

Managing Data Science Initiatives as Exploratory Projects – A new approach to Program Management

by Sandeep Mathur

Bachelor of Technology (Industrial Engineering)

Master of Technology (Industrial Engineering)

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of

Dr Shankar Sankaran

Dr Samuel MacAulay

Dr Ivor Tsang

University of Technology Sydney
Faculty of Design, Architecture and Building

June 2024

Certificate of Original Authorship

I, *Sandeep Mathur*, declare that this thesis is submitted in fulfillment of the requirements for the award of *Doctor of Philosophy*, in the *Faculty of Design, Architecture and Building* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed
prior to publication.

Signature: Sandeep Mathur

Date: 30th June 2024

Keywords

artificial intelligence
benefits realisation
business case
change management
data management
data science
deep learning
delivery framework
exploratory projects
innovation
machine learning
metrics
portfolio management
program management
scaled agile
value stream

Acknowledgments

Growing up in an academic family, I have always been inspired to pursue a PhD to bridge the gap between practice and academia in my profession. My father, a retired professor of botany, and my three siblings, all with PhDs in the sciences, instilled in me a passion for higher education. However, my busy schedule as a full-time portfolio and program manager, combined with volunteering work, left little time to pursue this dream. I had thought I would have to wait until retirement to make it a reality.

I had the opportunity to meet Dr Shankar Sankaran while serving as a board member of the Project Management Institute Sydney Chapter. During a conversation, I shared my desire to pursue a PhD with him. He suggested that data collection would be easier while working for an organisation rather than in retirement. Later, I met Dr Sam MacAulay who encouraged me to submit a PhD proposal. My experiences with the challenges of delivering data science initiatives (DSIs) at Transport for NSW (Transport) provided me with a relevant topic for my research. To integrate data science and program management, Shankar suggested that I talk to Dr Ivor Tsang, leading to my having three supervisors, each bringing a unique perspective to my research. They have continuously supported me throughout my PhD journey.

As a practitioner using agile methods for delivery, I approached my PhD project by breaking it down into sprints, a method that Shankar initially referred to as "PhD by sprints". While Sam was initially supportive of this approach, he grew increasingly sceptical of the pace at which I was trying to complete my PhD. He encouraged me to slow down and to spend more time thinking and reflecting rather than approaching the project as a task-oriented project that needed to be completed quickly.

The outbreak of COVID-19 in 2020 brought about a sudden change in circumstances, with everything going into lockdown. I found that, in this unusual situation, I was able to devote more time every morning to reading and writing, time which I would otherwise have spent commuting to work. Essentially, the pandemic created an opportunity for me to use this newfound capacity to further my PhD journey.

Thanks to professional editor, Hazel Baker, who provided copyediting services according to the IPEd guidelines for editing research theses.

I would like to express my gratitude to my wife Ruchi for her unwavering support throughout my PhD journey and for taking on additional responsibilities at home. I am also grateful for the encouragement and support from my daughters Utsa and Anuva. I also want to thank my teams at Transport for their participation in trying out and refining the DSI delivery framework. I am excited for the next chapter in my career and to see where this achievement takes me.

About the Researcher

Sandeep Mathur is a seasoned professional with more than 20 years of experience in project, program and portfolio management. He currently holds the position of Head of Data & Engineering at Greenpeace Australia Pacific, where he is developing and leading data and digital systems strategies and overseeing digital transformation of the organisation.. He is an Industry Fellow at the University of Technology Sydney (UTS), Chair of National Diversity and Inclusion Council at the Australian Computer Society (ACS) and a Fellow of the Australian Institute of Company Directors (FAICD). His prior experience includes roles at organisations such as Transport for NSW, First Data, Westpac, HSBC, Commonwealth Bank, Deutsche Bank, Société Générale, IBM, Fujitsu Australia and New Zealand and Tourism Australia.

He is a certified program management professional (PgMP) and project management professional (PMP) of the Project Management Institute, a certified SAFe 6 program consultant (SPC) and an agile project management (AgilePM) practitioner, a certified management of portfolio (MoP) practitioner and a certified Prosci change management consultant. He has also made significant contributions to the field of project, program and portfolio management through his authorship of or contribution to standards published by Standards Australia and the Project Management Institute.

Table of Contents

Certificate of Original Authorship	ii
Keywords	iii
Acknowledgments.....	iv
About the Researcher.....	vi
Table of Contents.....	vii
List of Figures.....	xi
List of Tables	xiv
List of Abbreviations	xvi
List of Publications and Courses.....	xvii
Abstract.....	xx
1. Introduction	1
1.1 Background.....	2
1.2 Context.....	2
1.3 Purpose	5
1.4 Research Questions.....	5
1.5 Significance, Scope and Definitions.....	6
1.6 Thesis Outline.....	10
1.7 Conclusion.....	11
2. Research Issues	12
2.1 Exploratory and Exploitative Projects	13
2.2 The Case for Data Science Initiatives.....	16
2.3 The Governance of Data Science Initiatives.....	22
2.4 The Delivery of Data Science Initiatives.....	33
2.5 Conclusion.....	41
3. Research Design.....	43

3.1	Methodology	43
3.2	Research Design.....	44
3.3	Participants.....	48
3.4	Instruments.....	49
3.5	Conclusion	51
4.	Data Collection and Analysis.....	52
4.1	Introducing the Transport Data Science Initiatives.....	52
4.2	Summary and Implications	62
5.	Unique Characteristics of DSIs	63
5.1	Abstract	63
5.2	Introduction.....	63
5.3	Literature Review.....	65
5.4	Research Setting and Methods	68
5.5	Data Collection and Analysis.....	72
5.6	Conclusion and RecommendationS	90
6.	Framework to Manage DSIs.....	95
6.1	Abstract	95
6.2	Introduction.....	95
6.3	Literature Review.....	97
6.4	Research Setting and Methods	104
6.5	Data Collection and Analysis.....	107
6.6	DSI Delivery Framework.....	107
6.7	Conclusion and Recommendation.....	117
7.	Minimum Viable Governance for DSIs.....	119
7.1	Abstract	119
7.2	Introduction.....	119
7.3	Literature Review.....	121
7.4	Research Setting and Methods	131

7.5	Data Collection and Analysis	134
7.6	Conclusion and RecommendationS	144
8.	Insights into Building Better Business Cases for DSIs.....	150
8.1	Abstract.....	150
8.2	Introduction	151
8.3	Literature Review	152
8.4	Research Setting and Methods.....	158
8.5	Data Collection and Analysis	159
8.6	Conclusion and RecommendationS	169
9.	Using DSIs for Benefits Tracking and Investment Decisioning – A Transport for NSW Case Study	174
9.1	Abstract.....	174
9.2	Introduction	175
9.3	Literature Review	176
9.4	Research Setting and Methods.....	180
9.5	Data Collection and Analysis	181
9.6	Conclusion and Recommendation	187
10.	Discussion and Conclusion	190
10.1	Introduction	190
10.2	Contributions to Theory.....	193
10.3	Contributions to Practice	206
10.4	Other Contributions	210
10.5	Research Limitations	212
10.6	Future Research	213
10.7	Conclusion.....	214
11.	Appendices	218
11.1	Appendix A: Analysis of Articles on Big Data	218
11.2	Appendix B: Semi-Structured Interview Guide.....	222

11.3	Appendix C: Navigating Hyperlinks.....	226
12.	Glossary	228
13.	References	230

List of Figures

<i>Figure 1-1. Program Life Cycle (Project Management Institute, 2017b, p. 5).</i>	8
<i>Figure 2-1. Project Classification using Diamond Framework (Shenhar et al., 2004)</i>	15
<i>Figure 2-2. Organisational Context of Portfolio Management (Project Management Institute, 2017a, p. 8)</i>	17
<i>Figure 2-3. Portfolio Management Framework (Jenner & Kilford, 2011, p. x)</i>	18
<i>Figure 2-4. Lean Startup Cycle (Scaled Agile, 2023)</i>	19
<i>Figure 2-5. Epics vs Projects (Scaled Agile, 2023)</i>	20
<i>Figure 2-6. Lean Business Case (Scaled Agile, 2023)</i>	21
<i>Figure 2-7. Governance Hierarchy (Project Management Institute, 2017b, p. 68)</i>	24
<i>Figure 2-8. MSP framework and concepts (Axelos, 2022)</i>	25
<i>Figure 2-9. Comparison of Traditional and Agile Project Governance Practices (Lappi et al., 2018, p. 55)</i>	27
<i>Figure 2-10. Traditional vs lean-agile Portfolio mindset (Scaled Agile, 2023)</i>	28
<i>Figure 2-11. SAE Lean-Agile Principles (Scaled Agile, 2023)</i>	29
<i>Figure 2-12. SAE Lean Portfolio Management (Scaled Agile, 2023)</i>	30
<i>Figure 2-13. SAE Lean Budget Guardrails (Scaled Agile, 2023)</i>	31
<i>Figure 2-14. SAE Lean Portfolio Management Events (Scaled Agile, 2023)</i>	31
<i>Figure 2-15. Weighted Shortest Job First formula (Scaled Agile, 2023)</i>	32
<i>Figure 2-16. Lean Portfolio Metrics (Scaled Agile, 2023)</i>	33
<i>Figure 2-17. Data Management Framework (Earley, 2017, p. 36)</i>	36
<i>Figure 2-19. Difference between frameworks for scaling agile (Atlassian, 2020)</i>	38

<i>Figure 2-20. CRISP-DM Data Science Processes (Chapman et al., 2000, p. 13)</i>	<i>40</i>
<i>Figure 2-21. Comparison of data science methodologies (Foroughi & Luksch, 2018, p. 10)</i>	<i>40</i>
<i>Figure 3-1. Research Methodology Approach.....</i>	<i>46</i>
<i>Figure 4-1. Transport DSIs Chronology</i>	<i>55</i>
<i>Figure 4-2. DSI Delivery Complexity</i>	<i>62</i>
<i>Figure 5-1. Vanguard Delivery Timeline in the Business Case (Oct 2016)</i>	<i>74</i>
<i>Figure 5-2. Vanguard Delivery Timeline using a data-centric approach (Jan 2018)</i>	<i>75</i>
<i>Figure 5-3. MPR Delivery Roadmap.....</i>	<i>77</i>
<i>Figure 5-4. Outcomes Delivered by Vanguard.....</i>	<i>80</i>
<i>Figure 5-5. Benefits Mapping and Dependencies of the six DSIs</i>	<i>82</i>
<i>Figure 6-1. Understanding the Knowledge Gap for DSIs</i>	<i>98</i>
<i>Figure 6-2. Comparison of KDD, SEMMA and CRISP-DM (Azevedo & Santos, 2008, p. 5).....</i>	<i>103</i>
<i>Figure 6-3. DSI Delivery Framework.....</i>	<i>109</i>
<i>Figure 7-1. Active Transport investment horizon budget guardrail</i>	<i>137</i>
<i>Figure 7-2. Active Transport Data and Analytics Roadmap</i>	<i>138</i>
<i>Figure 7-3. Average Walking and Cycling Counts in NSW.....</i>	<i>141</i>
<i>Figure 7-4. DSI Governance using Four-Governance Paradigms (Müller, 2016)...</i>	<i>147</i>
<i>Figure 8-1. Lean Business Case (Scaled Agile, 2023)</i>	<i>157</i>
<i>Figure 8-2. MPR Program Outcomes, Publications and Achievements.....</i>	<i>165</i>
<i>Figure 8-3. Active Transport Value Streams</i>	<i>170</i>
<i>Figure 9-1. Camera with AI/ML to capture cycling and walking counts</i>	<i>182</i>
<i>Figure 9-2. ML model inputs and outputs for cycling trips and kilometres (Saber, 2021).....</i>	<i>183</i>
<i>Figure 9-3. Average weekday cycling counts in Sydney metropolitan region</i>	<i>183</i>

<i>Figure 9-4. ML model inputs and outputs for walking trips and kilometres</i>	<i>184</i>
<i>Figure 9-5. Active Transport Voice of Customer</i>	<i>185</i>
<i>Figure 9-6. Cycleway Finder</i>	<i>186</i>
<i>Figure 9-7. Mobility Heatmap of Sydney</i>	<i>187</i>
<i>Figure 10-1. DSI Delivery Framework</i>	<i>197</i>
<i>Figure 10-2. DSI Governance using the Four-Paradigm Governance Model</i>	<i>204</i>
<i>Figure 10-3. Management and Governance of DSIs</i>	<i>217</i>

List of Tables

Table 1-1 <i>Key Reasons for taking Program-based approach for DSIs</i>	6
Table 1-2 <i>Program Life Cycle Deliverables and DSI Gaps and Issues</i>	8
Table 4-1 <i>Summary of Transport DSIs</i>	52
Table 4-2 <i>Benefits Summary of Transport DSIs</i>	56
Table 4-3 <i>Stakeholder Summary of Transport DSIs</i>	57
Table 4-4 <i>Source Data Summary of Transport DSIs</i>	58
Table 4-5 <i>Technology Summary of Transport DSIs</i>	59
Table 4-6 <i>Risks, Dependencies and Constraints of Transport DSIs</i>	59
Table 4-7 <i>Roles and Responsibilities across DSI Domains</i>	60
Table 4-8 <i>Skills Summary of Transport DSIs</i>	61
Table 4-9 <i>DSIs Delivery Complexity</i>	61
Table 5-1 <i>Program Management Artefacts and Ceremonies</i>	70
Table 5-2 <i>Vanguard – Planned vs Actual Delivery of datasets</i>	75
Table 5-3 <i>Vanguard Business Case Quantified Benefits</i>	79
Table 5-4 <i>Program Life Cycle Deliverables and DSI Gaps and Issues</i>	83
Table 5-5 <i>About Semi-Structured Interview Participants</i>	84
Table 5-6 <i>Interview Questions</i>	85
Table 5-7 <i>Unique Characteristics of DSIs</i>	92
Table 5-8 <i>Key differences between DSIs and ICT-enabled Programs</i>	93
Table 6-1 <i>Comparison of Change Management Models</i>	100
Table 6-2 <i>DSI Delivery Framework – Program Management Phases</i>	110
Table 6-3 <i>DSI Delivery Framework – Change Management Steps</i>	110
Table 6-4 <i>DSI Delivery Framework – Scaled Agile Processes</i>	111
Table 6-5 <i>DSI Delivery Framework – Data Management Processes</i>	114

Table 6-6 <i>DSI Delivery Framework – Data Science Processes</i>	115
Table 6-7 <i>DSI Delivery Framework – Summary of key processes across five domains and program phases</i>	116
Table 7-1 <i>Comparison of Traditional and Agile Project Governance Practices (Lappi et al., 2018, p. 55)</i>	126
Table 7-2 <i>Guiding Principles for Agile Governance</i>	127
Table 7-3 <i>Lean Portfolio Metrics (Scaled Agile, 2023)</i>	130
Table 7-4 <i>Active Transport Metrics</i>	141
Table 7-5 <i>Guiding Principles for Minimum Viable Governance of DSIs</i>	148
Table 8-1 <i>Vanguard Business Case Quantified Benefits</i>	161
Table 8-2 <i>Guiding Principles for Building Better Business Cases for DSIs</i>	172
Table 9-1 <i>Framework for per-km Active Transport benefits</i>	176
Table 9-2 <i>Financial impacts for per-km Active Transport benefits for urban and rural areas</i>	178
Table 10-1 <i>Summary of Research Contributions</i>	190
Table 10-2 <i>Unique Characteristics of DSIs</i>	195
Table 10-3 <i>DSI Delivery Framework – Summary of Key Processes across five domains and program phases</i>	198
Table 10-4 <i>Roles and Responsibilities across DSI Domains</i>	200
Table 10-5 <i>Guiding Principles for Minimum Viable Governance for DSIs</i>	202
Table 10-6 <i>Guiding Principles for Building Better Business Cases for DSIs</i>	206

List of Abbreviations

ADKAR	Awareness, Desire, Knowledge, Ability and Reinforcement
AI	Artificial Intelligence
ART	Agile Release Train
CRISP-DM	Cross-Industry Standard Process for Data Mining
DevOps	Development and Operations
DevSecOps	Development, Security and Operations
DSI	Data Science Initiative
ICT	Information and Telecommunications Technology
ML	Machine Learning
MVP	Minimum Viable Product
PMI	Project Management Institute
UTS	University of Technology Sydney
WSJF	Weighted Shortest Job First

List of Publications and Courses

Main Publications

As part of this research, seven papers were written and presented at conferences or published as a chapter in handbook. The content of these papers is included in the thesis. It is acknowledged that some content is intentionally repeated across chapters to maintain integrity and comprehensiveness of the research. This repetition serves the purpose of reinforcing key findings and conclusions and providing a holistic understanding of the research.

