DETERMINANTS OF BIG DATA INTEGRATION AMONG THE PUBLIC SECTOR ORGANIZATIONS OF MALAYSIA

ESMAT ABDULMAJEED SAEED WAHDAIN

UNIVERSITI SAINS MALAYSIA

2025

DETERMINANTS OF BIG DATA INTEGRATION AMONG THE PUBLIC SECTOR ORGANIZATIONS OF MALAYSIA

by

ESMAT ABDULMAJEED SAEED WAHDAIN

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

August 2025

ACKNOWLEDGEMENT

First and foremost, all praise be to Almighty Allah for His immense grace and abundant blessings, and for granting me the success to complete this research journey.

My deepest gratitude and appreciation to my beloved parents, who never hesitated to provide me with all forms of support, encouragement, and prayers, enabling me to overcome all obstacles and achieve my aspirations. Additionally, my heartfelt thanks to my devoted wife, who has been a great supporter, and a living embodiment of the phrase: together in good times and bad, and to my daughters and son, whose presence and prayers were invaluable to me during difficult times.

I would also like to express my deep gratitude to my supervisor, Dr. Nasuha Lee Abdullah, for her supervision, encouragement, guidance, constructive feedback, and continuous support. By Allah's grace, her invaluable support was crucial; without it, this challenging academic journey could not have been completed. Sincere thanks also to my former supervisor, Assoc. Prof. Dr. Ahmad Suhaimi Baharudin, for his invaluable support during the initial stages. I am profoundly grateful to my co-supervisor, Dr. Nursakirah Ab Rahman Muton, for her insightful advice and professional comments that significantly enhanced this work. I’m also grateful to my examiners for their invaluable feedback and diligent corrections, which significantly enhanced this thesis.

I extend my deepest gratitude to the Malaysian public sector participants for their invaluable insights, and to the experts who commendably validated this study's framework. School of Computer Sciences-USM staff for facilitating my studies. Finally, to my peer reviewers, lab-mates, colleagues, and friends, I am profoundly grateful for your unwavering support throughout this academic journey.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AHP	Analytic Hierarchy Process Method
ASEAN	The Association of SouthEast Asian Nations
BD	Big Data
BDA	Big Data Analytics
BDAC	Big Data Analytics Capability
BDI	Big Data Integration
BI	Business Intelligence
DC	Theory of Dynamic Capabilities
GOM	Government of Malaysia
ICT	Information Communication Technology
IDC	The International Data Corporation
K-economy	Knowledge-based Economy
M.O.U	Memorandum Of Understanding
MAMPU	Malaysian Administrative Modernisation and Management Planning Unit
MBPP	Penang Island City Council
MDEC	Malaysia Digital Economy Corporation
MENA	Middle East and North Africa Countries
P-economy	Production-based Economy
PMM	Penang Affordable Housing Scheme
RBV	Resource-Based View Theory
TAM	Technology Acceptance Model
TOE	Technology-Organization-Environment Framework

LIST OF APPENDICES

PENENTU INTEGRASI DATA RAYA DALAM KALANGAN ORGANISASI SEKTOR AWAM MALAYSIA

ABSTRAK

Kerajaan di seluruh dunia semakin menggunakan Analitik Data Raya (BDA) untuk meningkatkan prestasi sektor awam dan penyampaian perkhidmatan. Namun, usaha ini dihalang oleh cabaran Integrasi Data Raya (BDI) antara agensi. Walaupun BDI merupakan asas penting dalam kejayaan strategi BDA, kajian ilmiah terhadap penentunya masih terhad. Jurang ini, yang dikenal pasti melalui kajian awal, menjadi asas kepada penyelidikan ini. Tesis ini menyelidik penentu kejayaan BDI dalam organisasi sektor awam di Malaysia, dan membangunkan kerangka komprehensif faktor-faktor yang mempengaruhinya. Reka bentuk kajian kes kualitatif digunakan, dengan data diperoleh melalui temu bual bersama pegawai terlibat dalam BDI serta analisis dokumen awam. Analisis tematik menggunakan ATLAS.ti membolehkan pengekstrakan tema secara sistematik. Tema-tema ini membentuk kerangka yang menjelaskan sifat pelbagai dimensi BDI. Dapatan menunjukkan bahawa kejayaan BDI bergantung kepada empat dimensi utama: teknologi (kerumitan, keserasian, kesediaan data), organisasi (sumber kewangan, sokongan pengurusan atasan, kesedaran), persekitaran (kerjasama, sokongan peraturan/luaran), dan individu (kemahiran analitik, kesediaan berkongsi data). Kajian ini menyumbang kepada literatur dengan menawarkan panduan praktikal untuk pelaksanaan BDA dalam sektor awam.

DETERMINANTS OF BIG DATA INTEGRATION AMONG THE PUBLIC SECTOR ORGANIZATIONS OF MALAYSIA

ABSTRACT

Governments worldwide increasingly use Big Data Analytics (BDA) to boost public sector performance and service delivery. However, this shift is significantly impeded by Big Data Integration (BDI) across agencies. Despite BDI's foundational role in successful BDA strategies and its critical nature, scholarly inquiry into its specific determinants is scarce. This research gap, confirmed by a preliminary study, motivated the current investigation. This thesis systematically investigates the determinants of successful BDI within Malaysian public sector organizations, developing a comprehensive framework of influencing factors. A robust qualitative case study design was employed, gathering empirical data through interviews with BDI-involved officials and public document analysis. Thematic analysis, facilitated by ATLAS.ti, enabled systematic theme extraction and refinement. These themes informed the research framework, elucidating BDI's multidimensional nature. Findings address BDI's role in BDA success and its determinants, categorizing influencing factors into four interconnected dimensions: technological (complexity, compatibility, data readiness), organizational (financial resources, top management support, awareness), environmental (collaboration, regulatory/external support), and people (analytical skills, data sharing willingness). This research enriches literature by focusing on an under-examined area, offering valuable insights for best practices and guiding future BDA implementation in similar governmental contexts.

