*. 3–15. https://doi.org/10.1007/978-3-642-24148-2_1

Shuradze, G., & Wagner, H.-T. (2016). Governing for Agility and Innovation in Data-Rich Environments: The Role of Data Analytics Capabilities. *ECIS 2016 Proceedings*.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Smuc, M., Mayr, E., Lammarsch, T., Bertone, A., Aigner, W., Risku, H., & Miksch, S. (2008). Visualizations at first sight: Do insights require training? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5298 LNCS*, 261–280. https://doi.org/10.1007/978-3-540-89350-9-19

Spruit, M., & Sacu, C. (2015a). DWCMM: The data warehouse capability maturity model. *Journal of Universal Computer Science*, *21*(11), 1508–1534. https://doi.org/10.3217/jucs-021-11-1508

Spruit, M., & Sacu, C. (2015b). DWCMM: The data warehouse capability maturity model. *Journal of Universal Computer Science*, *21*(11), 1508–1534. https://doi.org/10.3217/jucs-021-11-1508

Sun, N., Morris, J. G., Xu, J., Zhu, X., & Xie, M. (2014). ICARE: A framework for big data-based banking customer analytics. *IBM Journal of Research and Development*, *58*(5/6), 4:1-4:9. https://doi.org/10.1147/JRD.2014.2337118

Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2019). Factors influencing effective use of big data: A research framework. *Information & Management*, *57*(1), 103146. https://doi.org/10.1016/j.im.2019.02.001

Tan, C., & Haji, M. (2017). Big Data Educational Portal for Small and Medium Sized Enterprises (SMEs). *Proceedings of the International Conference on Big Data and Internet of Thing - BDIOT2017*, 11–15. https://doi.org/10.1145/3175684.3175688

Tanwar, M., Duggal, R., & Khatri, S. K. (2015). Unravelling unstructured data: A wealth of information in big data. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2015*. https://doi.org/10.1109/ICRITO.2015.7359270

Tao, C., & Gao, J. (2016). *Quality Assurance for Big Data Application – Issuses, Challenges, and Needs*. 375–381. https://doi.org/10.18293/SEKE2016-166

Thirathon, U., Wieder, B., & Ossimitz, M.-L. (2018). Determinants of analytics-based managerial decision-making. *International Journal of Information Systems and Project Management*, *6*(1). https://doi.org/10.12821/ijispm060102

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Trieu, V. H. (2017). Getting value from Business Intelligence systems: A review and research agenda. *Decision Support Systems*, *93*, 111–124. https://doi.org/10.1016/j.dss.2016.09.019

Troilo, M., Bouchet, A., Urban, T. L., & Sutton, W. A. (2016). Perception, reality, and the adoption of business analytics: Evidence from North American professional sport organizations. *Omega (United Kingdom)*, *59*, 72–83. https://doi.org/10.1016/j.omega.2015.05.011

Troisi, O., D'Arco, M., Loia, F., & Maione, G. (2018). Big data management: The case of Mulino Bianco's engagement platform for value co-creation. *International Journal of Engineering Business Management*, *10*. https://doi.org/10.1177/1847979018767776

Truong, H. L., & Dustdar, S. (2014). Sustainability data and analytics in cloud-based M2M systems. *Studies in Computational Intelligence*, *546*, 343–365. https://doi.org/10.1007/978-3-319-05029-4_14

Tucker, I., Gil-Garcia, J. R., & Sayogo, D. S. (2017). Collaborative Data Analytics for Emergency Response: Identifying Key Factors and Proposing a Preliminary Framework. *10th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2017)*, 508–515. https://doi.org/10.1145/3047273.3047379

Vashisht, P., & Gupta, V. (2016). Big data analytics techniques: A survey. *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 264–269. https://doi.org/10.1109/ICGCIoT.2015.7380470

Veneberg, R. K. M., Iacob, M. E., Sinderen, M. J. V., & Bodenstaff, L. (2014). Enterprise Architecture Intelligence: Combining Enterprise Architecture and Operational Data. *IEEE International Enterprise Distributed Object Computing Conference (EDOC)*, 22–31. https://doi.org/10.1109/EDOC.2014.14

Vera-Baquero, A., Molloy, O., Colomo-Palacios, R., & Elbattah, M. (2015). Business process improvement by means of big data based decision support systems: A case study on call centers. *Int J Inf Syst Proj Manag*, *3*(1), 5–26. https://doi.org/10.12821/ijispm030101