Mathur, S. (2019). *Unique Characteristics of Data-Analytics Programs*. Paper presented at PMI Australia Conference, Gold Coast. (Mathur, 2019)

Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2024). *Unique characteristics of Data Science Initiatives*. Paper presented at the IPMA 28th International Congress on Project Management, Arandjelovac, Serbia. (Mathur, Sankaran, MacAulay, & Tsang, 2024)

Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2021). *A Framework to Manage Data Science Initiatives*. Paper presented at The 6th PMI Research and Academic Virtual Conference 2021. (Mathur, Sankaran, MacAulay, & Tsang, 2021)

Mathur, S. (2021). *A Framework to Manage Data Science Initiatives*. Paper presented at 21st International Conference on Knowledge, Culture and Change in Organizations, Auckland, New Zealand. (Mathur, 2021a)

Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2022). *Minimum Viable Governance for Data Science Initiatives – A Transport for NSW Case Study*. Paper presented at IPMA 10th Research Conference, Belgrade, Serbia. (Mathur, Sankaran, MacAulay, & Tsang, 2022)

Mathur, S. (2021). *Using Data Science Initiatives to deliver Smart Infrastructure and improve Customer Experience – A Transport for NSW Case Study*. Paper presented at the PGCS – Project and Program Management Symposium, Canberra 2021. (Mathur, 2021b)

Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2023). Minimum Viable Governance for Data Science Initiatives – A Transport for NSW Case Study. In *Research Handbook for the Governance of Projects*. (Mathur, Sankaran, MacAulay, & Tsang, 2023)

Working Papers

The paper listed below is being finalised and will be submitted for publication later.

Mathur, S. (2022). *When Should Data-Science Programs be Managed as Exploratory Projects? Insights for Building Better Business Cases.* Working Paper No. 1 (Mathur, 2024)

Courses

Courses that draw upon content from this thesis are listed below. The first course, which is delivered as a microcredential, has been run since 2021. The second course, while run since 2017, had DSI content included from 2020.

Mathur, S., & Sankaran, S. (2021). Management and Governance of Data Science Initiatives (UTS Course). Retrieved from <https://open.uts.edu.au/uts-open/study-area/built-environment/management-and-governance-of-data-science-initiatives2/> (Mathur & Sankaran, 2021)

Sankaran, S., Lawler, C., Ang, K., & Mathur, S. (2021). Governance, Portfolio and Program Management (Master's Course). Retrieved from <https://handbook.uts.edu.au/subjects/15346.html> (Sankaran, Lawler, Ang, & Mathur, 2021)

Related Publications and Papers

Other publications that draw upon the published contributions of the above-mentioned papers but do not form part of this thesis are as follows:

Mathur, S. (2005). *How PMBOK-RUP-ITIL integration contributes to successful software development projects.* Paper presented at PMI® Global Congress 2005—Asia Pacific, Singapore. Retrieved from <https://www.pmi.org/learning/library/pmbok-rup-itol-software-development-7573> (Mathur, 2005).

Mathur, S. (2006). *Project portfolio management techniques.* Paper presented at the PMI® Global Congress, Hong Kong, People's Republic of China. (Mathur, 2007).

Mathur, S., Sankaran, S., & MacAulay, S. (2020). *Project Delivery for General-Purpose Technologies: New Insights from Data Science Initiatives*. Working Paper No. 3 (Mathur, Sankaran, & MacAulay, 2023)

Cilliers, E. J., Sankaran, S., Armstrong, G., Mathur, S., & Nugapitiya, M. (2021). From Urban-Scape to Human-Scape: COVID-19 Trends That will Shape Future City Centres. *Land, 10(10), 1038*. (Cilliers, Sankaran, Armstrong, Mathur, & Nugapitiya, 2021)

Mathur, S., Ninan, J., Vuorinen, L., Ke, Y., & Sankaran, S. (2021). An exploratory study of the use of social media to assess benefits realization in transport infrastructure projects. *Project Leadership and Society, 2*, 100010. (Mathur, Ninan, Vuorinen, Ke, & Sankaran, 2021)

Devkar, G., Sankaran, S., Ke, Y., Ninan, J., Mathur, S., Tsang, I., & Vuorinen, L. (2021). Data Analytics to Evaluate Public Value from Megaprojects. *The 6th PMI Research and Academic Virtual Conference 2021*. (Devkar et al., 2021)

Lee, B., Rosario, L. D., Wu, H., Roberts, L., Arnold, T., Mathur, S., & Pettit, C. (2023). Assessing the monetary value of active transport and e-micromobility per unit of distance travelled: a systematic review. *Manuscript submitted for publication*. (B. Lee et al., 2023)

Ninan, J., Ke, Y., Sankaran, S., Mathur, S., Vuorinen, L., Devkar, G. (2022). Social media for improving metro rail project operations. In J. Ninan (2022). *Social media for project management*, CRC Press, Taylor and Francis. (Ninan et al., 2022)

Abstract

Data is becoming increasingly ubiquitous in organisational life, with many business investments using data to inform the development of their business case, to deliver core elements of scope, and to track the realisation of benefits. Data science initiatives (DSIs) are defined as related projects or programs that involve the application of data science techniques and methods to address complex problems, generate insights or create new products or services. DSIs typically involve the collection, analysis and interpretation of large and complex datasets and require expertise in data science, statistics, machine learning and programming. While DSIs have emerged as a popular mechanism for extracting value from data, their track record has drawn substantial criticism from sponsors. The success rate of delivering DSIs is perceived to be low, with Gartner estimating that 85% of projects fail (Asay, 2017).

The complexity of integrating various data sources, technologies and expertise; interdisciplinary nature requiring collaboration across multiple teams; strategic impact driving organisational transformation; long-term focus involving ongoing iterations and adaptations; and ongoing risk management of technical, regulatory and ethical considerations necessitated taking a program-based approach instead of project-based for their delivery. This highlights the need for effective program management frameworks to guide the successful delivery of DSIs and to realise the value they can potentially bring to an organisation.

Program managers require a methodology and framework to guide them in the successful delivery of programs. A review of literature shows a lack of frameworks that program managers can adopt to manage DSIs and navigate the exploratory and innovative nature of such investments. The existing literature is inadequate in addressing the end-to-end program life cycle of DSIs, including the construction and funding of business cases, setting up appropriate governance, effectively delivering DSIs and then realising the value of DSIs through benefits tracking and investment decision-making. This highlights the need for research that addresses these gaps and provides program managers with the tools and guidance they need to effectively manage and deliver DSIs.

This research posits that traditional practices used by program managers for conceptualising and managing ICT-enabled programs are not adequate for DSIs, which have unique characteristics. To build this argument, the research employed case studies of seven DSIs spanning six years at Transport for NSW (Transport), the statutory body responsible for transportation in the state of New South Wales in Australia. By analysing the seven case studies, the research aimed to understand the challenges and best practices in managing and delivering DSIs and to provide a framework for program managers to effectively navigate the unique characteristics of these initiatives. This research provides a valuable contribution to the field by addressing the gap in understanding how to effectively manage and deliver DSIs and by providing a practical framework for program managers to apply in their work.

The research delved into the difference between exploration and exploitation and how managing DSIs as “exploratory projects” could improve the success rate of implementations. It offers an evidence-based delivery framework using agile methods, which allows for flexibility and adaptability in the face of uncertainty and changing requirements. To address the governance of DSIs, the research proposes a minimum viable governance framework using lean portfolio management and product governance, moving away from traditional portfolio, program and project governance methods which may not be as effective in managing DSIs. Additionally, the research provides an example of how Transport is using DSIs for benefits tracking and investment decision-making and thereby demonstrates the potential value that DSIs can bring to an organisation. Overall, the research provides a comprehensive understanding of the unique characteristics of DSIs and the best practices for managing and delivering them successfully.

DSIs inherently involve a blend of exploration – experimentation, innovation, and discovery and exploitation – execution, efficiency, and optimisation of data-driven solutions. This duality is central to their success, as they navigate the uncertain terrain of emerging technologies and rapidly evolving business landscapes. By aligning with the theoretical framework of explorative and exploitative initiatives, this research elucidates how managing DSIs as “exploratory projects” within a program management context can enhance innovation outcomes while balancing the need for operational efficiency and value delivery. Furthermore, the research informs the

literature on exploitative and exploratory initiatives while contributing to practical outcomes such as efficiency and value delivery for DSIs.

In conclusion, this research provides valuable insights for both theory and practice. It contributes theoretically by integrating concepts of explorative and exploitative initiatives to confront innovation within the management of DSIs. Practically, it offers actionable insights for portfolio, program and change management practitioners, as well as decision-makers on how to effectively navigate the life cycle of DSIs and realise their value through benefits tracking and investment decision-making. Additionally, this research serves as a valuable resource for academics in portfolio, program and change management, who can use its findings to enhance their teaching and conduct further research to validate the challenges of using data and DSIs in other public and private sector organisations, both in Australia and internationally. Overall, this research addresses an identified need for academics and practitioners to better understand the emerging fields of big data and data science through a program management lens.

1. Introduction

This research focused on the analysis of DSI case studies at Transport for NSW (Transport) to understand the challenges faced and the methods used to overcome them. The aim of this research was to create a body of work that can assist data science practitioners in navigating the life cycle of DSIs, improving the success rate of such deliveries and realising the value of DSIs through benefits tracking and investment decision-making. Additionally, the research aimed to provide guidance for decision-makers on how to construct and manage business cases for DSIs, as well as being a useful resource for academics teaching program management and data science courses. The research findings can be used to inform best practices in managing and delivering DSIs and to provide a framework for practitioners and decision-makers to apply in their work.

This chapter presents an overview of the research, including its background and context, purpose, research questions, significance, scope and definition of terms used. In Section 1.1, the background of the research is discussed, including the current state of the field and the challenges faced in the delivery of DSIs. Section 1.2 provides the context of the research, including the specific organisation and case studies that were undertaken to understand the challenges faced in the delivery of DSIs. In Section 1.3, the purpose of the research is outlined, including the specific goals and objectives of the study. Section 1.4 presents the research questions, which guide the study and provided a clear focus for the research. Section 1.5 explains the significance and scope of the research, including the potential impact of the study on practitioners, decision-makers and academics, as well as the specific areas of the field that the research addresses. Finally, Section 1.6 provides an overview of the remaining chapters of the thesis, outlining the structure and content of the entire study.

1.1 BACKGROUND

While awareness of how to harness data continues to increase among organisations, data science and DSIs can still in the year 2023 be considered as an emerging discipline. DSIs offer some unique challenges in their delivery. While the high-level target outcomes are often known, the how-to is largely unknown because the data attributes, data quality and business rules are not known until a data discovery exercise is undertaken using sample data. Even then, not all data characteristics may emerge until complete data is received and examined. In addition, the time and cost estimates for development and delivery are also uncertain. These characteristics can often make it difficult to get approval for an investment business case and, even if this does happen, benefits realisation and tracking are challenged, making governance difficult. There is also reliance on internal and external operational systems to produce data and these systems undergo continuous change. Thus, the characteristics of the data also evolve over time. For example, if we are trying to identify crime hot spots in the Transport network, we initially need a few weeks of recorded data from Operational Systems to examine the structure and patterns. Once the initial data discovery is completed and the basic rules formulated, we can validate the assumptions and rules over larger datasets ranging from a few months to years. How this data is stored on a big data platform is also important because we want the business rules and visualisations to keep updating over time as systems evolve.

1.2 CONTEXT

The challenges described above underline both the innovative and exploratory nature of DSIs. There is a growing number of studies at the intersection of innovation and project studies that highlight similar challenges whereby organisations attempt a standardised and routinised approach to manage uncertainty instead of applying an adaptive model and treating innovative projects as “voyages of discovery” (Davies, Manning, & Söderlund, 2018, p. 969). Drawing on the research of prominent management scholar James March (March, 1991), one way to think about this is as an exploration of new possibilities and the exploitation of old certainties in the context of organisational learning. March defined exploration using terms such as search, variation, risk taking, experimentation, play, flexibility, discovery and innovation, and he defined exploitation as refinement, choice, production, efficiency, selection,

implementation and execution. Lavie, Stettner, and Tushman (2010) describe the distinction between exploration and exploitation as a matter of degree rather than kind and view exploration-exploitation as a continuum rather than a choice between discrete options.

The exploration-exploitation distinction has also shown promise within research on projects. In his study of initiatives relating to telematics services for automobiles, Lenfle (2008) identified a set of challenges where neither technology nor customer requirements were known at the start of the project. In the context of innovation, he questioned the definition of project management currently described in most textbooks and discourses on project management as the accomplishment of a clearly defined goal in a specified period, within budget and quality requirements. He argued that traditional and sometimes heavyweight project management models requiring detailed upfront planning and schedules were not suitable for innovation projects that were characterised by divergence and unforeseeable uncertainties. Rather, the process of innovation is messy and complex and cannot be reduced to a simple sequence of stages or phases, as implied by the stage gate model that is used in many organisations for managing innovation (Van de Ven, 2017). In fact, Lenfle called them exploratory projects and outlined five management principles to deliver them:

1. *setting up a dedicated organisation to manage the exploration,*
2. *the central role of tests (prototypes, testing, customer trials, etc.) in the management process,*
3. *the need for concurrent exploration,*
4. *satisfying two different dimensions of performance: the value of the products and accumulated knowledge, and*
5. *management tools that allow a reformulation of the objectives along the way.*

Another body of work based on research into a variety of innovation projects (Shenhar, 2001; Shenhar, Dvir, Morris, & Pinto, 2004) also supports a view that one project management approach does not fit all projects. Shenhar et al. (2004) developed a Diamond framework classifying projects based on novelty, technology, complexity and pace – proposing different ways to manage projects with these attributes. For

example, a project that is “derivative”, “medium-tech”, “system” and “fast/competitive” will need to be managed very differently than a project that is “breakthrough”, “high-tech”, “system” and “blitz” with the latter type of project requiring more sophisticated ways to manage the risks, resources, project team and the interactions with company executives.

Transport relies on data and technology to deliver services to customers and continues to invest in new datasets and innovate. The DSI case studies at Transport demonstrate reliance on emerging technologies such as artificial intelligence and big data. Matheus, Janssen, and Maheshwari (2020) showcase related use of data science and dashboards by governments to support their decision-making and policy processes as well as to communicate with the public in smart cities context. A review of the application of big data and data analytics shows very little research has been done on how to deliver DSIs. Appendix A: Analysis of Articles on Big Data in Section 11.1 show a gap exists in research on effective delivery of DSIs. Though big data has now become commonplace as a business term, there is very little published management research that tackles the challenges of using tools to deal with big data; or better yet, explores the promise and opportunities for new theories and practices that big data might bring about (George, Haas, & Pentland, 2014). A review of critical success factors of big data projects by Saltz and Shamsurhin (2016) shows lack of an agreed standard to execute them and a recommendation to improve process methodology and prioritises 33 success factors. Viaene and Van den Bunder (2011) have attempted to define skills that differentiate successful business analytics project managers to traditional project managers. They recommend a more adaptive approach with experimentation-based approaches emphasising less specific early planning, good-enough requirements and experimental and evolutionary design with significant ongoing learning and change is better suited to analytics projects. Franková, Drahosová, and Balco (2016) through a survey of 237 practitioners, found that agile and iterative delivery is more appropriate to big data projects. In summary, while DSI’s are becoming important there are still questions about how they can be delivered successfully.

My initial studies revealed that DSIs have characteristics that make using ICT-inspired program management approaches problematic. Table 1-2 *Program Life Cycle Deliverables and DSI Gaps and Issues* highlights these characteristics against each of

the program phases. I build on this work to argue there are occasions when DSIs should be managed as exploratory projects. Furthermore, the waterfall approaches to program management adopted by peak bodies such as the Project Management Institute (PMI), Axelos and the International Project Management Association (IPMA), set up structural tensions between business case development, program design, delivery and benefits realisations that decouple value creation from capture and thus undermine coherent governance across the investment life cycle of DSIs. I argue that the business case for DSIs should be constructed differently: instead of traditional focus on costs and benefits, the approvers should understand the rationale for such exploratory investment and how the benefits profile should be tracked across the investment life cycle. I further argue that as investment in big data and AI, including deep learning, ML, neural network, computer vision and natural language processing increases, the ability to deliver DSIs consistently and successfully needs to mature and I believe that this research will contribute to this development.

1.3 PURPOSE

This research expands on existing frameworks in the program management literature and practice and proposes a framework to deliver DSIs as exploratory and innovation projects. My experience of delivering DSIs over the past six years suggested to me that it was crucial to depart from traditional ICT delivery frameworks by integrating the program management, change management, scaled agile, data management and data science domains. My research aims to systematically analyse this unfolding experience to improve what is known about successfully managing the uncertainty and complexity associated with delivering DSIs.

1.4 RESEARCH QUESTIONS

To meet the challenge described in the section above, the thesis focuses on investigating the following key research questions:

***RQ:** Why do Data Science Initiatives face challenges delivering envisaged value when using traditional processes for managing ICT-enabled programs? How do these challenges manifest within the delivery process?*

The three supplementary research questions are:

SQ1: *What design principles should be incorporated in a Data Science Initiative delivery framework so that program managers can adopt a predictive path to realise value from such investments?*

SQ2: *What is the minimal viable governance required for Data Science Initiatives to ensure they deliver envisaged value efficiently and effectively?*

SQ3: *How should the business cases for Data Science Initiatives be structured?*

The research delivered the following outcomes:

- Identification of unique characteristics due to which DSIs face challenges delivering envisaged value when using traditional processes for managing ICT-enabled programs (Chapter 5 – Unique Characteristics of DSIs),
- Creation of a body of knowledge including a framework for delivery of DSIs acknowledging their exploratory and innovative nature and thus improving the likelihood of delivering their envisaged value (Chapter 6 – Framework to Manage DSIs),
- Proposing minimum viable governance for DSIs (Chapter 7 – Minimum Viable Governance for DSIs),
- Insights for writing business cases for DSIs (Chapter 8 – Insights into Building Better Business Cases for DSIs),
- Use case for DSIs to be used for Benefits Tracking and Investment Decisioning (Chapter 9 – Using DSIs for Benefits Tracking and Investment Decisioning – A Transport for NSW Case Study), and
- A microcredential course “Management and Governance of Data Science Initiatives” at the University of Technology Sydney targeted at practitioners and academics (Mathur & Sankaran, 2021).