CHAPTER 1

INTRODUCTION

1.1 Overview

With the emergence of social media, Internet of Things (IoT), and multimedia, the amount of data generated and captured by individuals, businesses, public organisations, industries and scientific research, has increased tremendously in recent years (Agarwal & Dhar, 2014; Barham & Daim, 2020; Goldstein et al., 2021; Sivarajah et al., 2017). These data are of different format be it textual data (either structured or unstructured), or multimedia, and it can be captured and shared on many platforms, e.g. social media sites, ubiquitous-sensing mobile devices (Favaretto et al., 2020; Zhong et al., 2016) traffic cameras, collected sensor data from satellite photos (Abraham, 2022) and machine to machine communications and IoT (Sivarajah et al. 2017). According to The International Data Corporation (IDC) estimates, the globe now produces more data every two days than humans did from the beginning of time until the year 2003. (Goldstein et al., 2021). Similarly, the most recent estimates from Statista indicate that the amount of data created daily is about 328.77 million terabytes, which results in 120 zettabytes of data to be generated in the year 2023 (1 zettabyte equals one billion terabytes), with the prediction for the figure to reach 181 zettabytes in 2025 (Fabio Duarte, 2023). Another figure that reflects the exponential increase in data generated is the example of Twitter, which now generates 50 GB of text data every day, as compared to a few terabytes in the whole year 1997 (Vuong et al., 2020). Data that is created at such an enormous rate is referred to as big data, and it has emerged as a globally recognized trend.

Big Data (BD) and analytics is cited as one of the top strategic technology directions in the last decade, and one of the most significant trends that is transforming organizational processes and operation throughout the world (Arya Bharti, 2021; Sivarajah et al., 2017), With the proven potential of Big Data, corporate investments are increasing significantly (De Luca et al., 2021). Statista, a market leader in providing business data, stated that the BD market size revenue is forecasted to increase exponentially from 7.6 billion USD in 2011, to 103 billion in 2027 (Figure 1.1) (Columbus, 2018), while the International Data Corporation (IDC) reported that the total sales for Big Data and analytics are expected to rise globally from 150.8 billion USD in 2017, to over 210 billion USD in 2020 (IDC, 2017). More recently, (Josh Howarth, 2022) reported that the global market for BDA is currently valued at 274 billion USD, which represents an increase of 62% from its value of 169 billion USD in 2018.

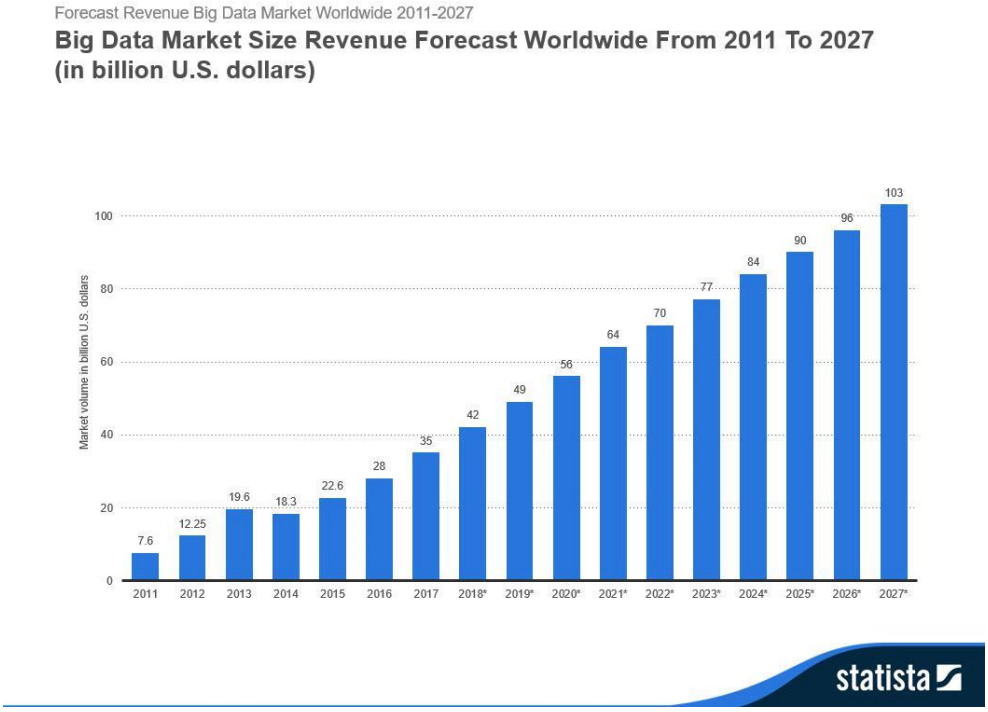

Figure 1.1 BD Market Size Revenues Forecasts Worldwide 2011-2027 (Columbus, 2018)

With this growth rate of Big Data adoption, prompt and cost-effective analytics over Big Data has become a key success factor for many businesses, scientific institutions, and government organizations (Barham & Daim, 2020; De Luca et al., 2021; Haafza et al., 2021; Rodríguez-Mazahua et al., 2016). Therefore, Big Data Analytics (BDA) has become a focus among practitioners, policy makers and scholars (Sheng et al., 2021; Wamba et al., 2017)

Governments can leverage Big Data Analytics to improve their efficiency, effectiveness and citizens well-being through several applications, such as: improving transparency of government agencies, Bureaucratic Reform, minimizing operational losses, sentiment analysis to understand the people's response to certain policies or projects, fraud detection, crime reduction, improve the safety and quality of public transportation, and monitoring and responding to epidemic outbreaks (Aggarwal, 2019; Chatfield et al., 2018; Fosso Wamba et al., 2015a; Janssen et al., 2017; Okuyucu & Yavuz, 2020; Rahmanto et al., 2021; Sheng et al., 2021).