Verma, N., & Voida, A. (2016). Mythologies of Business Intelligence. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, 2341–2347. https://doi.org/10.1145/2851581.2892379

Verma, S. (2017). Big Data and Advance Analytics: Architecture, Techniques, Applications, and Challenges. *International Journal of Business Analytics*, *4*(4), 21–47. https://doi.org/10.4018/IJBAN.2017100102

Verma, S., Bhattacharyya, S. S., & Kumar, S. (2018). An extension of the technology acceptance model in the big data analytics system implementation environment. *Information Processing and Management*, *54*(5), 791–806. https://doi.org/10.1016/j.ipm.2018.01.004

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626–639. https://doi.org/10.1016/j.ejor.2017.02.023

Vo, Q. D., Thomas, J., Cho, S., De, P., Choi, B. J., & Sael, L. (2017). *Next Generation Business Intelligence and Analytics: A Survey*. https://doi.org/10.1145/3278252.3278292

Wahdain, E. A., Baharudin, A. S., & Ahmad, M. N. (2019). Big Data Analytics in the Malaysian Public Sector: The Determinants of Value Creation. In F. Saeed, N. Gazem, F. Mohammed, & A. Busalim (Eds.), *Recent Trends in Data Science and Soft Computing* (pp. 139–150). Springer International Publishing. https://doi.org/10.1007/978-3-319-99007-1_14

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. fan, Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wang, F. L., Rischmoller, L., Reed, D., & Khanzode4, A. (2018). *Ad Hoc Data Analytics and Business Intelligence Service Framework for Construction Projects*. 1058–1068. https://doi.org/10.24928/2018/0535

Wang, Y., & Byrd, T. A. (2017). Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. *Journal of Knowledge Management*, *21*(3), 517–539. https://doi.org/10.1108/JKM-08-2015-0301

Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, *70*, 287–299. https://doi.org/10.1016/j.jbusres.2016.08.002

Wang, Y., Kung, L. A., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019

Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, *55*(1), 64–79. https://doi.org/10.1016/j.im.2017.04.001

Watson, H. J. (2018). Revisiting ralph sprague's framework for developing decision support systems. *Communications of the Association for Information Systems*, *42*(1), 363–385. https://doi.org/10.17705/1CAIS.04213

White, S. A. (2004). Introduction to BPMN. *IBM Corp.*, *2*(0), 0.

Willcocks, L., Venters, W., & Whitley, E. (2012). Cloud sourcing: Implications for managing the IT function. *Lecture Notes in Business Information Processing*, *130 LNBIP*, 142–163. https://doi.org/10.1007/978-3-642-33920-2_9

Willyard, C. H., & McClees, C. W. (1987). Motorola's Technology Roadmap Process. *Research Management*, *30*(5), 13–19. https://doi.org/10.1080/00345334.1987.11757057

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–39.

Wu, C., Buyya, R., & Ramamohanarao, K. (2016). Big Data Analytics = Machine Learning + Cloud Computing. In R. Buyya, R. N. Calheiros, & A. V. Dastjerdi (Eds.), *Big Data* (pp. 3–38). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-805394-2.00001-5

Yasin, A., Liu, L., Cao, Z., Wang, J., Liu, Y., & Ling, T. S. (2018). Big data services requirements analysis. *Communications in Computer and Information Science*, *809*, 3–14. https://doi.org/10.1007/978-981-10-7796-8_1

Yu, E., Deng, S., & Sasmal, D. (2012). Enterprise architecture for the adaptive enterprise—A vision paper. *Lecture Notes in Business Information Processing*, *131 LNBIP*, 146–161. https://doi.org/10.1007/978-3-642-34163-2_9

Zafeiropoulos, A., Fotopoulou, E., Gonzalez-Vidal, A., & Skarmeta, A. (2018). Detaching the design, development and execution of big data analysis processes: A case study based on energy and behavioral analytics. *2018 Global Internet of Things Summit, GIoTS 2018*. https://doi.org/10.1109/GIOTS.2018.8534525

Zahid, H., Mahmood, T., & Ikram, N. (2018). Enhancing dependability in big data analytics enterprise pipelines. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11342 LNCS*, 272–281. https://doi.org/10.1007/978-3-030-05345-1_23

Zhan, Y., Tan, K. H., & Perrons, R. K. (2018). A proposed framework for accelerated innovation in data-driven environments: Evidence and emerging trends from China. *Industrial Management & Data Systems*, *118*(6), 1266–1286. https://doi.org/10.1108/IMDS-11-2017-0542