1.5 SIGNIFICANCE, SCOPE AND DEFINITIONS

The inherent complexity, scope, and strategic importance within organisations justified the use taking a program-based approach instead of project-based for their delivery.

Table 1-1
Key Reasons for taking Program-based approach for DSIs

Key Reason	Comments
Complexity	DSIs involve the application of advanced data science techniques to tackle complex problems, often requiring the integration of various data sources, technologies, and expertise. This complexity surpasses that of typical projects and aligns more closely with the characteristics of programs.
Interdisciplinary Nature	DSIs typically require collaboration across multiple departments or teams, involving specialists in data science, statistics, machine learning, programming, and domain-specific knowledge. Managing such interdisciplinary efforts is better suited to the program management approach, which emphasises coordination and alignment of diverse stakeholders and resources over an extended period.
Strategic Impact	DSIs are often strategic initiatives aimed at driving organisational transformation, innovation, or competitive advantage through data-driven insights and decision-making. They typically span multiple projects and initiatives, each contributing to overarching organisational goals. Treating DSIs as programs helps ensure alignment with strategic objectives and facilitates holistic management of their various components.
Long-term Focus	Unlike projects with well-defined objectives and timelines, DSIs often have longer-term horizons and involve ongoing iterations and adaptations in response to evolving data, technology, and business needs. Program management provides a framework for continuous improvement and adaptation over time, ensuring the sustained value delivery of DSIs.
Risk Management	DSIs inherently entail significant risks, including technical, regulatory, and ethical considerations related to data privacy and security. Program management offers a more comprehensive approach to risk identification, assessment, and mitigation, enabling organisations to proactively manage the complexities and uncertainties associated with DSIs.

Due to reasons cited above, the thesis focuses on how existing program management frameworks support DSIs and how organisations can structure their activities to get benefits from DSIs. A review of the program life cycle (PMI, 2017) has identified gaps in using it for the delivery of DSIs. *Figure 1-1* shows a typical program life cycle followed by gaps and issues against deliverables for each of the three phases of a program in Table 1-2.

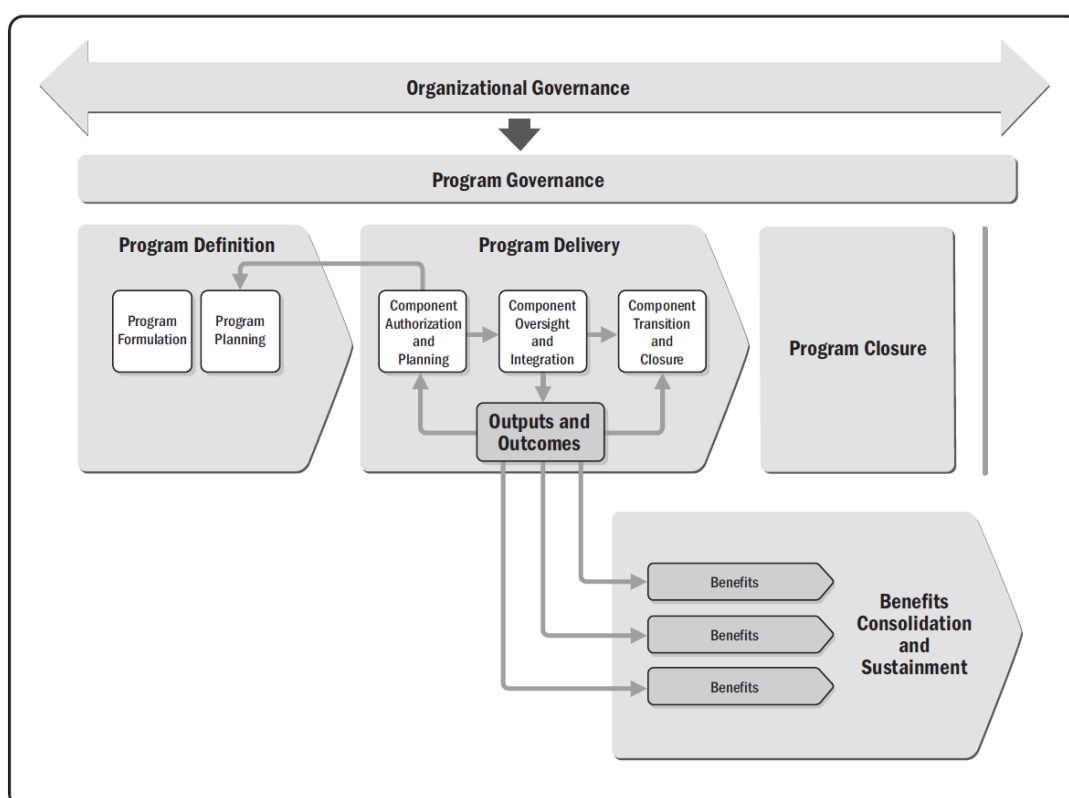


Figure 1-1. Program Life Cycle (Project Management Institute, 2017b, p. 5).

The gaps and issues identified in Table 1-2 are a summary of my observations over seven DSIs starting with ‘Vanguard’ which was the first DSI we delivered and then appeared as a recurring theme over other DSIs I have used as case studies in this research.

Table 1-2
Program Life Cycle Deliverables and DSI Gaps and Issues

Key Phase Deliverables	Gaps and Issues for DSIs
Program Definition Phase	
Key deliverables of this phase are Business Case, Program Charter and Program Management Plan.	For DSIs, risks associated with both costs and benefits are high. The Business Case which can take 6 to 12 months to develop and get approved in both public and private sectors can stay relevant for a traditional ICT investment but may not accurately reflect rapidly evolving. Requirements and technology landscape for DSIs. The creation of business case for innovation and exploration projects introducing new products has challenges and causes a lot of uncertainty and insecurity in the

decision-making process (Chirumalla, 2017).

- Unless the Program Management Plan stays at a high level, the accuracy of scope and schedule is low. The delivery mechanism can only evolve when the Components are identified and executed. This contrasts with traditional ICT initiatives where Components can be identified in advance and scheduled.

Program Delivery Phase

In this phase, individual Components are initiated, planned, executed, transitioned and closed while benefits are delivered, transitioned and sustained in accordance with the Program Management Plan.

- For DSIs, identification of all Components is difficult at the time Program Management Plan is developed and hence only limited planning can be done due to high degree of uncertainty.
- The Benefits will be discovered as the Components are planned and executed again due to high degree of uncertainty. This contrasts with traditional ICT initiatives where increase in revenue and service and/or costs savings can be identified earlier.

Program Closure Phase

In this phase, the Program Benefits are transitioned to sustaining organisation and program is closed.

- While sponsor and stakeholders are continuously communicated and kept informed on both the costs and benefits delivered, for an uninitiated stakeholder the value delivered by the program may be questionable. The outcomes are more enablers to organisational data-driven decision-making capability rather than absolute financial and non-financial metrics as in case of traditional ICT initiative.

Table 1-2 shows that there are many areas to investigate to establish baseline for delivery of DSIs to support the program life cycle. Without such baseline, organisations are at significant risk of wasting precious resources while not delivering outcomes which are possible. Thus, providing a new framework through this research has economic benefits while contributing to improve practice. It also contributes to the growing interest in theory of projects-in-practice (Clegg, Killen, Biesenthal, & Sankaran, 2018) by adding "practice-based" perspective of DSIs to Project portfolio management (PPM) research. Investment in DSIs which are exploratory and come

with uncertain benefits challenge the traditional PPM decision-making practice which focus on rational, top-down and structural aspects of prioritising projects in the portfolio and delivering strategy.

DSIs are defined as related projects or programs that involve the application of data science techniques and methods to address complex problems, generate insights or create new products or services. DSIs typically involve the collection, analysis and interpretation of large and complex datasets and require expertise in data science, statistics, machine learning and programming. DSIs can be found in various domains, such as healthcare, finance, education, transportation and social media and can have different objectives, such as improving operational efficiency, enhancing customer experience or developing innovative solutions.

1.6 THESIS OUTLINE

The thesis has used the “Thesis by Compilation” format (University of Technology Sydney, 2022) which allows a PhD thesis to be structured as a single manuscript that is a combination of chapters and published/publishable works. Therefore, the reader may find that the thesis is repetitive at places. Also, while decisions required consulting with appropriate stakeholders, this research is based on my lived experience delivering the initiatives uses as case studies (Van Maanen, 2015), Therefore, I have used first person throughout the thesis. This approach is further supported by the APA Style Manual which recommends the use of pronoun “I” when writing paper by oneself (American Psychological Association, 2019).

The thesis has this Introduction chapter followed by Research Issues (Chapter 2), Research Design (Chapter 3) and Data Collection and Analysis (Chapter 4). As part of this research, the papers I have written are integrated in the thesis as Unique Characteristics of DSIs (Chapter 5), Framework to Manage DSIs (Chapter 6), Minimum Viable Governance for DSIs (Chapter 7), Insights into Building Better Business Cases for DSIs (Chapter 8) and Using DSIs for Benefits Tracking and Investment Decisioning – A Transport for NSW Case Study (Chapter 9). The chapters are followed by a consolidated Discussion and Conclusion (Chapter 10).

1.7 CONCLUSION

This chapter laid the foundations for the thesis. It introduced the motivation for the research. Then the research was justified, definitions were presented, the research questions defined and the thesis outlined. On these foundations, the thesis can proceed with a detailed description of the research.

2. Research Issues

The focus of this research was to improve practices for the management and governance of DSIs¹ with the target audience being program managers of technology initiatives, portfolio managers/policy makers approving business cases and establishing governance mechanisms and academics teaching program management and conducting research in related topics. Some of the key issues motivating this research were:

- Why do 85% of big-data projects fail (Asay, 2017) when 73% projects (overall) met original goals (Project Management Institute, 2021a)?
- What makes DSIs unique compared to other ICT initiatives?
- How should we construct business cases for DSIs so that expectations are managed?
- What is an appropriate level of governance for DSIs to allow exploration and not stifle innovation?
- How are uncertainty in scope, schedule and benefits managed in ICT projects?
- How can DSIs be delivered effectively? and
- How do DSIs deliver value to organisations?

Significant literature exists covering various domains of ICT initiative delivery. However, a gap exists in the literature in relation to DSIs especially the lack in acknowledgement of the uncertainty, complexity and exploratory nature that characterise the management and governance of DSIs. The nascent stage of this field necessitates reliance on grey literature, such as industry reports and working papers, which provide timely insights not yet available in academic research. I start the review with an overarching discussion on exploratory and exploitative projects (section

¹ Data Science Initiatives (DSIs) are defined as related projects or programs that involve the application of data science techniques and methods to address complex problems and generate insights or create new products or services.

2.1). The rest of the review is then grouped into three themes relating to DSIs – business case (section 2.2), governance (section 2.3) and delivery (section 2.4), all of which leads us to the research questions. While the focus is on effective delivery of DSIs, a review of the technology related to DSIs has been specifically excluded as it is evolving so rapidly and does not affect the above-mentioned themes. Also excluded from review is project management, for reasons cited later in this section. Finally and as mentioned at the beginning of the paragraph, I discuss the knowledge gap that exists in the literature in relation to DSIs which raises questions about their management and governance (section 2.5).

2.1 EXPLORATORY AND EXPLOITATIVE PROJECTS

In this section, I review the differences between exploratory and exploitative projects and their applicability to DSIs. Exploratory projects can be characterised as projects for which neither the goals nor the means of attaining them are clearly defined from the outset (Lenfle, 2008). While Turner and Cochrane (1993) did not use the term “exploratory”, the Type-4 projects in their goals and methods matrix are those where neither the goals nor the methods are well defined. There is a growing number of studies at the intersection of innovation and project studies that highlight the challenges where organisations face when they attempt standardised and routinised approach to manage uncertainty instead of applying an adaptive model and treating innovative projects as “voyages of discovery” (Davies et al., 2018, p. 969). Drawing on management literature, one way to think about this is as the exploration of new possibilities and the exploitation of old certainties in the context of organisational learning (March, 1991). March defined exploration as search, variation, risk taking, experimentation, play, flexibility, discovery and innovation, and exploitation as refinement, choice, production, efficiency, selection, implementation and execution. Lavie et al. (2010) extended the exploration-exploitation framework of March (1991) and described the distinction between exploration and exploitation as matter of degree rather than kind and viewed exploration-exploitation as a continuum rather than a choice between discrete options.

In his study of initiatives relating to telematics services for automobiles where neither technology nor customer requirements were known at the start of the project, Lenfle (2008) identified a set of challenges as follows. In the context of innovation, he questioned the definition of project management as the accomplishment of a clearly defined goal in a specified period,

within budget and quality requirements, particularly as it is described in most textbooks and discourses on project management. He argued instead that traditional and sometimes heavyweight project management models, especially those requiring detailed upfront planning and schedules, were not suitable for such innovation projects that are characterised by divergence and unforeseeable uncertainties. The process of innovation is messy and complex and cannot be reduced to a simple sequence of stages or phases as implied with stage gate model that is used in many organisations for managing innovation (Van de Ven, 2017). Lenfle called innovation projects “exploratory projects” and outlined five management principles to deliver them:

1. *setting up a dedicated organisation to manage the exploration,*
2. *giving tests (prototypes, testing, customer trials, etc.) the central role in the management process,*
3. *the need for concurrent exploration,*
4. *satisfying two different dimensions of performance: the value of the products and accumulated knowledge, and*
5. *management tools that allow a reformulation of the objectives along the way.*

Another body of research into a variety of innovation projects (Shenhar, 2001; Shenhar et al., 2004) also supports a view that one project management approach does not fit all projects. Shenhar et al. (2004) developed a Diamond framework (*Figure 2-1*) that classifies projects based on novelty, technology, complexity and pace and he proposed different ways to manage projects with these attributes. For example, a project that is “derivative”, “medium-tech”, “system” and “fast/competitive” will need to be managed very differently from a project that is “breakthrough”, “high-tech”, “system” and “blitz” with the latter type of project requiring more sophisticated ways to manage the risks, resources, project team and the interactions with company executives.

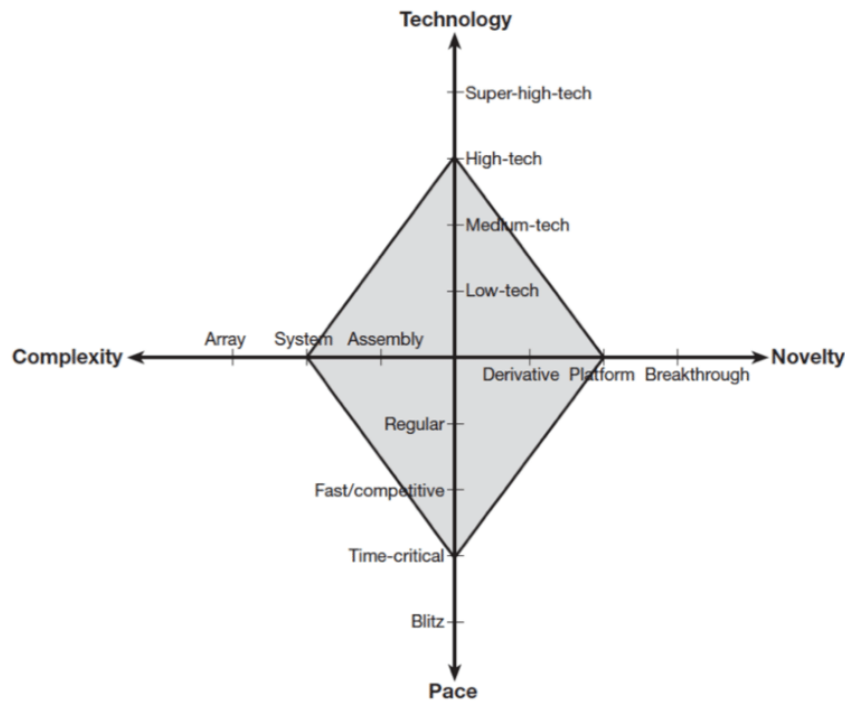


Figure 2-1. Project Classification using Diamond Framework (Shenhar et al., 2004)

Such projects cannot be delivered using standard project management methodology, which largely focuses on delivery of a defined scope, cost and schedule. The inclusion of an Agile Practice Guide in the PMBOK Guide – Sixth Edition (Project Management Institute, 2021b) is in my interpretation an acknowledgment by the PMI that waterfall methods do not allow for effective risk management of ICT-enabled projects, thus causing high failure rates in delivery. As a result of high failure rates, Agile adoption by software development teams increased significantly from 37% in 2020 to 86% in 2021 (digital.ai, 2021). CH Loch, DeMeyer, and Pich (2011) highlight concerns using standard risk management methods for novel projects with unforeseeable uncertainties, complexities and unknown unknowns. They argue that a project plan for novel projects is nothing but a starting point, an illusion or a simple sketch because a conventional project risk management approach cannot be followed. The governance of exploratory and innovative projects is another area requiring consideration, since supervising managers tighten control rather than loosen it when projects become more uncertain (Christoph Loch & Sommer, 2019). They discuss the need for managers to have flexible execution methods for exploratory projects by managers and offer suggestions on how to make exploratory projects more acceptable to top

management attempting to resolve the very fundamental managerial dilemma of wanting to apply more control when uncertainty prevails.