Recently, the COVID-19 outbreak has showcased the significant role that BDA can play in fighting the pandemic and saving human lives. Many governments have incorporated BDA into digital health to improve measures for prevention and control (Al-Ogaili et al., 2020; Azmi et al., 2022). BDA and Artificial Intelligence (AI) are applicable in all the phases of COVID-19 outbreak lifecycle, which are: detection, spread, pandemic management, and recovery (Aziz et al., 2021). Among those BDA's applications are: people monitoring and tracking, monitoring symptom, predicting outbreaks based on previous patterns analysis, quarantine observance, and facilitating treatment and vaccine discovery (Al-Sai et al., 2022; Aziz et al., 2021; Haafza et al., 2021). Moreover, McKinsey Global Institute contends that some sectors are capable to achieve greater benefits from using BD than others and based on the Big Data Value

potential index (an index that combines various quantitative metrics), public sector is among the highest sectors in terms of the potential value from using BD, as shown in Figure 1.2 below.

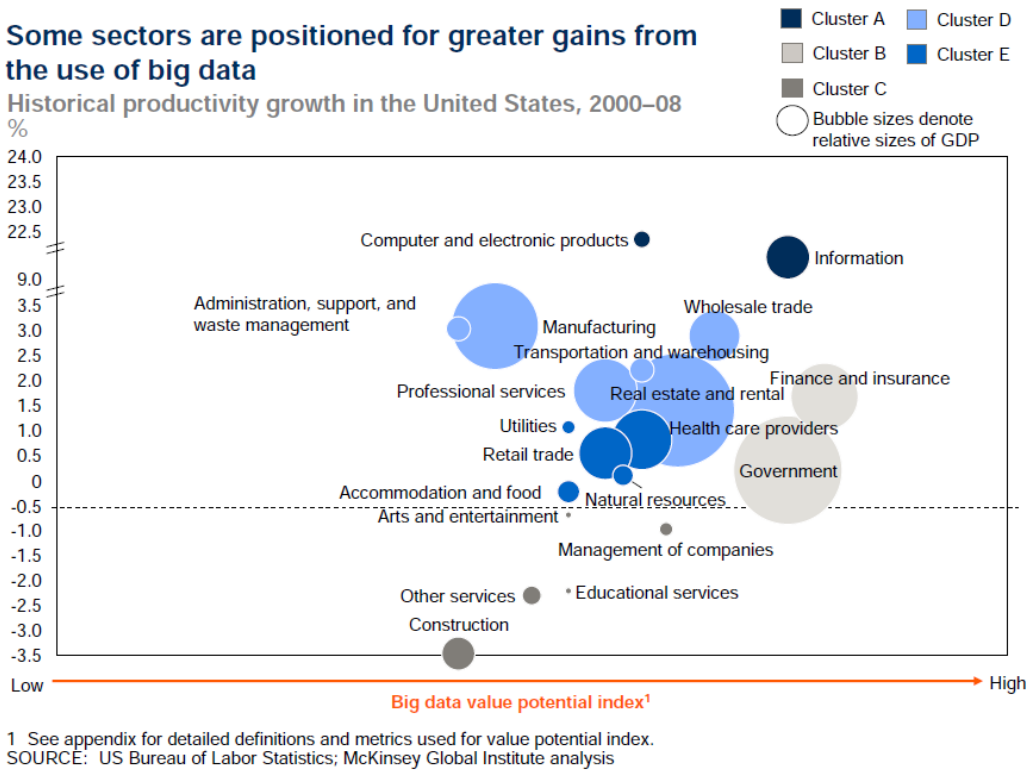

Figure 1.2 Potential of BD for the Government Sector, McKinsey Global Institute Analysis (Meikle, 2014)

As many governments around the world try to seize the BDA opportunities, its application in the public sector is emerging (Aggarwal, 2019); however, adopting Big Data Analytics in the public sector is a difficult task that involve some uncertainty and impose additional challenges that other sectors do not need to deal with (Kim et al., 2014; Klievink et al., 2017; Okuyucu & Yavuz, 2020; Rahmanto et al., 2021). This complex nature of applying BDA in public organizations calls for more research to be dedicated to investigating the issues and challenges related to implementing BDA in the

public sector (Aboelmaged & Mouakket, 2020; Cecilia Fredriksson et al., 2017; Chatfield et al., 2018; Okuyucu & Yavuz, 2020; Sheng et al., 2021; Talib et al., 2023) Among the challenges that face public sector BDA efforts, the issue of cross agency integration rises as one of the most important, common challenges that worth researching (Aggarwal, 2019; Anna & Nikolay, 2015; Hamzah et al., 2020; von Sanden & Neideck, 2021).

Despite the research need referred to by earlier studies, the topic is still growing and has not yet fully established, and further research is needed to offer a better understanding of how to adopt BDA to create value, taking into consideration the differences between various contexts (Aboelmaged & Mouakket, 2020; Alalawneh & Alkhatib, 2021; Guenduez et al., 2020; Lutfi et al., 2023).

Additionally, there is an imbalance in the existing literature, as the majority of extant literature is quantitative in nature (Aboelmaged & Mouakket, 2020). Moreover, a significant portion of the previously published studies have investigated BD and BDA in private organizations; however, in many respects, business analytics cannot support the application of analytics in the public sector, which creates the need for more research to be dedicated for the public sector's use of BD and BDA to fill in this gap (Cecilia Fredriksson et al., 2017; Lutfi et al., 2023; Okuyucu & Yavuz, 2020; Talib et al., 2023)

Even more, this growing, yet insufficient academic scholarship on BD in the public sector focuses mainly on the public sector of most developed countries, mostly in Europe and North America, with much fewer studies are dedicated to the developing countries' context (Alalawneh & Alkhatib, 2021; Jonathan & Kuika Watat, 2023; Joubert et al., 2023).

Because of the contextual differences between developed countries and developing countries, results from studies like the above-mentioned examples, cannot

be generalized to the case of developing economies (Agrawal, 2015; Alalawneh & Alkhatib, 2021; Furtado et al., 2023; Jonathan & Kuika Watat, 2023; Joubert et al., 2023)

Among the issues that are pertinent to Big Data Analytics adoption in the public sector, Big Data Integration (BDI) is considered a core element in BDA strategy, and it is one of the most challenging steps (Azmi et al., 2022; Coulthart & Riccucci, 2021; Fillinger et al., 2019; Olama et al., 2014). The organizations' capacity to integrate data is an essential element for Big Data Analytics (Dong & Rekatsinas, 2018; Fillinger et al., 2019), and it plays key role in improving the organizations' efficiency (Aberdeen, 2020; Kadadi et al., 2014). Therefore, this study will try to shed some light on this important issue by studying its different aspects in the context of public organizations of Malaysia.