Zhao, J., Liu, Y., & Zhou, P. (2018). Framing a Sustainable Architecture for Data Analytics Systems: An Exploratory Study. *IEEE Access*, *6*, 61600–61613. https://doi.org/10.1109/ACCESS.2018.2872673

Zhu, J., Zhuang, E., Fu, J., Baranowski, J., Ford, A., & Shen, J. (2016a). A Framework-Based Approach to Utility Big Data Analytics. *IEEE Transactions on Power Systems*, *31*(3), 2455–2462. https://doi.org/10.1109/TPWRS.2015.2462775

Zhu, J., Zhuang, E., Fu, J., Baranowski, J., Ford, A., & Shen, J. (2016b). A Framework-Based Approach to Utility Big Data Analytics. *IEEE Transactions on Power Systems*, *31*(3), 2455–2462. https://doi.org/10.1109/TPWRS.2015.2462775

Zimmermann, A., Schmidt, R., Sandkuhl, K., El-Sheikh, E., Jugel, D., Schweda, C., Möhring, M., Wißotzki, M., & Lantow, B. (2016). Leveraging analytics for digital transformation of enterprise services and architectures. *Intelligent Systems Reference Library*, *111*, 91–112. https://doi.org/10.1007/978-3-319-40564-3_6

Ziora, A. C. L. (2015). The Role of Big Data Solutions in the Management of Organizations. Review of Selected Practical Examples. *Procedia Computer Science*, *65*, 1006–1012. https://doi.org/10.1016/j.procs.2015.09.059

Zorrilla, M., & García-Saiz, D. (2013). A service oriented architecture to provide data mining services for non-expert data miners. *Decision Support Systems*, *55*(1), 399–411. https://doi.org/10.1016/j.dss.2012.05.045

Zuber-Skerritt, O., & Fletcher, M. (2007). The quality of an action research thesis in the social sciences. *Quality Assurance in Education*, *15*(4), 413–436. https://doi.org/10.1108/09684880710829983

CURRICULUM VITAE

Kerem Kayabay

E-mail: kerem.kayabay@metu.edu.tr

ORCID: 0000-0002-3333-4248

Web of Science ResearcherID: ABA-2271-2020

BIOGRAPHY

Kerem Kayabay received his B.S. degree in Computer Science and MBA degree at Bilkent University. During his Ph.D. study in Information Systems, he worked as a system administrator and teaching assistant for graduate-level courses at METU Informatics Institute. Currently, he works as a senior researcher at The Scientific and Technological Research Council of Turkey (TÜBİTAK) as part of the EuroCC project. His primary responsibilities are facilitating industrial and academic access to high-performance computing (HPC) resources in high-performance data analytics and artificial intelligence and increasing HPC adoption in the industry.

AREAS OF INTEREST

Data Science, Big Data, High-Performance Data Analytics, Artificial Intelligence, Digital Transformation, Industry 4.0, Cloud Computing, Data-Driven Organizations, Technology Management, Technology Roadmapping, Data Strategy

EDUCATION

Degree	Institution	Year of Graduation
PhD	Middle East Technical University, Information Systems	2022

MBA	Bilkent University, Business Administration	2014
BS	Bilkent University, Computer Science	2011

WORK EXPERIENCE

Year	**Place**	**Enrollment**
01/2021-Present	TÜBİTAK ULAKBİM, Network Technologies Department, TRUBA HPC Center	Senior Researcher
10/2014-01/2021	Middle East Technical University, Informatics Institute	Research Assistant
06/2017-10/2017	University of Cambridge, Institute for Manufacturing	Visiting Researcher
01/2012-10/2014	Anchora Games	Co-Founder and Full Stack Developer
06/2010-09/2010	Turkish Aerospace Industries, Avionics Department	Part-Time Developer
06/2009-07/2009	Presidency of Strategy and Budget, IT Department	Intern

TEACHING ASSISTANTSHIPS

IS547 Cloud Computing: Technology and Business

IS587 Big Data

IS777 Technology Entrepreneurship and Lean Startups

IS786 Data-Driven Organizations

IS788 Digital Transformation: Management, Technology, and Organization

PUBLICATIONS

Indexed Journals

Kayabay, K., Gökalp, M. O., Gökalp, E., Eren, P. E., & Koçyigit, A. (2022) Data Science Roadmapping: An Architectural Framework Helping Organizations Become Data-Driven. Technological Forecasting & Social Change

Gökalp, M. O., Gökalp, E., **Kayabay, K.**, Gökalp, S., Koçyiğit, A., & Eren, P. E (2022). A Process Assessment Model for Big Data Analytics, Computer Standards and Interfaces

Gökalp, M. O., Gökalp, E., **Kayabay, K.**, Koçyiğit, A., & Eren, P. E. (2021). Data-driven manufacturing: An assessment model for data science maturity. Journal of Manufacturing Systems, 60, 527-546.