Duncan (1976) introduced the concept of ambidextrous organisation arguing that organisations needed to shift structures to initiate and execute innovation. Ambidextrous organisations possess the ability to simultaneously pursue both incremental and discontinuous innovation by hosting multiple contradictory structures, processes, and cultures (O'Reilly III & Tushman, 2013). This concept has been extensively researched in the context of project management, shedding light on how organisations effectively manage innovation initiatives. For instance, Zhang, Le, Liu, and Chen (2021) explore how ambidextrous innovation strategies can be fostered in large infrastructure projects, emphasising the importance of team heterogeneity. Moreover, in the realm of data science initiatives, researchers have applied the concept of ambidexterity to understand organisational capabilities and strategies for leveraging big data. Studies such as Aljumah, Nuseir, and Alam (2021) and Reddy, Bhattacharjee, Mishra, and Mandal (2022) delve into the relationship between organisational performance, data analytics capabilities, and ambidextrous behaviour, offering valuable insights into the factors enabling successful data science strategies. By integrating the concept of ambidextrous organisations into the discussion, the research gains a deeper understanding of how organisations balance exploration and exploitation in the context of DSIs, enhancing the relevance and comprehensiveness of the analysis.

2.2 THE CASE FOR DATA SCIENCE INITIATIVES

In this section, I review portfolio management practices and the structure of business cases that could support DSIs. The majority of the organisations have either a formal or informal portfolio management function to approve project investments, including in DSIs. Portfolios and portfolio management help deliver organisational strategy (Jenner & Kilford, 2011; Project Management Institute, 2017a). According to the Project Management Institute (2017a), a portfolio is “a collection of projects, programs, subsidiary portfolios and operations managed as a group to achieve strategic objectives”. Similarly, Jenner and Kilford (2011) state “Portfolio management is a coordinated collection of strategic processes and decisions that together enable the most effective balance of the organisational change and business as usual”. Portfolio management is a bridge between organisational strategy and program, projects and operations, as shown in *Figure 2-2* and

ensures that the organisation is doing the “right things” compared to project management which is about doing “things right”.

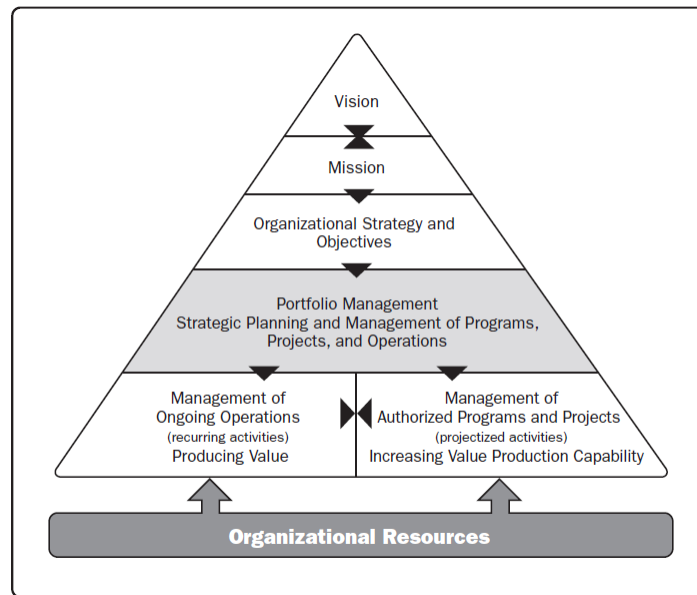


Figure 2-2. Organisational Context of Portfolio Management (Project Management Institute, 2017a, p. 8)

Jenner and Kilford (2011) define cycles of Portfolio Definition and Portfolio Delivery as shown in *Figure 2-3*. The “Understand” practice provides a clear and transparent view of what is in the current portfolio and the project development pipeline. The “Categorise” practice aligns portfolio components against organisation approved categories. The categories include “Run the Business” which is non-discretionary and “Change the Business” which is discretionary and requires a business case for investment approval. While the prioritisation of investments takes place within categories, portfolio balancing is across all portfolios.

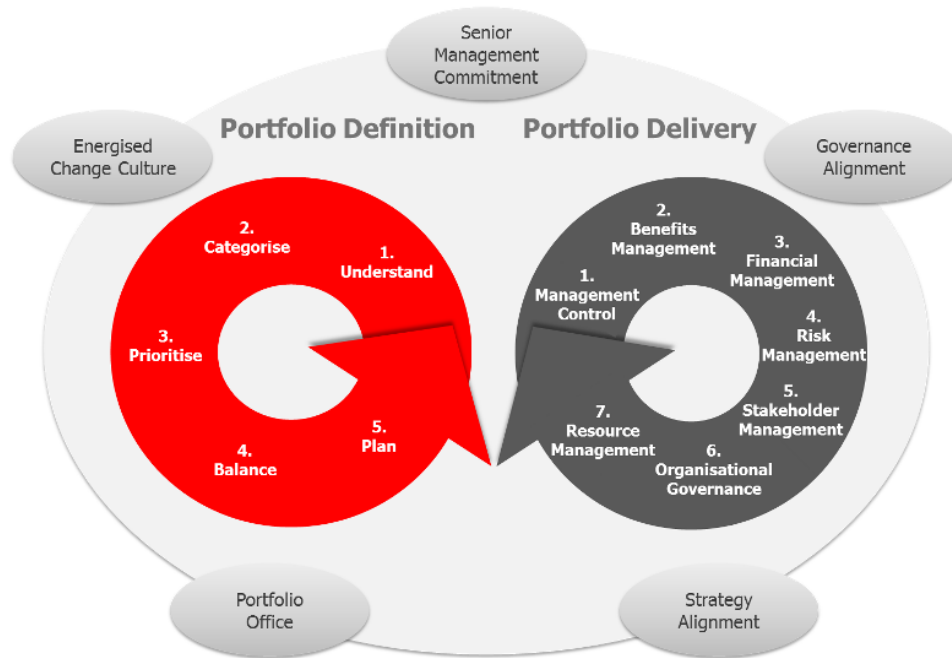


Figure 2-3. Portfolio Management Framework (Jenner & Kilford, 2011, p. x)

This research reviewed portfolio management practices and how they address the exploratory and innovative nature of DSIs in terms of obtaining business case approval and tracking of benefits realisation. Organisations traditionally use business case criteria to make project portfolio decisions that is a business case articulates the business drivers, costs and benefits and the factors affecting their achievability (Axelos, 2022; ISACA, 2010). The benefits should be identified in sufficient detail to allow their contribution to the organisation's strategic measures to be assessed against other potential investments. While financial metrics such as net present value (NPV) and hurdle rates of return are commonly used in decision-making, Christensen, Kaufman, and Shih (2008) describe them as "innovation killers" as they create a systemic bias against successful innovation and focus on short-term shareholder value creation. McGrath and McManus (2020) propose a discovery-driven planning (DDP) approach to integrate agility and innovation into an organisation while minimising disruption to existing business.

Organisations that use lean portfolio management² or that becoming product-centric are increasingly using product lifecycle as an alternative approach to traditional project lifecycle business cases. SAFe (Scaled Agile, 2023) recommends an iterative build-measure-learn cycle for product innovation and strategic investment as shown in *Figure 2-4*.

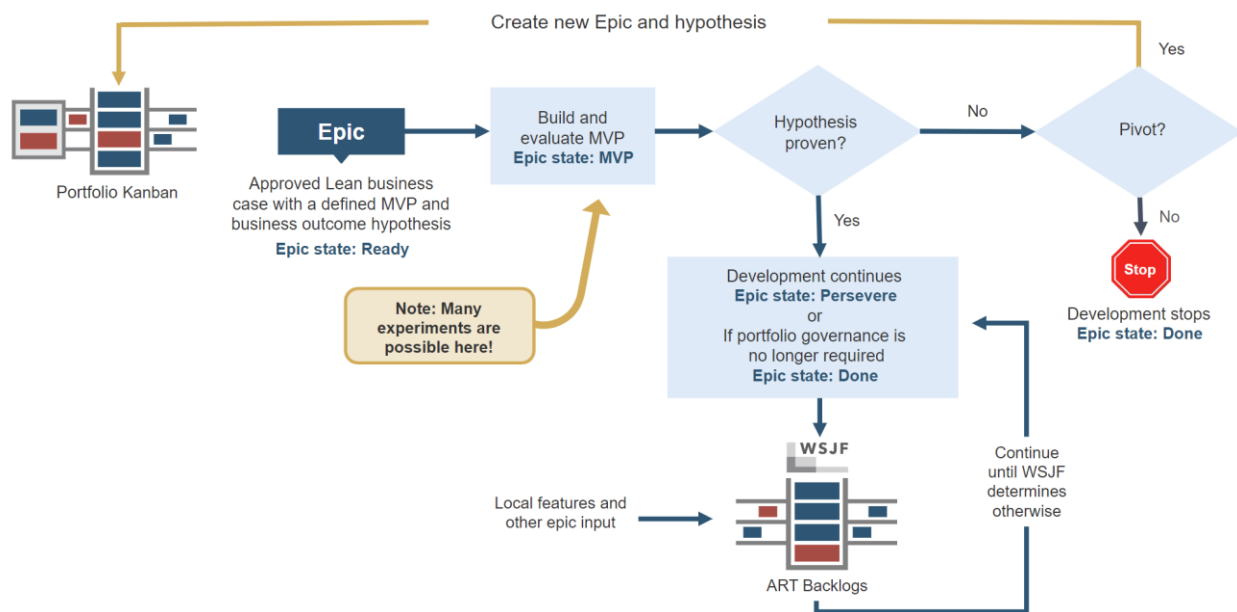


Figure 2-4. Lean Startup Cycle (Scaled Agile, 2023)

SAFe uses Epic instead of projects as a container for a significant solution development and requires a definition of minimum viable product (MVP). An Epic is a large body of work that can be broken down into several smaller stories and generally follows a product roadmap and initiatives hierarchy. The differences between Epic and Projects are outlined in *Figure 2-5*.

² Lean portfolio management is a process by which strategy is aligned with execution using a lean approach and agile portfolio operations and governance. Lean project management is part of an agile methodology that aims to increase customer value by removing project waste.

Epics	≠	Projects
Implemented by stable, cross-functional Value Streams and ARTs.		Implemented by temporary teams, which disband after work is completed.
No definitive start and end date; scope is variable. Continue until WSJF says otherwise.		Definitive start and end date; scope is fixed. All scope must be implemented.
Progress is measured as outcomes against the benefit hypothesis.		Progress is measured based on task completion.
Lean Business Case, based on benefit hypothesis and definition of an MVP.		Overly detailed business case, based on speculative ROI.
Implementation follows the build – measure – learn SAFe Lean Startup Cycle.		Implementation typically follows phase-gated, sequential (waterfall) process.
After the Lean Business Case is approved, commitment is to the evaluation of the MVP.		After business case is approved, up-front commitment is made to the entire project scope.

Figure 2-5. Epics vs Projects (Scaled Agile, 2023)

SAFe proposes creation of a lean business case as illustrated in *Figure 2-6* for review by the lean portfolio management function.

than a project lifecycle. At the same time, program management is required to provide a lateral view of an organisation in classifying, analysing and coordinating the interdependencies among products, projects and other vital elements functioning within the company. Kersten (2018) challenged the traditional project management methods that have served us well for over a century and introduced flow framework, flow metrics and value stream management to both measure the performance of projects and structure the delivery teams. He proposed setting up persistent and fixed teams to support continuous delivery of value instead of project-based teams which get disbanded at the end of the project. Setting up high performance teams takes time and is expensive and letting the product and organisational knowledge walk out the door at the end of a project is inefficient. For this reason, I compared a traditional business case to deliver an ICT project against a value-based business case to set up a permanent team to support a DSI through its entire product lifecycle.

Whether it supports project or product lifecycle, a business case captures the costs and benefits of any investment. Flyvbjerg (2007a) mentions optimism bias and proposes caution in project development as studies in major infrastructure projects in UK and US have shown cost underestimation and benefit overestimation to ensure inverted Darwinism that is "survival of the unfittest". Lovallo and Kahneman (2003) discuss "Delusional optimism: we overemphasise projects' potential benefits and underestimate likely costs, spinning success scenarios while ignoring the possibility of mistakes." Kotter (2007) estimates up to 70% of change initiatives fail to deliver the benefits that they set out to achieve. As part of this research, therefore, I investigated what metrics should be included in DSI business cases so that success upon delivery can be measured.

2.3 THE GOVERNANCE OF DATA SCIENCE INITIATIVES

Considering the significant failure rate, in this section I explore what type of governance is suitable for DSIs to improve their success rate. As governance of DSIs is done within the context of existing governance structures in the organisation, I use the hierarchy of the Corporate, Portfolio, Program and Agile types of governance. Corporate governance deals with the mechanisms by which the stakeholders of a corporation exercise control over corporate insiders and management such that the stakeholders' interests are protected (John & Senbet, 1998). John and Senbet (1998) define the stakeholders of a corporation as equity holders, creditors and other

claimants who supply capital, as well as such stakeholders as employees, consumers, suppliers and the government. Claessens (2006) defines corporate governance as the rules under which firms operate, with the rules coming from such sources as the legal system, financial markets and labour markets as well as corporate social responsibility and sustainability. I expect all DSIs to operate within the boundaries of corporate governance.

Portfolio governance is the set of practices, functions and processes within a portfolio management framework based on fundamental norms, rules, or values that guide portfolio management activities in order to optimise investments and meet organisational strategic and operational goals (Project Management Institute, 2017a). Effective portfolio governance provides clarity about what decisions are made, where and by whom and about what criteria are used in reaching these decisions and they ensure alignment with the wider organisational governance structure (Jenner & Kilford, 2011). I see portfolio governance as a mechanism for selecting the DSIs that render optimum value for money to the organisation rather than as a means to deliver those DSIs “correctly”.

Program governance³ comprises the framework, functions and processes by which a program is monitored, managed and supported in order to meet organisational strategic and operational goals, with a focus on delivery of program benefits by establishing the systems and methods by which a program and its strategy are defined, authorised, monitored and supported by its sponsoring organisation (Project Management Institute, 2017b). A traditional view of governance hierarchy is reflected in *Figure 2-7* where component projects follow program governance and portfolio provides governance policies, oversight, control, integration and decision-making functions and processes to programs within the portfolio. Key program governance roles are program sponsor, program steering committee, program management office, program manager, project manager(s) and other stakeholders.

³ Note: I discuss program governance and not project governance because DSIs are not delivered as a one-off project and generally are part of a larger initiative.

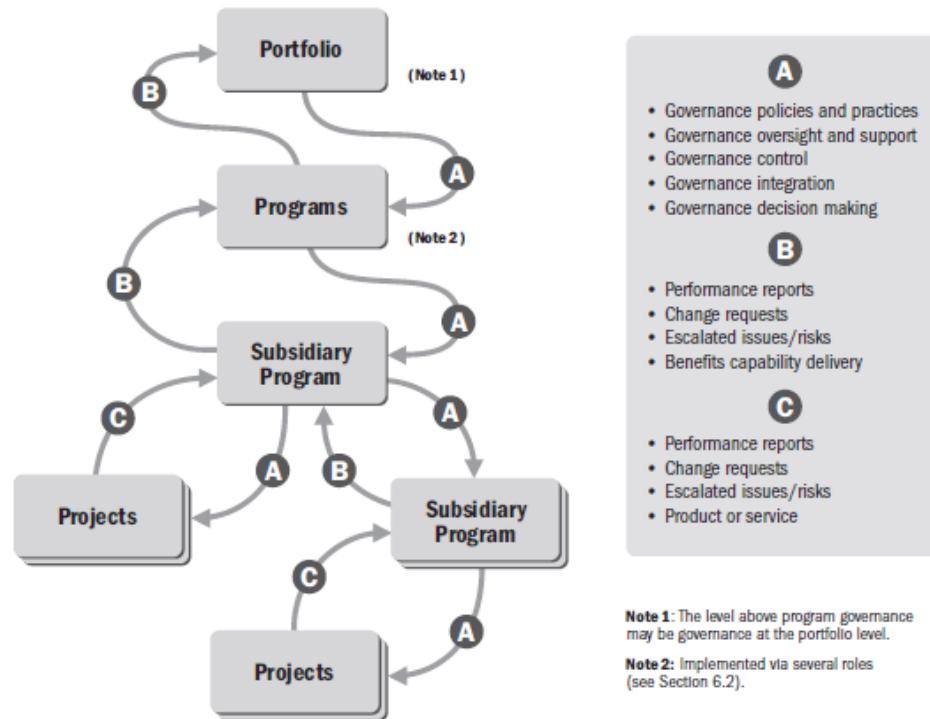


Figure 2-7. Governance Hierarchy (Project Management Institute, 2017b, p. 68)

As an alternative to PMI, Axelos (2022) defines nine governance themes which need to be considered when delivering a program of transformational change. These are program organisation, vision, leadership and stakeholder engagement, benefits management, blueprint design and delivery, planning and control, business case, risk and issue management and quality and assurance management, as shown in *Figure 2-8*.