The next section delves into this topic within the Malaysian public sector context, and then the following section, Section 1.3, discusses the preliminary study conducted at the early stage of this research, explaining the rationale behind it and presenting its findings. After that, the research problem is highlighted in Section 1.4, followed by the research objectives and research significance in Sections 1.5 and 1.6, respectively. Finally, the scope of the study is defined, and the structure of the thesis is described.

1.2 Background

Malaysia has one of the world's most open economies, with a trade-to-GDP ratio that has averaged more than 130% since 2010 (The World Bank, 2022). In terms of GDP, the Malaysian economy ranks as the fifth largest in Southeast Asia and holds the 38th position globally (International Monetary Fund, 2021), and recently, the Malaysian

economy has been ranked as the 27th most competitive economy in the world in the Global Competitiveness Report of 2023 (Bernama, 2023).

To realize transformation towards a developed nation, Malaysia is eager to achieve significant economic and technological progress. In line with that, the Malaysian Administrative Modernisation and Management Planning Unit (MAMPU) was established in 1977 to improve professionalism in public services. Among its major roles is to lead the development of ICT for the public service sector, and to provide consultancy in management organisation and ICT for the public service sector (Ijab et al., 2017; MAMPU, 2016b).

The governmental interest in leveraging technology to achieve transformational goals was reflected in the consecutive Malaysian plans, including the latest 12th plan initiated in 2021 (ITA, 2021). Moreover, in 2021 the government of Malaysia (GOM) launched the MyDigital initiative as a holistic approach to anchor the Malaysian digital economy by 2030.

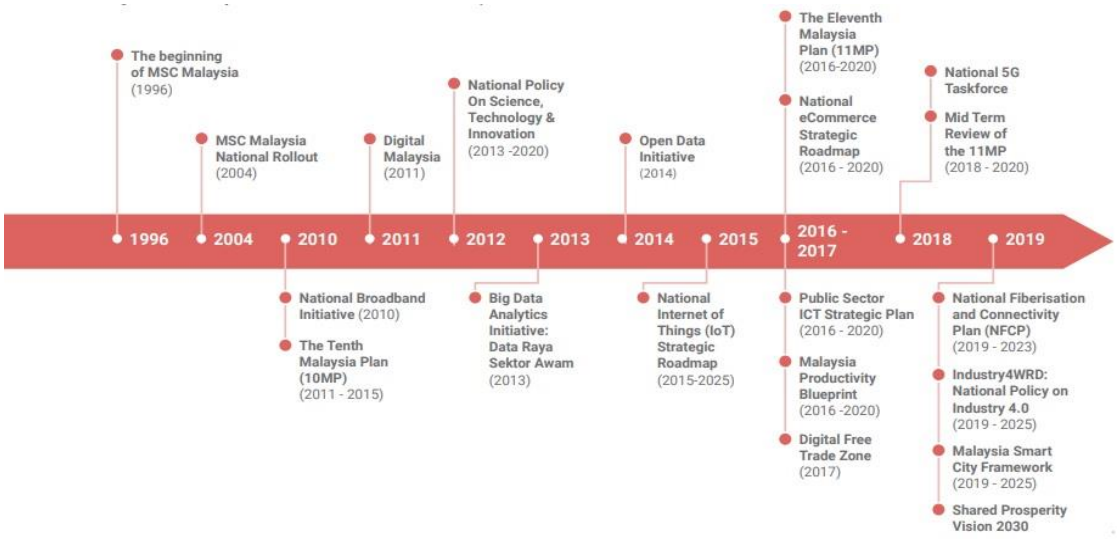

Figure 1.3 The Evolution of ICT Related Policies in Malaysia.

MyDigital aims to serve the GOM plans to transform Malaysia to a digitally driven economy, a digital economy leader in the region, and a nation with high average income. (Figure 1.3 shows the evolution of GOM policies related to ICT). As a result, the ICT industry becomes one of the fastest-growing in the Malaysian market, accounting for 19.1% of the country's GDP in 2019 and predicted to expand to 22.6% by 2025 (Economic Planning Unit, 2021). This ICT industry growth, supported by the GOM policies that encourage the adoption of digital technologies has led to massive data generation due to the increased adoption rates of different digital technologies such as: smaller ICT devices and gadgets, growing amounts of interactive multimedia content based on mobile technology, cloud computing, IoT devices, and social media.

With the massive rise in data generation due to the above-mentioned technological changes, big data analytics has become crucial for both public and private organisations to create value for effective decision making, and to improve productivity and innovation (Barham & Daim, 2020; De Luca et al., 2021; U.S. Embassies abroad., 2017). The Government Chief Information Officer also declared big data analytics (BDA) as one of three key priorities for ICT in the public sector of Malaysia and was incorporated into the Eleventh Malaysia Plan (Big community, 2015). Accordingly, four government BDA pilot projects were initiated. These projects covered: sentiment analysis, crime prevention, infectious disease prevention, and price watch (Anna & Nikolay, 2015; Hamzah et al., 2020).

In accordance with the announced plan, Malaysia managed to complete the first phase of BDA initiative in 2016 by implementing the four pilot projects successfully (Hamzah et al., 2020; MyGovernment, 2019). The next phase was the expansion of public sector BDA projects, in which 12 new projects were implemented in 12 different agencies in collaboration with MAMPU.

With the emergence of COVID-19, the need for BDA to confront the outbreak and mitigate its impact has increased (Azmi et al., 2022; Mohd Arif & Choo Ta, 2022). Malaysia was among the first nations in Southeast Asia to develop digital applications for managing and controlling the COVID-19 epidemic; therefore, it has been ranked among the top 50 countries in the United Nation's E-Government Development Index for the year 2020. Among those applications that was mandated by GOM is MySejahtera, which is a QR based mobile app for contact tracing. MySejahtera registers the daily Check-Ins to keep track of the users' movement history, along with the health assessments reported by the users(Mohd Arif & Choo Ta, 2022; MySejahtera Unit, 2022).