Gökalp, M. O., Gökalp, E., **Kayabay, K.**, Koçyiğit, A., & Eren, P. E. (2021). The development of the data science capability maturity model: A survey-based research. Online Information Review

Gökalp, M. O., **Kayabay, K.**, Gökalp, E., Koçyiğit, A., & Eren, P. E. (2021). Assessment of process capabilities in transition to a data-driven organisation: A multidisciplinary approach. IET Software

Book Chapters

Gökalp, E., **Kayabay, K.**, & Gökalp, M. O. (2021). Leveraging Digital Transformation Technologies to Tackle COVID-19: Proposing a Privacy-First Holistic Framework. Emerging Technologies During the Era of COVID-19 Pandemic, 348, 149.

Gökalp, M. O., **Kayabay, K.**, Akyol, M. A., Koçyiğit, A., & Eren, P. E. (2018). Big data in mHealth. In Current and Emerging mHealth Technologies (pp. 241-256). Springer, Cham.

International Conferences

Polat, G., Polat, E. I., **Kayabay, K.**, & Temizel, A., Polyp Detection in Colonoscopy Images using Deep Learning and Bootstrap Aggregation, 3rd International Workshop and Challenge on Computer Vision in Endoscopy (EndoCV2021)

Kayabay, K., Gökalp, M. O., Gökalp, E., Eren, P. E., & Koçyigit, A. (2021, March). Data Science Roadmapping: Towards an Architectural Framework. In 2020 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD) (pp. 1-6).

Gökalp, M. O., **Kayabay. K.**, Gökalp, E., Koçyiğit, A., & Eren, P. E. (2020, September). Towards a Model Based Process Assessment for Data Analytics: An Exploratory Case Study. In European Conference on Software Process Improvement (pp. 617-628). Springer, Cham.

Gökalp, M. O., **Kayabay, K.**, Zaki, M., Koçyiğit, A., Eren, P. E., & Neely, A. (2019, November). Open-source big data analytics architecture for businesses. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-6). IEEE.

Kayabay, K., Gökalp, M. O., Eren, P. E., & Koçyiğit, A. (2018, November). [WiP] A Workflow and Cloud Based Service-Oriented Architecture for Distributed Manufacturing in Industry 4.0 Context. In 2018 IEEE 11th Conference on Service-Oriented Computing and Applications (SOCA) (pp. 88-92). IEEE.

Gökalp, M. O., **Kayabay, K.**, Çoban, S., Yandık, Y. B., & Eren, P. E. (2018, November). Büyük Veri Çağında İşletmelerde Veri Bilimi. In 5th International Management Information Systems Conference (pp. 94-97).

Kayabay, K., Akyol, M. A., Gökalp, M. O., Koçyiğit, A., & Eren, P. E. (2018). Current research topics in industry 4.0 and an analysis of prominent frameworks. Industry 4.0 from the MIS Perspective.

Akyol, M. A., Gökalp, M. O., **Kayabay, K.**, Eren, P. E., & Koçyiğit, A. (2017, September). A Context-Aware Notification Architecture Based on Distributed Focused Crawling in the Big Data Era. In European, Mediterranean, and Middle Eastern Conference on Information Systems (pp. 29-39). Springer, Cham.

Gokalp, M. O., **Kayabay, K.**, Akyol, M. A., Eren, P. E., & Koçyiğit, A. (2016, December). Big data for industry 4.0: A conceptual framework. In 2016 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 431-434). IEEE.

Kayabay, K., Gökalp, M. O., Akyol, M. A., Koçyigit, A., & Eren, P. E. (2016). Big Data for Future Enterprises: Current State and Trends. In 3rd International Management Information Systems Conference, Izmir (pp. 298-307).

Technical Reports

Gökalp, M. O., **Kayabay, K.**, Zaki, M., Koçyiğit, A., Eren, P. E., & Neely, A. (2017). Big-Data Analytics Architecture for Businesses: a comprehensive review on new open-source big-data tools. Cambridge Service Alliance: Cambridge, UK.