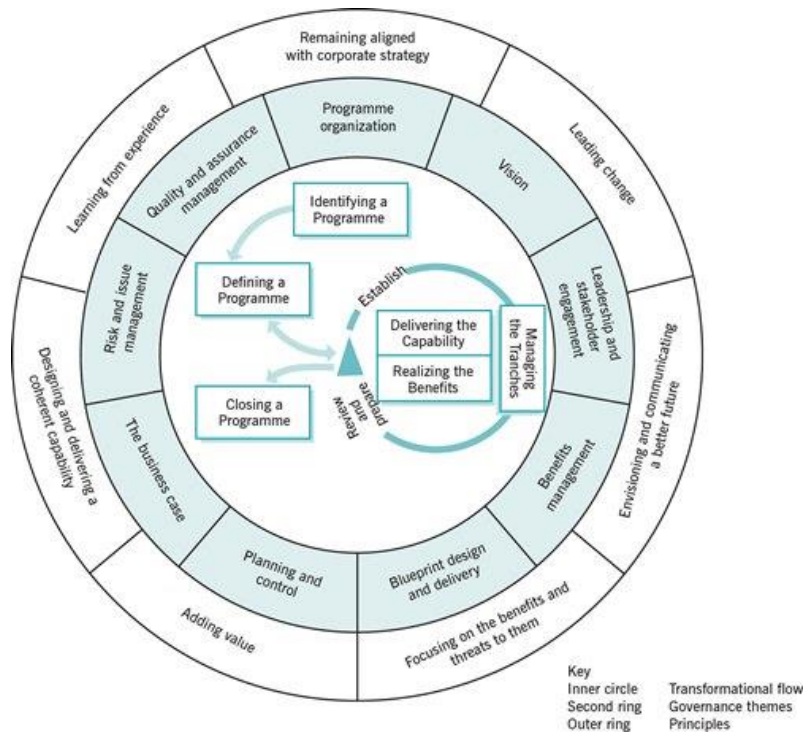


Figure 2-8. MSP framework and concepts (Axelos, 2022)

In the project management literature, Müller (2009, 2016) identified four models of organisational project governance: Process Models, Governance and Governmentality-based Models, Nested Models and Layered Models, all of which cover agency, institutional, stakeholder and transaction cost theories. He then describes the model of four-governance paradigms linking corporate governance with organisational project governance through an overlay of two dimensions. The first is corporate governance orientation from shareholder to stakeholder perspective and second is the control structures in a project's parent organisation from a behaviour-to-outcome perspective. The conformist paradigm assumes efficiency is maximised by trusting process over the capabilities of individuals. The Flexible Economist paradigm aims for a flexible application of the most effective project management methods, tools, techniques and management approaches, often through a project management office (PMO). The Versatile Artist paradigm is characterised by organisations that employ senior and experienced project managers who develop their methodologies and work practices in accordance with project needs. The Agile Pragmatist paradigm emphasises process compliance with a time-phased delivery of functionalities.

The demand for agility imposes challenges for portfolio management and governance as best-practice frameworks for portfolio management (Jenner & Kilford, 2011; Project Management

Institute, 2017a) and IT governance (Agutter, 2020; ISACA, 2018) are more suited to stable environments where traditional command-and-control settings (Horlach, Schirmer, & Drews, 2019) work well. Taborda and Mohammed (2021) highlight how agile principles that promote self-organisation, iterative, value-driven delivery challenge the traditional top-down view of governance and present seven conceptual dimensions differentiating traditional and agile governance. Traditional project governance focuses on successful delivery of the iron triangle – time, cost and quality (Pollack, Helm, & Adler, 2018) which are used in measuring success. In the case of innovation and exploratory projects, which generally follow agile delivery methods, scope and goals are difficult to determine upfront and instead evolve in a continuous and iterative process, also referred to as ‘progressive elaboration’ (Collyer & Warren, 2009). This is because project requirements tend to be ambiguous and often change due to uncertainty. This in turn leads agile teams to use four different categories for iteration objectives – functionality, schedule, quality and team satisfaction (Drury-Grogan, 2014). Lappi, Karvonen, Lwakatare, Aaltonen, and Kuvaja (2018) compared traditional and agile project governance practices (as shown in *Figure 2-9*) to show which practices that exist in traditional governance can be transferred to Agile either as they are or with some modification and hence become a new Agile practice. It can be seen from their suggestions that customer-centricity and empowerment of teams makes Agile governance more suited to the changing demands faced by digital economy.

Traditional Practices That Can Be Transferred to Agile as Such	Traditional Practices That Can Be Modified to Agile	Emerging Agile-Specific Practices with No Regard to Traditional	Remaining Gaps Between Agile and Traditional Project Governance
Distinguished front-end phase for goal setting and business case	Scope setting is visionary and iterative process, starting with product vision and user stories	Inclusion of customer within the project team for decision making and validation efficiency	Linking and aligning the agile project objectives with organizations' strategy
Long-term post-project benefit realization	Short-term value realization through utilization of deliverables	Using validated—tested and accepted—deliverables as project performance and incentives basis	Connection between formal and flexible budgeting
Clear organizational and ownership structure	Flexible, value-based project budgeting and contracting	Change management is embedded into iterative planning and delivery	Impact of collaborative project models to agile project governance
External audits by third parties	Motivation through peer and team performance-related incentives	Agile philosophy and principles as project culture foundation	Conflict escalation and resolution within agile project
Project stakeholder management	Visual, real-time project monitoring through agile specific and traditional measures	Project team empowerment for autonomy, flexibility, and decision-making power	Application and impact of formalized project governance structures such as steering groups, portfolios, and programs in agile context
Project risk management	Product and sprint backlogs are the main planning and communication means	Project team composed of multitalented, self-organized, and cross-functional individuals, both experts and beginners	Project human resource and capacity management in agile
Necessary documentation for external validation and communication	Onsite, informal, and real-time communication (internal)	Agile mindset and technical capabilities as competence requirements	
Project capability management on organizational level through HR and PMOs	Project management daily activities are shared on team level	Project manager role changes to administrative and facilitative	
	Sprint retrospectives used as interproject learning method	Agile coach and scrum master manage agile project team performance	

Figure 2-9. Comparison of Traditional and Agile Project Governance Practices (Lappi et al., 2018, p. 55)

The fact that portfolio planning and budget cycles in most organisations are annual impedes such organisation from adapting to market changes and digital disruptions that require a more dynamic approach. The problem is compounded when agile delivery methods are used especially for DSIs where scope and design decisions are required in a typical two-week sprint cadence. Following the Goldilocks principle of being just right, Governance that is neither too light nor too overbearing is recommended (Knapp, Zeratsky, & Kowitz, 2016; Mullaly, 2009; Mullay, 2010; Williams, 2017).

In agile projects, the focus of the coordination and communication is to provide sufficient real-time information about the project's status to the empowered project team and involved customer representatives visually and in real time. To enable this, agile projects prefer product vision and backlogs instead of a formal, rigid project plan (Sheffield & Lemétayer, 2013). Scaled Agile frameworks that use a lean-agile approach are found to be more suitable for DSIs than traditional approaches to portfolio management, as illustrated in *Figure 2-10*.

Traditional Approach	Lean-Agile Approach
People organized in functional silos and temporary project teams	People organized in value streams and ARTs; continuous value flow
Fund projects and project-cost accounting	Fund value streams, Lean budgets and guardrails
Big up-front, top-down, annual planning and budgeting	Value stream budgets adjusted dynamically; participatory budgeting
Centralized, unlimited work intake; project overload	Strategic demand managed by portfolio Kanban; decentralized intake by value Streams and ARTs
Overly detailed business cases based on speculative ROI	Lean business cases with MVP, business outcome hypothesis, agile forecasting and estimating
Projects governed by phase gates; waterfall milestones, progress measured by task completion	Products and services governed by self-managing ARTs; objective measures and milestones based on working solutions

Figure 2-10. Traditional vs lean-agile Portfolio mindset (Scaled Agile, 2023)⁴

SAFe (Scaled Agile, 2023) is based on 10 underlying lean-agile principles which shape the governance and practices of SAFe (*Figure 2-11*). Some of these principles are relevant to management and governance of DSIs as guiding principles.

⁴ The Agile Release Train (ART) is a long-lived Agile team that incrementally develops, delivers, and often operates one or more solutions in a value stream. Each ART is a virtual organisation (typically 50 – 125 people) organised around the enterprise's significant Development Value Streams and exist solely to realise the promise of that value by building and delivering solutions that benefit the customer.

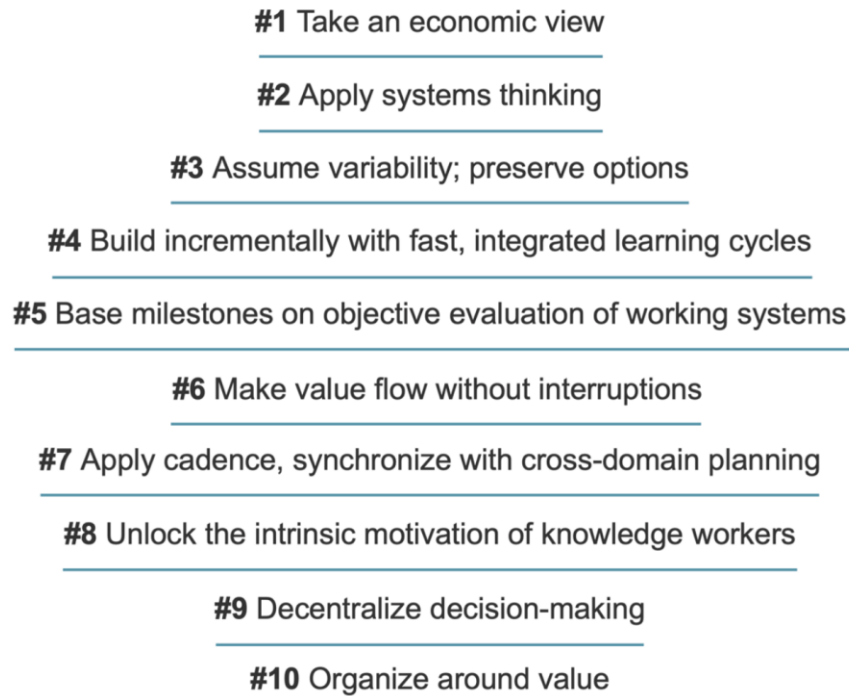


Figure 2-11. SAFe Lean-Agile Principles (Scaled Agile, 2023)

In SAFe, the portfolio aligns strategy to execution via a collection of development value streams which focus on building the right things with the appropriate level of investments in solutions for the portfolio to meet its strategic objectives, as shown in *Figure 2-12*. Key concepts used in SAFe portfolio management which can also be applicable to DSIs are:

- Value Streams – Every value stream must fund the people and resources necessary to build solutions that deliver value to the internal or external customer.
- Lean Budgets – Lean budgeting allows fast and empowered decision-making, with appropriate financial control and accountability through guardrails.
- Portfolio Kanban – The portfolio Kanban system makes the work visible and uses Work-in-Process (WIP) limits to help ensure that demand is matched to the actual capacity of value streams and ARTs.
- Portfolio Vision – The portfolio vision is a description of the future state of a portfolio's value streams and solutions and describes how they will cooperate to achieve the portfolio's objectives and the broader aim of the enterprise.

- Portfolio Canvas – The portfolio canvas provides critical inputs to the portfolio vision, Portfolio Backlog and Lean Budgets.

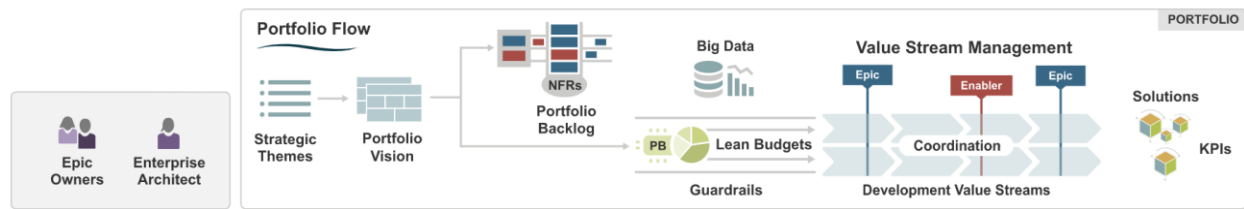


Figure 2-12. SAFe Lean Portfolio Management (Scaled Agile, 2023)

SAFe establishes guardrails to ensure that the mix of investments address both near-term opportunities and long-term strategy, that investments in technology, infrastructure and maintenance are not routinely ignored and that large investments are approved appropriately. As we shift the decision-making lower in organisational hierarchy, these guardrails can play a role in delivery of DSIs. Figure 2-13 illustrates the lean budget guardrails for:

- Guiding investments by horizon,
- Applying capacity allocation to optimise value and solution integrity,
- Approving significant initiatives, and
- Continuous business owner engagement

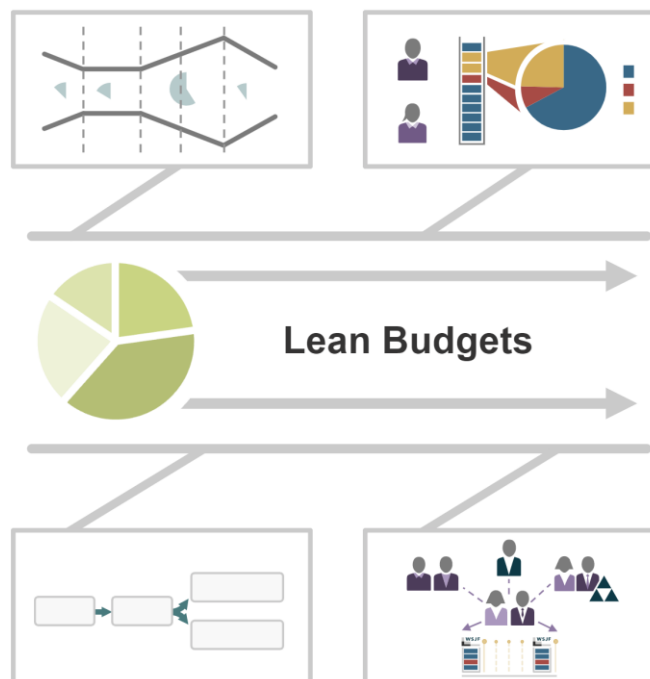


Figure 2-13. SAFe Lean Budget Guardrails (Scaled Agile, 2023)

SAFe Lean Portfolio Management relies on three significant events (*Figure 2-14*):

- Strategic Portfolio Review – typically held on a quarterly cadence, provides continuous strategy, implementation and budget alignment,
- Portfolio Sync – typically held on a monthly cadence, provides visibility into how well the portfolio is progressing towards meeting its objectives, and
- Participatory Budgeting – typically held on a six-monthly cadence, allows stakeholders to decide how to invest the portfolio budget across solutions and epics.

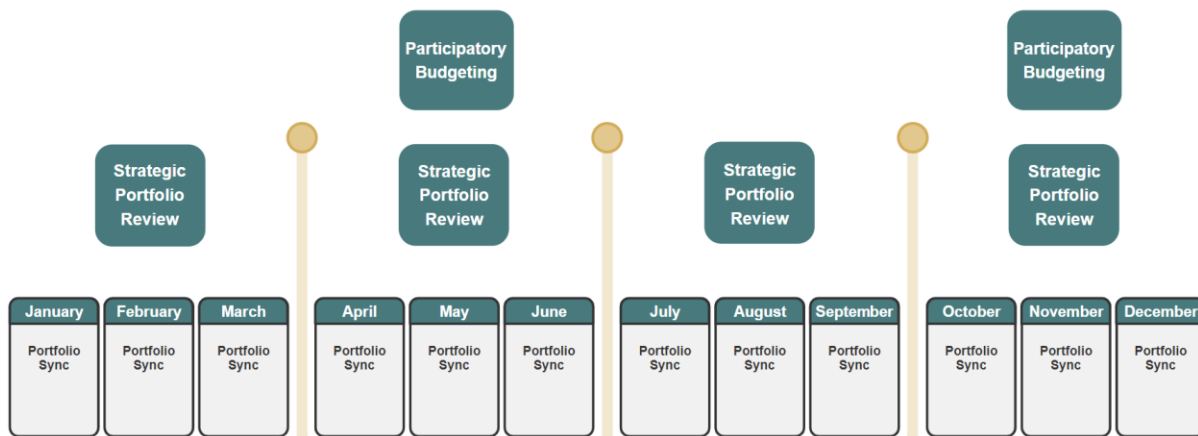
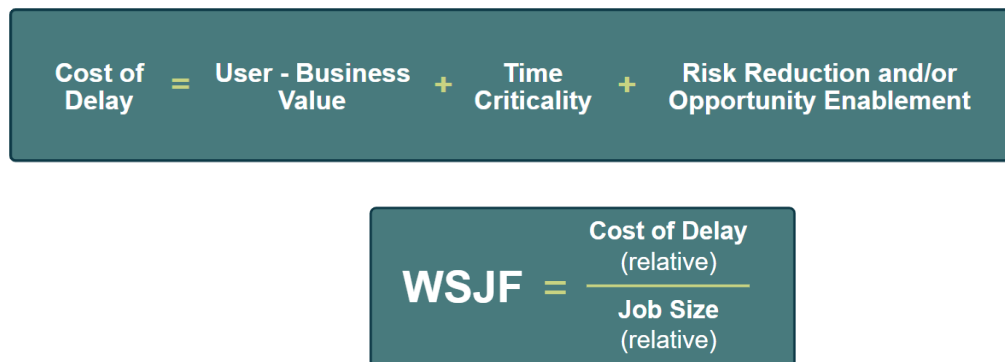


Figure 2-14. SAFe Lean Portfolio Management Events (Scaled Agile, 2023)

In Active Transport portfolio, I used these three events for delivering DSI. I found that while this was a shift from conventional portfolio management governance, it was quite effective. SAFe has program and solution backlogs as repositories for all the upcoming work that can affect the behaviour of the solution. The backlogs are a short-term holding area for features and capabilities that have gone through their respective Kanban systems and have been approved for implementation. SAFe has adopted the Weighted Shortest Job First (WSJF) using the cost-of-delay (CoD) concept (Reinertsen, 2009) as a method of prioritising. Cost of delay is calculated as the sum of user-business value, time criticality and risk reduction and/or opportunity enablement. WSJF makes a relative estimate by concentrating on one item at a time, evaluating the item with the smallest value as "one," and assigning the remainder to the relative value. Relative values are the Fibonacci numbers used in planning poker. Finally, the WSJF value can calculate the CoD as the job size (job duration) and the item with the highest WSJF value has the highest priority. At

the time of writing this thesis, Active Transport was using WSJF as shown in *Figure 2-15* as one of the lenses to prioritise initiatives covering engineering and DSIs.



$$\text{Cost of Delay} = \text{User - Business Value} + \text{Time Criticality} + \text{Risk Reduction and/or Opportunity Enablement}$$

$$\text{WSJF} = \frac{\text{Cost of Delay (relative)}}{\text{Job Size (relative)}}$$

Figure 2-15. Weighted Shortest Job First formula (Scaled Agile, 2023)

For performance monitoring, SAFe has three measurement domains – Outcomes, Flow and Competency which are defined as follows:

- Outcomes: How well do the portfolio’s solutions meet customers’ needs and provide the expected results for the business?
- Flow: How efficient is the portfolio at delivering a continuous flow of value to its customers and the desired outcomes for the business?
- Competency: How proficient is the organisation in the practices that enable business agility?