The application includes other features such as: the daily update of COVID-19 cases, hot spots tracking, and health declaration for travellers, among others. Information about confirmed COVID-19 cases from MySejahtera, along with the reports from other health providers are registered in e-COVID-19, which is a web-based system for disease reporting and surveillance. This information is used by the National Public Health Laboratory and Crisis Preparedness and Response Centre (CPRC) to plan and allocate hospital resources as needed based on the case reports(Azmi et al., 2022). Moreover, by analysing the data gathered from e-COVID-19 and other platforms and systems, the MOH was able to confront the pandemic by issuing daily case updates, imposing the suitable SOPs (Standard Operating Procedures), conducting citizens' sentiment analysis, and predicting future COVID-19 cases (Azmi et al., 2022; Mohd Arif & Choo Ta, 2022).

Turning to analytics to face the pandemic provides additional evidence of the significance of BDI. Health authorities leveraged BDA to analyse the data integrated from multiple systems originated from different agencies to overview and monitor the

COVID-19 situation in real time (Aziz et al., 2021; Azmi et al., 2022; Mohd Arif & Choo Ta, 2022). Furthermore, the need for inter-agency partnership has been a necessity to facilitate the application development; and MySejahtera as an example was the result of a collaboration between multiple ministries and agencies, including MOH, MAMPU, Ministry of Science, Technology, and Innovation (MOSTI), and Malaysian Communications and Multimedia Commission (MCMC), and the National Security Council (NSC).

Today, BDA maintain its position as a key priority for the GOM as it represents one of domains of the Fourth Industrial Revolution (4IR), and a leading ICT sector to achieve the GOM's transformational plans. (Economic Planning Unit, 2021; ITA, 2021). However, despite this government eagerness to leverage BDA to achieve the desired transformations, the full potential of BDA adoption and usage in the public sector of Malaysia remains largely unrealized (Fan et al., 2021; Harun et al., 2022). That is, regardless of achieved successes, BDA initiatives still struggle to overcome the challenges that usually hinder BDA adoption in public sector, such as data integration and sharing among agencies (Aziz et al., 2021; Ijab, Surin, et al., 2019; Unit, 2021), and the shortage of analytical skills (BIGIT, 2021; Zulkarnain et al., 2019).

Problems in big data integration among different systems in the public sector occur due to the diversity of data types being integrated and the lack of harmonization between various systems (Shah et al., 2024; Zian et al., 2024). Several previous studies refer to the significance of the issue of BDI in the Malaysian context, highlighting the need for more research to investigate this topic in the Malaysian context (BIGIT, 2021; Harun et al., 2022; Saadan et al., 2024).

Furthermore, the BDI issue has not been covered sufficiently in literature as will be discussed further in this research, resulting in the lack of high-level BDI frameworks

that can guide public sector organizations in their transformation into data-driven organizations (Shah et al., 2024).

1.3 Preliminary Study

The main objective of this study is to explore and confirm the research problem by exploring the study domain in depth. Due to the shortage of published studies that cover the topic of BDA in the public sector of developing countries similar to Malaysia, and because of difficulty to find sufficient documentation about the status of BDA in the Malaysian public sector, this qualitative preliminary study is obligated to be conducted, in order to explore the study domain, and to confirm the research problem.

In such exploratory researches whereby there is an emergent phenomenon, (in the case of this study adopting BDA in the public sector), qualitative investigation is preferred to explore the different aspects of the phenomenon (Schutt et al., 2018; Shufutinsky, 2020).

Despite the fact that collecting data qualitatively is much more time consuming than following the quantitative approach, the information gathered qualitatively are richer information, and the insights gained about the phenomenon under investigation, are deeper (Awan et al., 2023; DeVaney et al., 2018; Hazri Jamil, 2015b) . The coming subsections will describe in detail the methodology followed to conduct this preliminary study, showing the different stages of the study, as well as the findings of it.

1.3.1 Preliminary Study Methodology

Firstly, the interview guide was developed after deep discussions between the researcher and the research supervisor. Because of the exploratory nature of the study, semi-structured type has been selected due to its flexibility, which is advantageous in

this type of exploratory inquiry (Creswell, 2014b; Ruslin et al., 2022), hence, many of the questions are open ended to allow the interviewee to explain the issues in detail.

The interview guide's questions are designed to get as much information as possible in regard with the research problem. The interview questions are aimed to serve two purposes; the first is to confirm the existence, and significance of the research problem, and secondly, to explore the antecedents of that problem.

The following questions are directed to the informant:

1- We read about the 4 pilot BDA projects launched by the Malaysian government, at what stage are they now?
2- What is current level of BDA adoption among Malaysian public organizations, how many organizations have implemented BDA?
3- Are you satisfied with the pace of the implementation process?
4- What are the main obstacles that hinder BDA diffusion among public Malaysian agencies?
5- What are the measurements taken by MAMPU to overcome those obstacles?
6- Drawing on your experience, what seems to be the most influential facilitators?
7- Do you consider BDI as one of the main issues in BD projects?
8- What are the reasons of data integration difficulties, or what are the factors that affect BDI?
9- What are the solutions for that issue?

For the sake of this exploratory stage of the research, the key characteristic of the potential informant is to have a holistic view of the BDA efforts in the Malaysian public sector; this is to get the most useful insights about the ongoing BDA projects by interviewing the least possible number of informants.

As mentioned earlier, there is lack of documentation about the on-going BDA projects in Malaysia, which makes it difficult for any researcher to have a clear image about those projects and the related issues that worth study.

Therefore, the required informant must be one of the higher positions' members of the governmental bodies that lead the BDA initiatives in Malaysia's public sector, hence, MAMPU is the government department that leads BDA dissemination among ministries and departments. For that, it is decided that the key informant in this stage must be one of MAMPU's high position officials whom his/her job involves planning and/or leading and/or supervising BDA projects in the public sector organizations, and she was interviewed in her office at MAMPU.

Moreover, to enrich the perception about the research topic and research problem, the researcher felt the necessity to explore the BDA-related issues from different points of view and at different levels; thus, the researcher decided to interview one of the officials who has the same characteristics as the first informant, but at the state level this time. The second respondent has been selected based on the information obtained from Penang's state government website, along with the recommendation from the first informant who is knowledgeable about the state level BDA projects as well, hence, the second informant is from the ICT department at Penang Island's state government secretary office, located at George Town.

In total, three interviews are conducted at this stage, two of them with the first informant from MAMPU, and one interview with the informant from Penang state ICT department, the interviews were conducted between January 2018, and October 2018.

1.3.2 Findings of the Preliminary Study

The following broad themes are identified from the respondents' replies, namely: the status of BDA adoption, factors influencing BDA projects, the efforts taken to facilitate BDA adoption, and the role of big data integration.

Theme 1: Status of BDA Adoption

Theme 1 (Status of BDA Adoption) represents the information shared by respondents about the ongoing BDA adoption efforts in the Malaysian public sector. It includes statistics of completed projects, information about project initiation, and plans for future projects.

In general, the shared statistics shows that the adoption level is slower than expected, the following table summarizes the key ideas extracted under this theme:

Table 1.1 key Ideas Extracted from Theme1

Theme	Example Data Extract	Conclusions
Status of BDA adoption	- "After the completion of the 4 pilot projects, another 12 projects were initiated in 12 different agencies". 1st interview. - "…12 agencies come onboard, but only six agencies managed to complete the BDA project, the other six are half the way, three quarter the way…". 1st interview - "BDA was initiated in Penang in 2016 with the help of MAMPU", and "Six projects were initiated, and at the time of the interview(in 2 years time) two of them were completed successfully". 3rd interview.	- BDA adoption pace is slower than expected. - talking about status is pertinent to the talk about the difficulties associated with BDA in public sector, or about the requirements that public agencies need to fulfil to succeed in BDA adoption.

These findings go in line with the literature that confirms the uncertainties related to BD adoption in public sector, and the effect of those uncertainties on both: the adoption rates, and adoption successes (Anna & Nikolay, 2015; Klievink et al.,

2017; Kuoppakangas et al., 2019). This is also consistent with BDA literature that discussed the special challenges that BDA adoption brings to public organizations (Chatfield et al., 2018; Gamage, 2016; Heeks et al., 2021; Kim et al., 2014; Okuyucu & Yavuz, 2020).

Theme 2: Factors Influencing BDA Projects.

Talking about barriers and facilitators constitute a considerable part of each of the conducted interviews. The participants have discussed different issues, hence, the open-ended questions have allowed for more elaboration from the participants which serves the exploratory nature of the interviews and resulted in discovering some issues that are not covered in BDA literature. For example, the issue of employees' training discontinuation is referred to as one of the main issues that affect competency development efforts. The factors shared by participants are different in nature, which reflects the multifaceted nature of BD that have been emphasized by many studies, such as (Agrawal, 2015; Farzaneh et al., 2018; Löfgren & Webster, 2020; Motau & Kalema, 2016; Raguseo et al., 2018). Those factors are categorized as: technological, organizational, environmental, and people related factors.

The following Table 1.2 summarizes the main factors that influence BDA implementation and their categories.

Table 1.2 Categories and Subcategories of the Influential Factors

Theme	Category	Subcategory
Factors Influencing BDA Projects in Malaysia	Technological	readiness of data in digital format
	Organizational	- funding - top management support
	Environmental	- legal constraints on data sharing among agencies.
	People	- shortage of analytical skills.

Theme 3: Efforts Taken to Facilitate BDA Implementation.

This theme summarizes the measures taken by the interviewees' organizations to pave the way for successful BDA implementation. Those efforts are either to overcome the difficulties encountered by BDA projects, or to fulfil some requirements needed by BDA projects. An example of the former is the coaching program that was established by MAMPU to solve the problem of data analytics skills shortage, whereas conducting workshops that target the top management of all departments to raise their awareness of BDA is an example of the latter.

The following quote from the 1st interview highlights the efforts on competency development: *"We call agencies to come on-board and we monitor their progress until they are able to develop their BDA"*. While the following quote is from the 3rd interview *"A workshop was conducted with the participation of the top management of all departments to raise the awareness about BDA importance and potentials"*.

Theme 4: The Role of Data Integration in BDA Success

The issue of integrating data among different agencies, abbreviated as Big Data Integration (BDI), has occupied significant space of the respondents' discussions. They refer to it as one of the main challenges that face BDA, and they also reveal some factors that cause that challenge. Furthermore, BDI is recognized by some respondents as a key requirement and an important success factor for BDA projects, such as the interviewee in interview #1 who commented *"in some cases, the visualization was poor because not all the required data was available, some involved agencies cannot share data with other parties"*.

Moreover, BDI is also present when the respondents describe the efforts made to disseminate BDA in public agencies, for instance, the following quote is from the 1st interview *"We are moving towards issuing an act to enhance data sharing among different agencies, the current act that puts some constraints on data sharing is dated back to the seventies".* Hence, the issue of BDI is connected to all the other themes, and it constitutes a large part of the respondents' discussions, which reflects its pivotal role in BDA successful adoption, and that agrees with the literature that emphasized the cruciality of BDI, such as (Aberdeen, 2020; Dong & Rekatsinas, 2018; Kadadi et al., 2014; Lutfi et al., 2023; Patel, 2019), and confirms the significance of investigating BDI determinants. In conclusion, the results of the preliminary study are beneficial for the research. It provides useful insights about the BDA adoption status in the Malaysian public sector and helps in identifying number of BDA influential factors in the Malaysian context, which is one of the study's objectives. More importantly, the initial findings confirm the significance of BDI as a worth studying research problem, and as a key determinant of the successful BDA adoption. Finally, the findings shed some light on the factors that influence BDI in the Malaysian context, which is helpful in developing the research framework in the next research stages.