These three measurement domains can be applied at all levels of a SAFe enterprise to measure performance, be it SAFe Portfolio, Solution Train, Agile Release Train or Agile Team. SAFe further proposes minimum metrics, as in *Figure 2-16*, to measure portfolio performance against strategy implementation, spending alignment with the agreed boundaries and continuous improvement. Organisations can adapt these metrics to suit their specific domain requirements when delivering DSIs.

Goals	Measures	Desired Benefits
Employee Engagement	Employee surveys; Agile People Operations data	Improved employee satisfaction, higher employee engagement, and better business outcomes
Customer Satisfaction	Net Promoter Score (NPS)	Improved customer experiences and loyalty
Partner Health	Partner and vendor surveys	Improved ecosystem relationships
Business Agility	Business Agility self-assessment	Improved ability to respond to market changes, emerging opportunities and threats
Portfolio Performance	Objectives and Key Results (OKRs); LPM Self-Assessment	Better alignment and improved progress toward the portfolio's strategic themes, and portfolio performance
Value Stream Performance	Value Stream KPIs	Improved performance against the forecasted business outcomes
Program Predictability	Program predictability measure	Improved actual business value achieved
Time-to-Market	Number of releases; feature cycle time	More frequent releases; faster feature delivery
Relentless improvement	Self-assessments for each level of the Framework	Relentless improvement in team, program and portfolio performance
Quality	Defect count and cycle time; support call volume; escaped defects	Improved customer satisfaction; reduced support call volume, and lower product development costs

© Scaled Agile, Inc.

Figure 2-16. Lean Portfolio Metrics (Scaled Agile, 2023)

2.4 THE DELIVERY OF DATA SCIENCE INITIATIVES

While domains such as program management, data management, data science and development-operations (DevOps) were well established for us to explore, development-security-operations (DevSecOps) emerged as the best option considering the need for securing the data published on the internet. While exploring DevSecOps, I compared the Waterfall and Agile Delivery Methods for Software Delivery to see how complex interdependent initiatives might influence the application of these methods. As the teams grew from one to multiples very quickly, I reviewed scaled agile frameworks for large-scale agile delivery. The need to explore change management came up later as I realised that while organisations were investing heavily in DSIs, I questioned if they were understood the benefits of such investments or were just following the next technology trend.

I start with definitions of big data, business intelligence and data science to introduce them as foundational elements of DSIs, to identify gaps and lead up to my research questions. The term “big data” has become ubiquitous and several authors have attempted to standardise its definition (De Mauro, Greco, & Grimaldi, 2016; Ward & Barker, 2013). An appropriate definition in the

context of this research is “big data is the information asset characterised by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value.” (De Mauro et al., 2016). However, a comprehensive literature review and synthesis of critical success factors of big data projects by Saltz and Shamshurin (2016) shows lack of an agreed standard to execute them and a recommendation to improve process methodology and prioritises 33 success factors. Surbakti, Wang, Indulska, and Sadiq (2020) identified 41 factors and used thematic analysis to arrive at seven main themes of factors, namely organisational aspects; systems, tools and techniques; people aspects; data privacy and security and governance; data quality; process management and perceived organisational benefit in the effective use of big data. Similarly, Al-Sai, Abdullah, and Husin (2020) in their systematic literature review, identified 74 CSFs for big data and proposed a classification schema and framework in terms of five categories, namely Organisation, Technology, People, Data Management and Governance. Business intelligence is an umbrella term that includes applications, tools, infrastructure and practices to enable access and analysis of information to optimise performance and decision-making (Gartner, 2023; Halper, 2015; Larson & Chang, 2016). Data science is the study of the generalisable extraction of knowledge from data (Dhar, 2013) and can be defined as the collection, preparation, analysis, visualisation, management and preservation of large collections of information to generate actionable insights (Donoho, 2017; Saltz & Stanton, 2017). Whereas data analytics is primarily focused on understanding datasets and obtaining insights which can be turned into actions, Data Science is centred on building, cleaning and organising datasets. Data science is a discipline that is built on a foundation of critical thinking (Stobierski, 2021). DSIs often include implementation of artificial intelligence (AI) and machine learning (ML). AI applies advanced analysis and logic-based techniques, including ML, to interpret events, support and automate decisions and take actions (Gartner, 2020). ML is an application of AI that provides systems with the ability to automatically learn and improve from experience without being explicitly programmed. ML focuses on the development of computer programs that can learn from data, identify patterns and either assist or make decisions with minimal human intervention.

For the DSIs we delivered, I found that it was never about bringing in data from one source and closing the project. Every DSI ingested a series of interrelated datasets to generate value and I questioned if we should treat them as a project or a program. I saw DSIs being typically

implemented as a program on a continuous spectrum rather than a single one-off project and I focused on program management rather than project management processes for delivery.

A program is defined as related projects, subsidiary programs, and program activities managed in a coordinated manner to obtain benefits not available from managing them individually (Project Management Institute, 2017b). Program management is defined as the application of knowledge, skills and principles to a program to achieve the program objectives and to obtain benefits and control not available by managing program components individually (Project Management Institute, 2017b). A review of typical program lifecycle has identified gaps in using it for DSI delivery. *Figure 1-1* shows a typical program lifecycle followed by gaps and issues against deliverables for each of the three phases of a program in Table 1-2.

Managing Successful Programmes (MSP) is another program management framework whereby large, complex change can be broken down into manageable, interrelated projects (Axelos, 2022). The MSP framework is based on three core concepts: MSP Principles, MSP Governance Themes and MSP Transformational Flow. Both the PMI and Axelos standards are widely accepted and used in the industry with PMI being principle-based and Axelos providing detailed guidance on program management. Any DSI delivery framework must be flexible to integrate the program management standard the organisation has implemented.

Value realisation for any program occurs when the product and service created are adopted successfully by users. This is usually achieved when the changes required to adopt the outputs from the projects create the desired outcomes. Change management is a systematic approach that includes dealing with the transition or transformation of organisational goals, core values, processes or technologies. Kotter's Change Management Model (J. Kotter, 2007), McKinsey's 7-S Change Management Model (Lorenzi & Waterman, 1985), the ADKAR change management model (Hiatt, 2006) and the Kübler-Ross Five Stage Change Management Model (Kübler-Ross, 2009) are some of the widely adopted change management models in industry. Access to big data and the availability of analytics solutions have provided organisations with new ways of working especially for benefits tracking and investment decision-making. I see effective use of a change management model in DSI as essential to value realisation because each dataset brings additional insight, complexity and need for change, all of which needs to be embedded in the organisation.

Data management is the development, execution and supervision of plans, programs and practices that deliver, control, protect and enhance the value of data and information assets throughout their lifecycles (Earley, 2017). *Figure 2-17* defines the 11 data management knowledge areas with data governance at the centre of wheel and other knowledge areas, while necessary, can be implemented at different times depending upon the requirements of the organisation. I see the 11 knowledge areas of data management as foundational elements for the DSI delivery team.

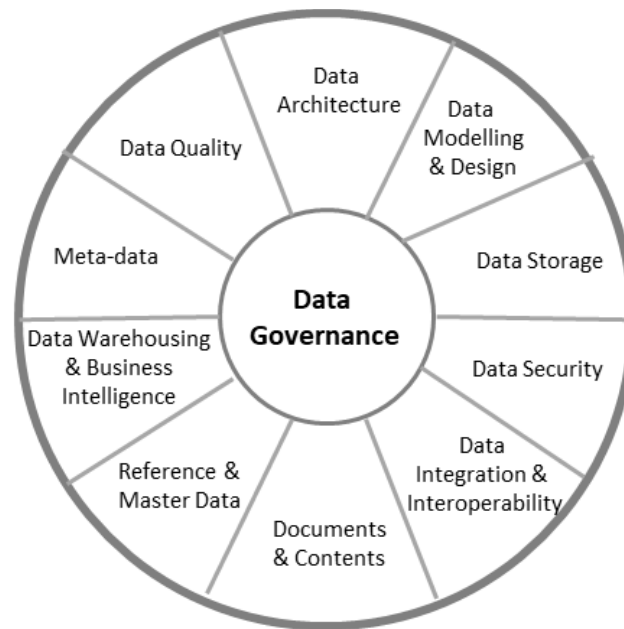


Figure 2-17. Data Management Framework (Earley, 2017, p. 36)

The two major methodologies used in software development are waterfall and agile. Waterfall projects are plan-driven and completed sequentially whereas agile projects are delivered iteratively in a cycle. Waterfall has specific deliverables for each phase and a review process while agile allows continuous delivery of value throughout a project. The agile manifesto and the 12 principles of agile software (The Agile Alliance, 2001) are widely adapted and used in software development. The agile manifesto equips software development teams with a flexible framework to guide their project management processes and uphold agile best practices. It promotes individuals and interactions over process and tools, working software over comprehensive documentation, customer collaboration over contract negotiation and responding to change over following a plan. Belling (2020) mentions that many organisations operate on a continuum between Waterfall and Agile methods and discusses a growing number of AgileFall, ScrumBan, Wagile, LeanBan and ScrumBut hybrid models of project management which currently exist.

Viaene and Van den Bunder (2011) have attempted to define the skills that differentiate successful business analytics project managers to traditional project managers. They recommend a more adaptive experimentation-based approach emphasising limited early planning, good-enough requirements and experimental and evolutionary design with significant ongoing learning and change is better suited to analytics projects. Through a survey of 237 practitioners, Franková et al. (2016) found that agile and iterative delivery is more appropriate to big data projects. In summary, evolving research points to using agile methods and not waterfall methods to deliver DSIs effectively.

A need to align the priorities respectively of the software development and operations teams for successful project execution has led to the concept of DevOps which is described as the conceptual and operational merging of development's and operations' needs, teams and technologies (Cois, Yankel, & Connell, 2014). With growing realisation that software solutions must embed security as part of design, the concept of DevSecOps evolved to create and include modern security practices in the fast and agile world of DevOps (Myrbakken & Colomo-Palacios, 2017). I argue that any DSI delivery framework should consider having a DevSecOps domain in its core to counter cyber-security threats as well as align the development and operations teams.

A requirement for large projects, particularly those with globally distributed teams needing to collaborate and coordinate, has led to the popularity of such frameworks as Scaled Agile Framework (SAFe), Large-Scale Scrum (LeSS) and Lean Scalable Agility for Engineering (LeanSAFE). (Ebert & Paasivaara, 2017; Leffingwell, 2007). I see scaling as important in the context of DSIs for the same reason: multiple geographically-spread teams within an organisation are often involved in delivering data science outcomes. A comparison of various five scaled agile frameworks as shown in *Figure 2-18* shows the strengths across 16 criteria (Atlassian, 2020).

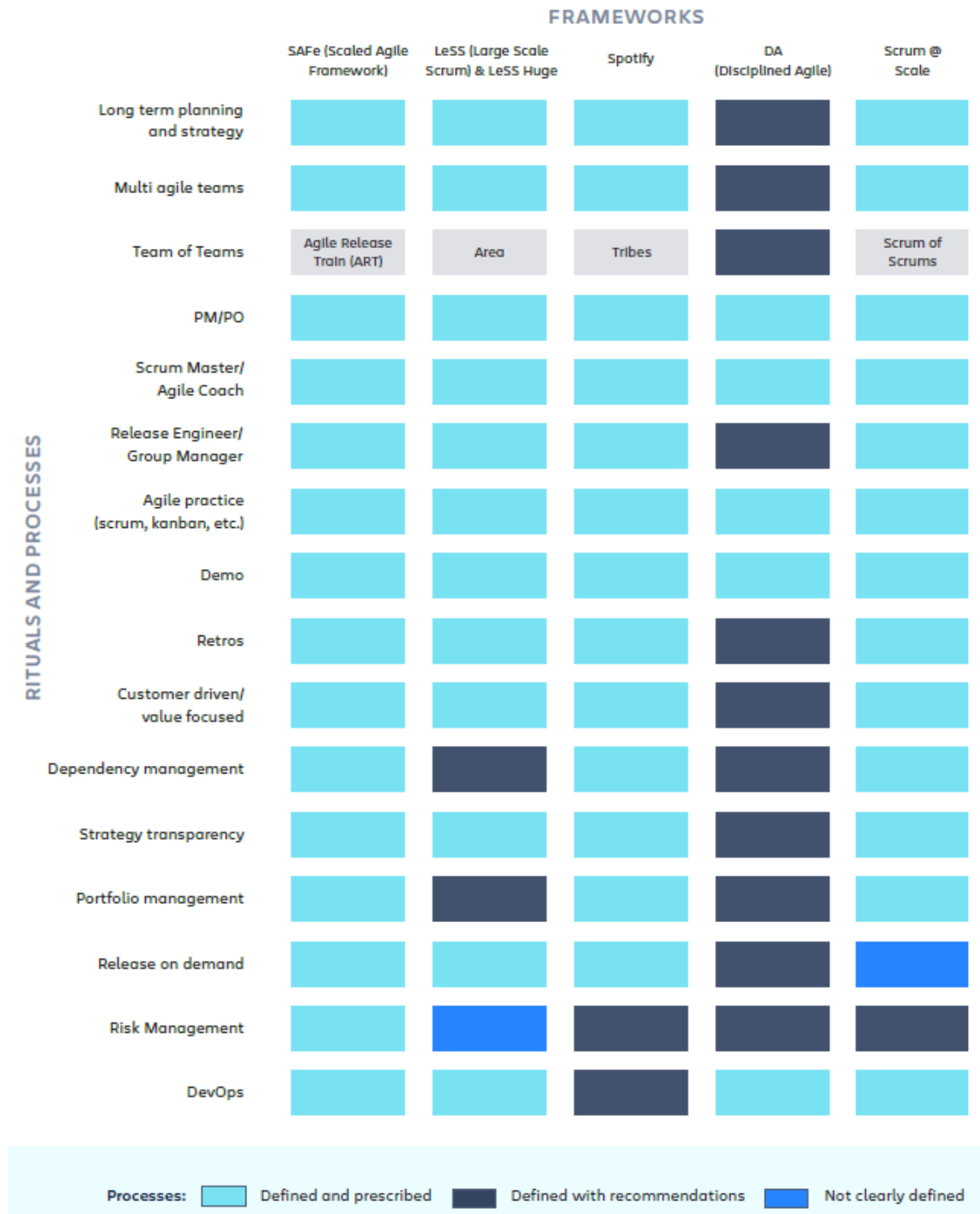


Figure 2-18. Difference between frameworks for scaling agile (Atlassian, 2020)

The five frameworks mentioned above are both rich and mature and have been adopted by various organisations globally. While there is no right way to scale agile, Atlassian (2020) describes these as the top agile frameworks with which many organisations have had great success

in evolving their processes, teams and cultures. Due to the depth covered by these frameworks, that is, from portfolio management and project management to product management, organisations generally pick and choose elements of the framework based on strategic priority. For example, we implemented SAFe in Transport, starting with an Essential SAFe component and at the time of writing we were implementing lean portfolio management. As these frameworks are detailed with processes covering portfolio, program and team levels and not every process may contribute to successful delivery of an initiative at the same level, Khoza and Marnewick (2023) propose a conceptual framework for scaled agile success. Their framework determines the perceived indicators to be achieved by organisations at each level and allows management and practitioners to identify specific processes to implement.