1.4 Problem Statement

Among the issues that face BDA adoption in the public sector, the issue of big data integration and sharing data among different agencies raises as one of the pivotal issues that has been highlighted as a core element of BDA strategy, and one of the major challenges for any BDA project (Fillinger et al., 2019; Olama et al., 2014; Putrama & Martinek, 2024; Shah et al., 2021). Without proper integration, data from different government agencies and departments remains in silos (Abdullah et al., 2017; Bruintjies

& Njenga, 2024; Shah et al., 2021; Yahya et al., 2019), in contrast, the value of data explodes when it can be linked to and combined with other data from other sources (Aberdeen, 2020; Dias et al., 2021; Dong & Srivastava, 2013; Patel, 2019; Sonia Johnson, 2019). The importance of BDI increases with the constantly expanding data environments, and the increasing heterogeneity of data sources (Aberdeen, 2020; IBM Corporation, 2020; Karanam et al., 2021; Michael Lock, 2021; Rozony et al., 2024), which signify the essentiality of overcoming the Big Data Integration challenge in order to achieve high levels of information quality (Valle & Kenett, 2018; Dias et al., 2021; Lutfi et al., 2023; Shah et al., 2021), and fully realise the potential of Big Data (Dias et al., 2021; Dong & Srivastava, 2013; Löfgren & Webster, 2020).

Furthermore, data integration is the most significant impediment to organisations attempting to become data-driven through the use of BD, AI and machine learning (Bruintjies & Njenga, 2024; Liu & Hsieh, 2017; Patel, 2019), and lack of integration leads to poor decision making, which has a negative impact of the efficiency of the organization (Corsi et al., 2021; Kadadi et al., 2014; Munné, 2016; Patel, 2019; Xiao et al., 2020).

Moreover, BDI holds great importance in the context of public sector, not only because the public sector produces unprecedented amounts of data from different sources (Ali et al., 2016; van Ooijen et al., 2019); but also due to their cross departmental nature of work. This kind of work requires data to be transferred across institutions, states, or even countries (Ali et al., 2016; Corsi et al., 2021; Löfgren & Webster, 2020), which in turn, makes data integration a challenging task (Aggarwal, 2019; Hassan & Ahmad, 2019; Putrama & Martinek, 2024; Sanden & Neideck, 2021). In fact, BDI is more challenging in public sector than in other sectors (Ali et al., 2016; Hassan & Ahmad, 2019; Kadadi et al., 2014), due to difficulties in data collection

(Hassan et al., 2020; Kadadi et al., 2014), the massive amount of data to be integrated (Munné, 2016; Shah et al., 2024), stricter data governance policies (Hassan et al., 2020; Kingsley et al., 2018), existence of data silos (Aggarwal, 2019; Bourdeau et al., 2019; Shah et al., 2024).

Several studies have reported the significance and difficulties of data integration in the public sector (Aggarwal, 2019; Anna & Nikolay, 2015; Desouza & Jacob, 2017; Gamage, 2016; Hassan et al., 2020; Janssen, Konopnicki, et al., 2017; Karanam et al., 2021; Kim et al., 2014; Klievink et al., 2017; Morabito & Morabito, 2015; Shah et al., 2021, 2024; Smith et al., 2023; Zulkarnain et al., 2019).

Despite that, BDI is still one of the issues that are rarely covered in detail as Aboelmaged & Mouakket, (2020) reported in their bibliometric study, which agrees with (Xiao, Zhu, et al., 2020) who contend that the studies investigating the factors affecting BDI in organizations are minimal, and Putrama & Martinek, (2024)'s systematic review that highlighted many gaps in the BDI area, as well as several unresolved necessitating further investigation. Also, although technical developments have potentially made data integration and the pursuit of deeper insights possible, this is not yet a seamless practice (Aberdeen, 2020; Butts-Wilmsmeyer et al., 2020).

More importantly, the issue of data integration is very relevant to the case of Malaysian public sector. According to Ijab, Surin, and Nayan (2019), the lack of data integration and sharing among different agencies in the Malaysian public sector is the reason behind the relatively low ranking of Malaysia in the open data barometer, comparing to its neighbouring countries such as the Philippines, Singapore, and Indonesia. The significance of this issue has been emphasized by the preliminary study's interviews with the Malaysian public sectors' officials, and it is referred to by different

reports and publications. For example, Dr.Yusminar Binti Yunus, the director of Digital Government Division at MAMPU highlights that lack of readiness to share data among agencies is one of main challenges for the government's ongoing BDA efforts (Yunus, 2018). This is also supported by other research works such as Hamzah et al., (2020) that confirms the need for more work to improve the BDI status in the Malaysian public sector. Moreover, the Malaysia's AI Blueprint annual report listed lack of quality data, and lack of data sharing policies as one of the key hurdles of BDA maturity in Malaysia (BIGIT, 2021). Likewise, the Twelfth Malaysia Plan (from 2021 to 2025) states: "*Limited data sharing and lack of big data analytics (BDA) among agencies affect overall development planning and implementation.*", among the issues that lead to ineffective implementation and governance (Unit, 2021). The plan has also attributed the fragmented development plans in some states to the absence of data integration and inefficient coordination between different government agencies. Therefore, improving data integration is mentioned as one of the strategic priorities to achieve inclusive digitalization, and enhance service delivery efficiency (Unit, 2021).

More recently, Saadan et al. (2024) investigated the existing inefficiencies in the e-payment process in the Malaysian public sector, and attributed it to the limitations existing in the current system which is characterized by the lack of effective integration. The authors quoted the 2022's report of the audit general which stated that over RM 6.0 billion of funds collected by the government were delayed due to the inefficient data integration among the banks and collection agents.

Moreover, the insufficiency of literature in this research area is manifested by the shortage of studies that present comprehensive big data integration frameworks that encompass both the facilitators and barriers of successful BDI in the public sector; such frameworks are necessary to serve as guidelines helping public sector organization in

paving their way towards leveraging BD optimally (Orenga-Roglá & Chalmeta, 2018; Shah et al., 2024).

Because of the importance of the issue of BDI from the academic perspective as one of the most important, yet least studied topics in the area of BD, and from the practical perspective as one of significant determinants of the success or failure of any BD project, this research will investigate the factors that influence BDI, as well as the role BDI plays in successful BD adoption.

Based on the issues highlighted above, the following research questions have been developed:

1. What are the issues hindering to BDA adoption, particularly regarding data integration, in the public sector of Malaysia?
2. What are the factors that influence Big Data Integration (BDI) among government agencies in the Malaysian public sector?
3. What is the most suitable framework for successful Big Data Integration (BDI) in the context of the public sector?
4. How to validate the proposed BDI framework?