Before I discuss DSIs in detail (2.5), I introduce here data science and its associated delivery processes. Data science can be defined as the collection, preparation, analysis, visualisation, management and preservation of large collections of information to generate actionable insights (Dhar, 2013; Saltz & Stanton, 2017). Data analytics and business intelligence have been used by organisations to provide answers to specific questions based on existing data. Data science goes one step further by parsing large datasets, sometimes in unstructured ways, to expose insights and undertake predictive analysis. Delivery of DSIs requires that practitioners and management have a good understanding of data mining and data science delivery processes. Six widely-used frameworks in the delivery of DSIs are the Knowledge Discovery in Databases (KDD) model (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), Cross-Industry Standard Process for Data Mining (CRISP-DM) as shown in *Figure 2-19* (Chapman et al., 2000), Sample, Explore, Modify, Model and Assess (SEMMA) (SAS Institute, 2009), Obtain, Scrub, Explore, Model and Interpret (OSEMN) (Mason & Wiggins, 2010), Team Data Science Process (TDSP) (Severtson, Franks, & Ericson, 2017) and Foundational Methodology for Data Science (FMDS) (Rollins, 2015).

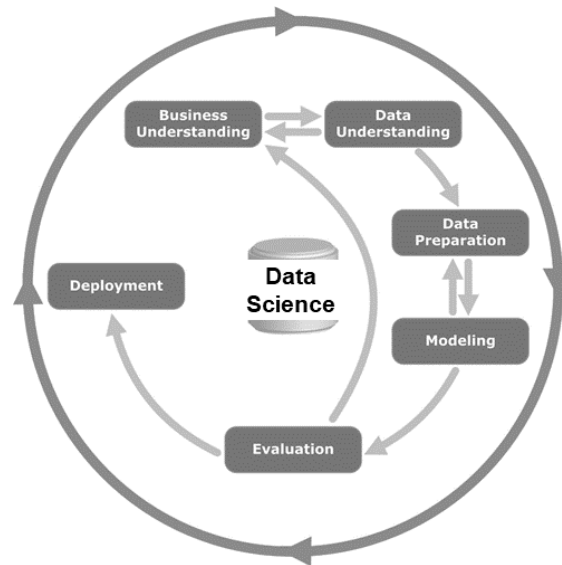


Figure 2-19. CRISP-DM Data Science Processes (Chapman et al., 2000, p. 13)

Foroughi and Luksch (2018) compared the KDD, CRISP-DM, FMDS and TDSP processes against four common iteratives stages of problem definition/formulation, data gathering, data modelling and data production as shown in *Figure 2-20*. They found that the KDD process did not cover the business understanding and deployment phases that the CRISP-DM methodology did. However, the CRISP-DM did not have the analytic approach, identification of suitable data collection strategy and data resources and the feedback phases of FMDS. While FMDS and TDSP were very similar, the detailed stages of FMDS would have been more useful for a wide range of projects. TDSP used a specific set of Microsoft tools and infrastructure to deliver intelligent applications by deploying ML or AI models.

KDD Process	CRISP-DM	FMDS	TDSP
-	Business Understanding	Business Understanding	Business Understanding
-	-	Analytic Approach	
-	-	Data Requirements	
Selection	Data Understanding	Data Collection	Data Acquisition and Understanding
Pre Processing		Data Understanding	
Transformation		Data Preparation	
Data Mining	Modelling	Modelling	Modelling
Interpretation/ Evaluation	Evaluation	Evaluation	
-	Deployment	Deployment	
-	-	Feedback	Customer Acceptance

Figure 2-20. Comparison of data science methodologies (Foroughi & Luksch, 2018, p. 10)

Clearly, the six data mining and data science frameworks have their advantages and disadvantages in their use. In their systematic literature review, Schröer, Kruse, and Gómez (2021) found that CRISP-DM was the de facto standard for an industry-independent process model for data mining projects. At Transport, CRISP-DM was widely chosen by teams delivering DSIs and I have included it in the DSI Delivery Framework for Data Science Processes domain.

2.5 CONCLUSION

Considering the uncertainty and ambiguity that surrounds DSIs, I commenced this chapter with an overarching discussion on exploratory and exploitative projects. I then expanded the review into three themes that impact the management and governance of DSIs, namely, business case, governance and delivery, which lead us to the research questions. In the business case section, I reviewed both traditional and lean business cases to examine their suitability for DSIs as part of portfolio management. In the governance section, I provided the context of corporate, portfolio and program governance in which DSIs operate to ensure that organisations implement a governance structure that will improve the success rate of DSIs. In the delivery section, I reviewed various domains essential to developing a program delivery framework for DSIs. My framework required an end-to-end program management system from business case through to benefits realisation. Therefore, I reviewed change management to ensure that investment in DSI delivers value to the organisation. I reviewed several scaled agile frameworks to assess their suitability for both smaller and large-scale DSI implementations. Since an understanding of data management is essential to all DSIs, I reviewed DAMA's DMBok. I then reviewed several of the data mining and data science methods used since late 1990s to find that the FMDS and TDSP from IBM and Microsoft respectively were built upon the solid foundations of KDD and CRISP-DM. The five domains covered in this chapter—Program Management, Change Management, Scaled Agile, Data Management and Data Science Processes—provide us with a comprehensive set of tools and techniques to deliver DSIs.

Through the review of related literature, it became evident that existing practices often fall short in adequately addressing the uncertainty, complexity, and exploratory nature inherent in DSIs. While current approaches may provide some level of guidance, they often lack the agility and adaptability required to navigate the dynamic landscape of DSIs effectively.

The inadequacy of traditional project management frameworks and governance structures in accommodating the unique challenges posed by DSIs underscores a critical research gap. Despite the growing importance of DSIs in driving organisational transformation and innovation, there remains a notable lack of understanding regarding how best to define and execute them within portfolios and programs. Existing practices, while well-established in managing more predictable projects, struggle to cope with the inherent ambiguity and evolving nature of DSIs. This gap not only impedes the successful delivery of DSIs but also undermines their potential to generate value for organisations.

By elucidating the limitations of current practices in addressing the complexities of DSIs, this research aims to bridge the identified gap and contribute to advancing the management and governance of DSIs. Through a comprehensive examination of the research questions outlined in Section 1.4 Research Questions, this thesis seeks to provide actionable insights and frameworks that empower organisations to navigate the uncertainties of DSIs more effectively. By integrating elements of program management, change management, scaled agile, data management, and data science processes, the proposed approach offers a holistic framework tailored to the unique needs of DSIs.

In summary, this chapter highlights the pressing need to enhance our understanding of the uncertainty, complexity, and exploratory nature inherent in DSIs. By elucidating the shortcomings of current practices and articulating the research gap, this research sets the stage for a more robust and adaptive approach to managing and governing DSIs. It is imperative to reframe the research gap as frontier of portfolio and program management and the need for a paradigm shift in practices to accommodate these emerging complexities adequately. Through the exploration of key themes and research questions, this thesis seeks to drive meaningful advancements in the field, ultimately enabling organisations to unlock the full potential of DSIs and drive sustainable value creation.

3. Research Design

This chapter describes the design adopted for this research to achieve the aims and objectives of developing an effective framework for management and governance of DSIs, as stated in section 1.3 of Chapter 1. In this chapter, section 3.1 discusses the methodology used in the study; section 3.2 covers the stages by which the methodology was developed, and the research design; section 3.3 details the participants in the study; section 3.4 lists all the instruments used in the study and justifies their use; section 3.4.2 details validity and reliability of the framework and the final section, 3.4.3, discusses how the data was analysed. Ethical considerations undertaken in the research include obtaining Ethics Approval from UTS, ensuring participant confidentiality and obtaining informed consent from participants.

3.1 METHODOLOGY

The research combined the learnings from delivering DSIs using a multi-methods approach (Hunter & Brewer, 2015; Straits & Singleton, 2018) with case studies and semi-structured interviews to develop a DSI delivery framework. This approach was chosen to ensure a comprehensive understanding of the complexities involved in managing and governing DSIs. The framework was developed through case studies undertaken at Transport, providing rich insights into the real-life experiences and challenges faced in DSI delivery and has been implemented in the Active Transport branch of Transport since 2021. These case studies were complemented by semi-structured interviews with practitioners and PMO/portfolio managers, allowing for a deeper exploration of key themes and issues. The active involvement of participants ensured that they not only provided valuable insights but also played a crucial role in validating and refining the proposed framework. Overall, the methodology employed in this research aimed to enhance transparency and rigor, thereby strengthening the credibility of the findings.

Throughout the thesis, my personal perspective and decision-making have significantly influenced the research process and its outcomes. While this chapter delineates the methodology, it mirrors a personal voyage of exploration, learning, and decision-making, all of which have

substantially contributed to crafting and honing the research framework. For instance, the selection of case studies and participants for semi-structured interviews, as well as the development of the DSI delivery framework, minimum viable governance principles, and business case guidelines, have been informed by a blend of existing literature, practical experience, and feedback from stakeholders. In essence, while maintaining an academic tone, the thesis is fundamentally shaped by personal experiences, viewpoints, and choices that have permeated every facet of the research journey.

3.2 RESEARCH DESIGN

The research was underpinned by a pragmatic research philosophy, as the reality of managing and governing DSIs is based on the real-life experiences and views of stakeholders involved in them at Transport. The requirement to deliver seven DSIs at Transport and the challenges faced while delivering them compared to other ICT programs provided the motivation and inspiration for this research and the model for the case study method.

The research methodology followed the approach shown in *Figure 3-1*, involving participants in both understanding the problem and developing potential solutions. Participants played an active role in providing insights that informed the initial framework and subsequently tested the proposed framework during the pilot and final implementation phases.

In the Problem Definition phase, the research embarked on a comprehensive exploration of the underlying challenges and dynamics of DSIs within the Transport context. Through an in-depth analysis of Transport case studies, the study elucidated the ontological and epistemological assumptions inherent in DSIs, striving to achieve theoretical saturation. Concurrently, semi-structured interviews (Appendix B: Semi-Structured Interview Guide) with practitioners from diverse organisational backgrounds provided invaluable insights and validation, ensuring a robust foundation for subsequent solution development.

During the Solution Development phase, the research translated the rich insights garnered from the problem definition into actionable frameworks to address the identified challenges. A meticulously crafted DSI delivery framework emerged, encapsulating key aspects such as program lifecycle management, agile methodologies and change management practices. Simultaneously,

minimum viable governance principles and business case guidelines were formulated, aiming to strike a delicate balance between innovation and governance in the management of DSIs.

The Validate Trustworthiness & Reliability phase marked the transition from theory to practice, as the developed frameworks underwent rigorous validation and refinement. Through a pilot implementation within the Transport organisation, the efficacy and feasibility of the DSI delivery framework were thoroughly evaluated in real-world contexts. Iterative feedback loops facilitated continuous refinement, ensuring the adaptability and responsiveness of the framework to evolving organisational needs and challenges.

Finally, the Finalise Findings phase culminated in the dissemination and application of the validated frameworks. The finalised DSI delivery framework was published, making it accessible for broader adoption and implementation across diverse organisational contexts. Through practical use cases and demonstrations, the research effectively showcased the value proposition of DSIs, empowering stakeholders to make informed investment decisions and drive organisational success in the dynamic digital landscape.

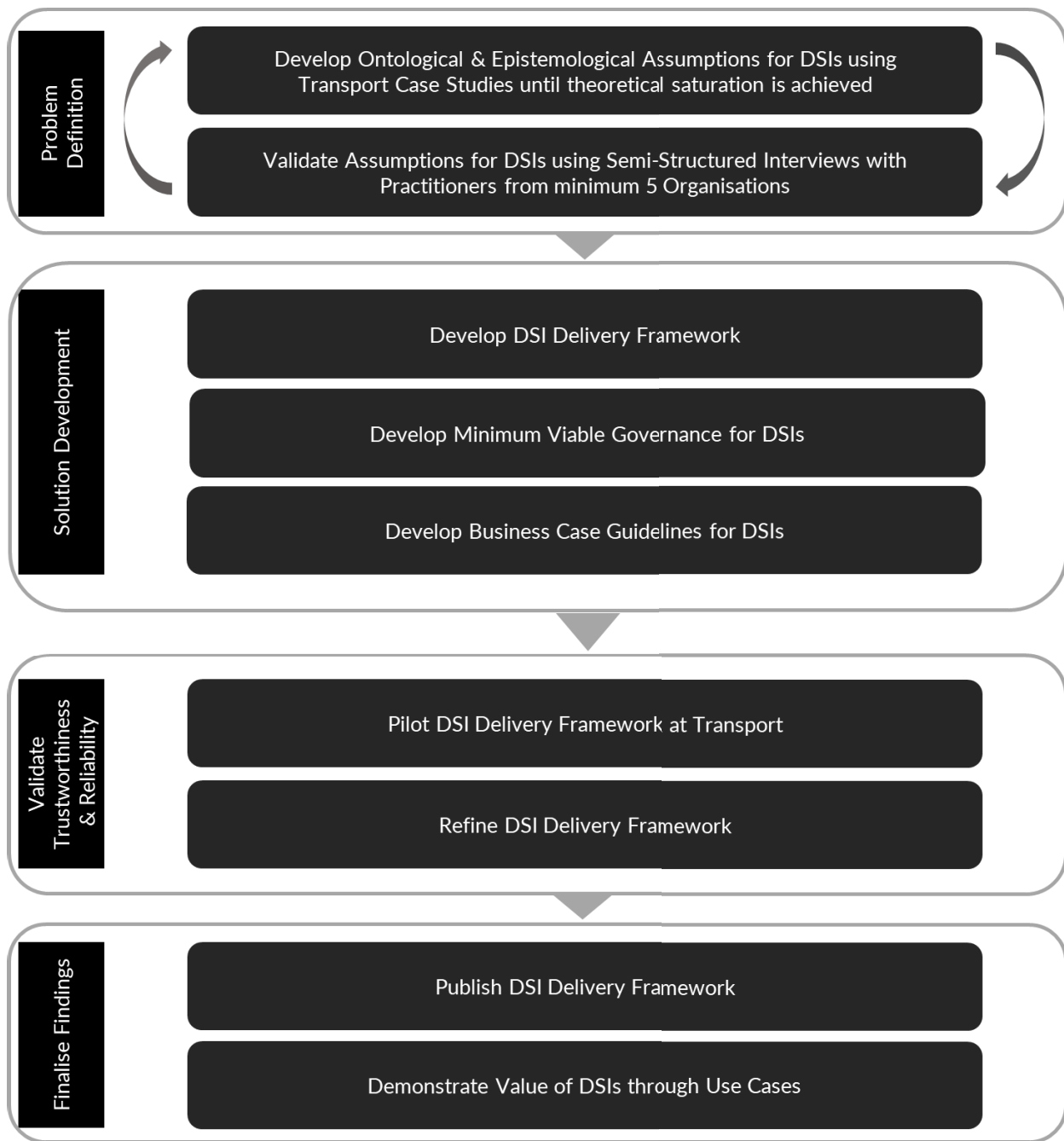


Figure 3-1. Research Methodology Approach

My lack of success in delivering DSI while having a successful track record otherwise in delivering ICT programs triggered my search for knowledge, beginning with contacting UTS to undertake doctoral research on this topic. Seven DSIs delivered by and for Transport between 2017 and 2022 presented the opportunity to use Eisenhardt (1989) case study method for the initial qualitative phase of the study and the ensuing research. As such, they reflect the chronology and

maturity of Transport in delivering DSIs (*Figure 4-1*). The number of cases selected follows the recommendations made by Small (2009) and Yin (2018) for setting up a multiple case study design, with the extreme cases contributing to theoretical replication (predicting contrasting results) and the semi-structured interviews with five different organisations providing literal replication (finding similarities). In this way, the ontological and epistemological findings on DSIs derived from the case studies were validated with DSI practitioners from a minimum of five different organisations. The two processes in the problem definition phase were repeated until theoretical saturation was achieved, at which point the phase was terminated.

Using data from the problem definition phase, this research refined the existing frameworks proposed in the program management literature to develop a suitable framework to manage and govern DSIs as innovation projects. The new framework covered the complete program life cycle of program definition, program delivery and program closure. Apart from program management, the framework also included the change management, scaled agile, data management and data science process domains to arrive at a comprehensive framework. Bearing in mind that too much governance can stifle innovation in organisations and too little governance can waste precious organisational resources, this research investigated agile governance at project, program and portfolio level for DSIs and developed guiding principles that focus on product and portfolio governance. Considering the exploratory and innovation aspects of DSIs, the research reviewed how to build better business cases by moving from a project to a product mindset, funding value streams and acknowledging the exploratory nature of DSIs; this indeed allowed Transport to build better business cases.

Reliability is concerned with questions of stability and consistency and validity refers to the congruence of “goodness of fit” (Straits & Singleton, 2018). In this research, the delivery framework was expected to support efficient delivery of DSIs to other parts of Transport as well as to external organisations. To validate the reliability of the delivery framework, therefore, a pilot was conducted at Transport. The building of any framework requires multiple iterations, and the pilot allowed me to validate the delivery processes and underpinning assumptions. It allowed me to answer the questions of what the final product should look like, whether the product was feasible and were there any potential issues (Laukat & Wedel, 2020). The delivery framework was refined according to the continuous feedback from the pilot, feedback from the conferences where it was

presented, and feedback from the “Management and Governance of DSI” microcredential course at UTS where the framework was taught to both academics and practitioners. While investment in DSIs continues to increase, demonstration of the value they deliver is important. This research captured cases from Transport that show how organisations can leverage the power of DSIs to track benefits and make investment decisions.