1.5 Research Objectives

This study aims to achieve the following objectives:

1. To explore the issues hindering BDA adoption, with an emphasis on BDI issues, in the public sector organizations in Malaysia.
2. To identify the factors that influence Big Data Integration (BDI) among government agencies in the Malaysian public sector.

3. To formulate a Big Data Integration framework that includes all influential factors.

4. To validate the proposed BDI framework using expert review method.

1.6 Significance of the Research

From the academic point of view, the current research will try to fill in a literature gap by investigating the challenges that face BDA adoption among the public organizations of one of the fast-growing Asian economies, and a leading Association of South East Asian Nations (ASEAN) country in terms of using ICT in the government sector: Malaysia. Also, the study will focus on the issue of big data integration (BDI) which is one of the least studied topics in the area of BD, and a key determinant of successful BDA implementation, and a major, common challenge for every public sector's BDA project, including Malaysia's public sector.

From the practical perspective, it is very significant to investigate the challenges related to BDI from the point view of the people who are involved in the initial stage of BDA adoption in Malaysia, and to identify factors that influence BDI and BDA as a result, as well as the lessons learned so far. This is important not only because of the proven significance of these issues in the Malaysian context, but also because of their prominence as a common challenge for every BDA project.

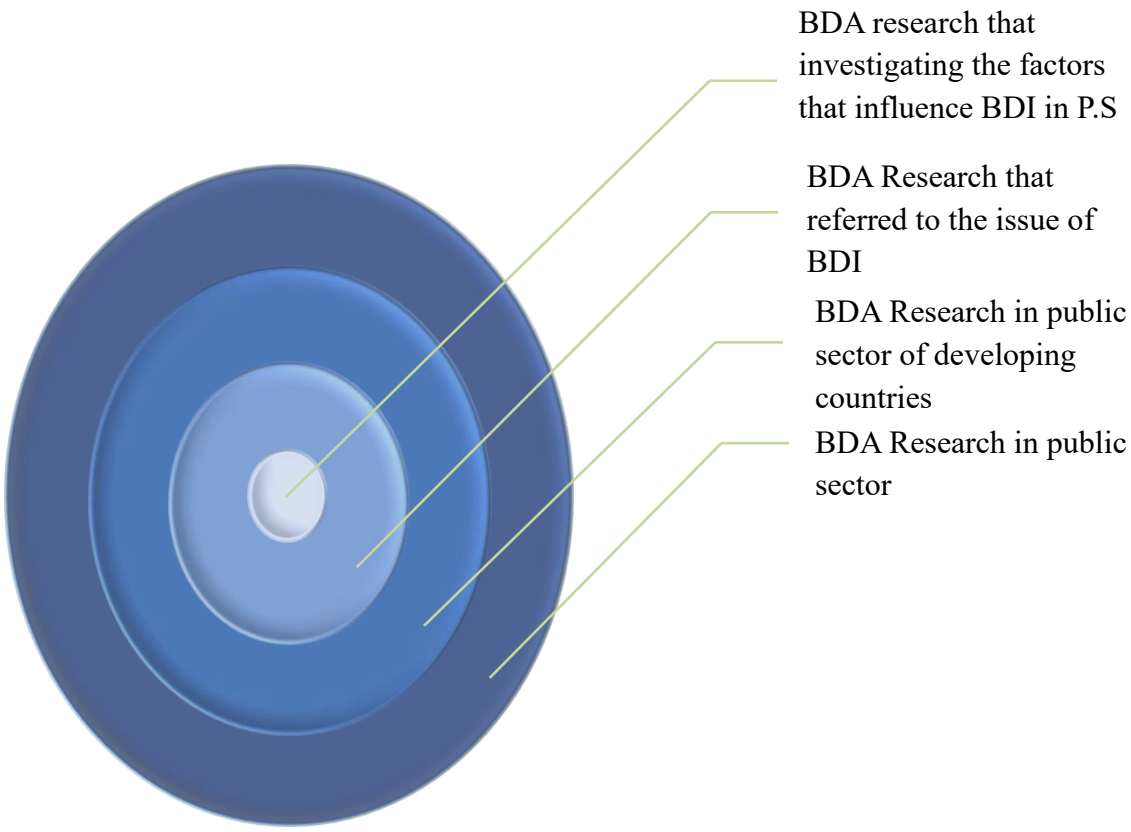

Figure 1.4 Research Gap.

1.7 Scope of the Study

The study domain for this research is the public sector of Malaysia, hence, the targeted organizations must be involved in implementing BDA projects. Those organizations can be either federal level organizations or state level.

As for the respondents, they must be public sector employees (technical or subject matter experts) who have been engaged in the BDA projects in their organizations or other public organizations in Malaysia.

1.8 Definition of Terms

The terms used frequentlay in this thesis are described below:

Big Data Integration (BDI) represents "the advanced data integration processes developed to manage the enormous volume, variety, and velocity of big data, and combines this data from sources such as web data, social media, machine-generated data, and data from the Internet of Things (IoT), into a single framework" (OmniSci, 2020)

Complexity is defined as "Degree to which Big Data Integration is perceived as relatively difficult to understand and use" (Verma & Bhattacharyya, 2017).

Compatibility indicates "Whether an innovation is compatible with current: practices, values, and needs" (Sam & Chatwin, 2019).

Data readiness refers to "The existence of data in digital format that is ready to be handled by the BDA software" influenced by (Neil D. Lawrence, 2017).

Financial resources addresses "Whether the organization has adequate financial resources to support the expenses of BDI implementation and ongoing operation"(Sam & Chatwin, 2019).

Top management support is defined as "the degree to which top management understands the importance of BDI and is involved in BDI activities" (Verma & Bhattacharyya, 2017)

Awareness explains "the level of managerial understanding of the ultimate benefits of data integration to reap the benefits of BDA" (Meikle, 2014)

Regulatory support specifies "Whether the existing policies encourage the organizations to embrace BDA by providing supportive legal environment that support