3.3 PARTICIPANTS

As part of the semi-structured interviews for the problem definition, validate trustworthiness and reliability phases, I recruited two sets of participants, practitioners and PMO/portfolio managers. An ideal practitioner was someone who had a minimum of five years’ experience in delivering projects/programs and a minimum of two years’ experience in delivering DSIs. An ideal PMO/portfolio manager participant was someone who had a minimum of five years’ experience in managing portfolios with DSIs. These criteria ensured that survey participants had the experience to reflect upon the initial content of the thesis. Due to the practicalities of cost, access and safety during the Covid-19 pandemic, all interviews were conducted online. I was mindful of engaging participants who represented diversity in terms of age, gender, company type and background experience that at the same time reflected a DSI ecosystem. While this research did not attempt to seek any statistical representation, it monitored the risk of being skewed if it did not represent a DSI ecosystem.

During the validation phase, it is important to acknowledge the limitations of the study. One such limitation is the relatively small sample size of the semi-structured interviews, consisting of only 5 respondents. While these interviews provided valuable insights into the perspectives of practitioners and PMO/portfolio managers, the small sample size may limit the generalisability of the findings. Additionally, the small sample size introduces the potential for biases, as the perspectives of a larger and more diverse group of participants may not have been fully represented. Despite these limitations, efforts were made to ensure that the participants were selected based on their expertise and experience in delivering DSIs, thereby enhancing the credibility of the insights gathered. By acknowledging these weaknesses, I aim to maintain transparency and uphold scholarly integrity in my research endeavour.

3.4 INSTRUMENTS

The instruments used were based on the four phases of research methodology, as described below.

3.4.1 Problem Definition Phase Instruments

This phase used seven DSIs at Transport as case studies followed by validation of assumptions with practitioners through semi-structured interviews.

My data collection and analysis used field-studies standards (Gibbert & Ruigrok, 2010; Yin, 2018) to collect data from seven Transport DSIs. These were selected because I was involved in their end-to-end delivery and thus had access to the data. The timing of this research allowed both real time and retrospective data collection. While the scale of the DSIs was different, together they paint a good picture of unique characteristics and business cases.

The data collection included observation of daily activities and review of the program artefacts produced throughout the program. The program artefacts used as instruments included a program management plan, a schedule, a register of risks, assumptions, issues and dependencies (RAID), Jira Align, Jira and Confluence⁵, field notes, communications and minutes for the data collection. I was copied on all key program emails. To quantify the artefacts, some (program management plan, schedule, and raid register) were created at the beginning of the DSI and updated monthly at a minimum; stories and sprint performance reports were updated fortnightly, communication was generally milestone based and formal governance packs were created monthly. I used the OneNote tool for field notes throughout the period (2017 to 2022), especially during key ceremonies, to record my observations. Outputs from the retrospectives ceremony done at the end of each bi-weekly Agile sprint were categorised to gain insights into the sentiments of the delivery team at different points of the DSI delivery. These outputs were in the form of

⁵ Jira and Confluence are collaboration tools from Atlassian used by agile practitioners, Jira is a project management tool to plan, track and release software. Confluence is a team workspace to create, capture and collaborate on a project or idea.

comments from team members and stakeholders attending the ceremony captured on a digital board such as Mural.

As part of the problem definition phase, participants were involved in providing insights into the unique characteristics of DSIs, delivery methods, and learnings from the delivery of DSIs. This input was crucial in forming the basis of the initial framework.

The semi-structured interviews with practitioners were conducted to validate the following:

- Definition of DSI initially created using the Transport context
- Unique characteristics of DSIs
- Delivery methods used, and
- Learnings from delivery of DSIs.

3.4.2 Validate Trustworthiness and Reliability Phase Instruments

This phase piloted the DSI delivery framework and validated the framework through its use on all seven DSIs and the continuous feedback received from stakeholders. During the pilot, the approach used was the same as that used for the data collection of the seven DSI case studies, that is, retrospectives, program artefacts and field notes. The following components of the solution were validated:

- Understanding of the proposed DSI delivery framework,
- Identification of gaps and coverage, and
- Identification of risks in implementing the proposed DSI delivery framework.

Participants continued to play an active role during the pilot phase, providing feedback and insights that were instrumental in refining and validating the proposed DSI delivery framework. Their continuous engagement ensured the reliability and trustworthiness of the framework as it was iteratively improved.

3.4.3 Analysis

The information collected from the seven DSIs via the retrospectives, program artefacts and field notes was reviewed and analysed in both raw and coded form. As data was collected over six years, a great reliance was placed on the use of coding and analysis. Analysis was undertaken with

the support of software tools (primarily Microsoft Excel and Atlassian Confluence) and was an advocated method for examining quantitative data in business research, where I was seeking transferable, dependable, credible and generally trustworthy reporting (Pope, Ziebland, & Mays, 2000; Straits & Singleton, 2018; Webb, Campbell, Schwartz, & Sechrest, 1999; Yin, 2018).

Coding data is the process of systematically categorising the collected data into manageable patterns, allowing for the identification of recurring themes and insights.. It is a means to synthesise large volume of program artefacts and field notes to a stage where patterns, observations and understanding may be grounded (Pope et al., 2000). Initially, the data was reviewed in its raw form to gain a comprehensive understanding of the content. Subsequently, codes were applied to relevant sections of the data, enabling the identification of commonalities, differences, and emerging patterns. This process was iterative with codes being refined and adjusted as new insights emerged. Through constant comparison and validation, themes began to emerge organically from the coded data, providing a deeper understanding of the nuances and complexities within the dataset. The themes were then further analysed and synthesised to develop a coherent narrative that informed the development and validation of the DSI delivery framework. It allowed me to review and implement minimum viable governance and develop guidelines for building better business cases for DSIs. It also allowed me to capture examples of how DSIs were used to track benefits and investment decision-making at Transport. While the coding process was integral to the analysis, it is important to note that its effectiveness relied on my judgment, expertise, and constant engagement with the data throughout the research journey.

3.5 CONCLUSION

This chapter describes the overall design adopted by the research to develop an effective framework for management and governance of DSIs. The research process combined a multi-methods approach involving participants not only in understanding the problem but also in validating and developing the final framework. This approach ensured that participants were actively engaged in shaping the problem definition and refining the proposed framework. Their involvement enabled the creation of a robust framework that reflects the real-life experiences and views of stakeholders involved in DSIs at Transport. These foundations allowed the research to proceed to data collection and detailed analysis.

4.Data Collection and Analysis

4.1 INTRODUCING THE TRANSPORT DATA SCIENCE INITIATIVES

The seven DSIs chosen as case studies represent contemporary data science phenomena within the real-world context (Yin, 2018) at Transport. These were particularly useful in addressing the research questions because the organisation needed to better understand the unique characteristics of DSIs so as to better manage and govern them. While the data presented in Table 4-1 and subsequent tables is attributed to the delivery team, it is essential to clarify my individual contribution to the analysis process. My role primarily involved synthesising and interpreting the data collected by the delivery team, applying qualitative assessments, and deriving insights relevant to the research questions. Table 4-1 summarises the essential characteristics of the DSIs, with details obtained through qualitative assessments by the delivery team, including delivery uncertainty and dominant program types.

Table 4-1
Summary of Transport DSIs

Program	Description	Period	Budget (AUD)	Delivery Uncertainty	Dominant Program Type
Vanguard	Consolidate and disseminate data & information to contribute to a public transport network where customers and staff feel safe & always travel with a valid ticket.	Jan 2017 - Mar 2019	\$5.14m	High	Exploration
Ferry	Implement evidence-based Ferry Contract Management & improved customer experience.	Apr 2018 - Dec 2020	\$4.8m	High	Exploration
CTABS	Enable data analytics and verification of Provider self-reporting.	Oct 2017 - Mar 2018	\$289k	High	Exploration
PTIPS Analytics	Conduct a proof of concept of Azure big-data platform by using PTIPS (Public Transport & Information Priority System) which supports operational requirements of all public transport buses in Metropolitan NSW.	Apr 2019 - Jun 2019	\$357k	High	Exploration
Light Rail Priority	Provide priority to Light Rail at traffic intersections shared with other road users.	Feb 2019 - Mar 2020	\$1.49m	Medium	Exploitation
MPR	Ensure data management and architectural consistency of Operational Data Lake (ODL) across multiple performance reporting business cases.	Jan 2020 - Jun 2021	\$4.4m	Medium	Exploitation
Active Transport	Deliver Active Transport network & portfolio insights and data-driven decisioning capability	Jul 2021 - Dec 2022	\$3.25m	Medium	Exploitation

The chronology of DSIs outlines the progression from exploration to exploitation stages, reflecting Transport's evolving understanding of DSIs and the corresponding adaptations in delivery processes. While the data originates from the delivery team's assessments and stakeholder feedback, my contribution lies in the synthesis and interpretation of this information to delineate the organisational journey and its implications for managing DSIs effectively. When the first DSI (Vanguard) commenced delivery as a traditional ICT program, I faced several challenges in managing the schedule, which resulted in the planned milestones not being met. In hindsight, it became clear that the organisation was not aware of the exploratory nature of DSIs. However, as we progressed with Vanguard, we started realising the unique characteristics and made changes to the delivery processes. At a macro level, I mapped this initial stage to Transport's "exploration" stage. As we delivered more DSIs, the organisation started accepting the uniqueness of DSIs, the adaptations required for the delivery processes and the building skills needed to deliver DSIs successfully. The "exploitation" stage refers to a mature state where the organisation accepts that datasets come with uncertainty, that agile methods are practiced and improved upon and management accepts DSI business cases without any measurable benefits. In the context of

Transport, the exploration stage delivery timeline could be mapped to Vanguard, CTABS, Ferry and the Public Transport and Information Priority System (PTIPS) Analytics and the exploitation stage could be mapped to Light Rail Priority, MPR and Active Transport. *Figure 4-1* provides a chronology of the DSIs used as case studies. This includes the delivery timeline, the complexity of each project and the project stages. The last in particular indicates each team's journey from the uncertainty and frustration of not being able to deliver program outcomes as per the schedule to acceptance of the exploratory nature of DSIs, the need to plan for their uncertainty and the need to engage stakeholders effectively with that uncertainty in mind. A key insight from this was that as we transitioned from the Exploratory to the Exploitative stage, the exploratory nature of DSIs did not go away, it was just that the degree of uncertainty was reduced.

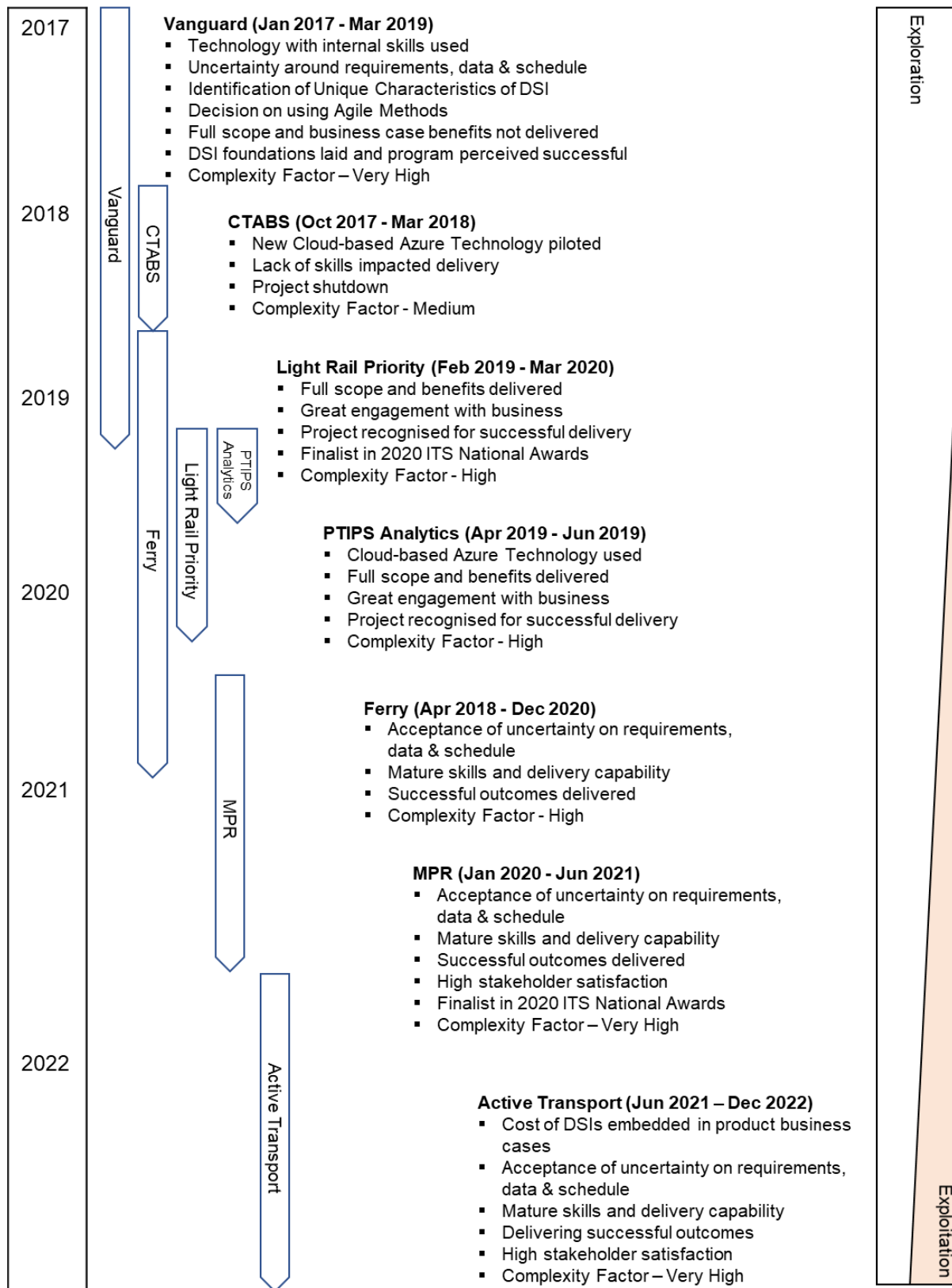


Figure 4-1. Transport DSIs Chronology

Table 4-2 to Table 4-8 provide insights into various aspects of DSI delivery, including benefits summary, stakeholder engagement, technology complexity, and risk assessment. While the data in these tables stems from the organisation and the delivery team's assessments, my role involved analysing this information, identifying patterns, and drawing conclusions pertinent to the research objectives. For instance, the assessment of stakeholder complexity and technology requirements offers valuable insights into the organisational dynamics surrounding DSI implementation, contributing to a nuanced understanding of the challenges and opportunities involved.

A summary of the benefits in each business case and benefits delivered along with the benefits complexity is captured in Table 4-2. The Benefits Complexity in column 4 was based on a qualitative assessment by the delivery team to collect and track data post-closure of the program.

Table 4-2
Benefits Summary of Transport DSIs

Program	Business Case Benefits	Benefits Delivered	Benefits Complexity
Vanguard	Increase revenue through improved fare compliance & improve customer satisfaction & security outcomes.	Partial delivery of benefits. Dashboards delivered to paint a picture of fare evasion & security by ingesting six of possible twenty-one data sources. Also, laid the foundation of data management & DSI delivery.	Very High
Ferry	Deliver five dashboards to monitor operator performance. Also, deliver Microsoft Azure-based Operational Data Lake (ODL) platform to current and future needs.	Partial delivery of benefits. On-boarded a new Operator on Transport systems and delivered performance reporting dashboards on Azure-based ODL. Issues with Ferry data could not be resolved due to external	High
CTABS	Obtain visibility of community transport services in NSW; understand the customers (who/how/why/where); understand the trips & travel patterns; assess service quality; investigate opportunities to improve service delivery; (vi) determine if CTABS has resulted in operational efficiencies; and (vii) assist in managing contracts	Project terminated as both solution and benefits could not be delivered.	Medium
PTIPS Analytics	Validate analytics solution using Azure Operational Data Lake; provide self-service capability to Bus Contract Managers & Operators with minimum six months of PTIPS data; and determine the operational expenditure (OPEX) requirements.	All benefits delivered including ten complex PowerBI dashboards with high stakeholder satisfaction.	High
Light Rail Priority	Support optimising Sydney Light Rail journey time; provide light rail, enhanced level 3 priority at intersections; increase visibility of Light Rail vehicles to TMC, RMS and SCATS; support decrease in Sydney congestion; and implement a hardware free solution for all SCATS intersections.	Partial benefits delivered. Technology solution delivered but some benefits were dependent on other systems and could not be directly attributed to this project. This was an enabler project.	Medium
MPR	Delivery of consistent ODL architecture and Bus (Metro), Bus (Regional), Ferry, Light Rail, Sydney Metro, Community Transport, OnDemand and Zero Emission Buses performance reporting.	Program consisting of ten projects delivered all benefits. Performance dashboards being used by Contract Management teams to identify and resolve operational issues.	High
Active Transport	* Delivery of real-time consistent walking & cycling volumes on infrastructure delivered. * Delivery of real-time voice of customer for qualitative data. * Delivery of Active Transport benefits by statistical areas, councils and by state. * Delivery of single source of truth of Active Transport portfolio. * Data-driven decisioning capability for future investment decisions.	Benefits delivery in-progress. Three proof-of-concept successfully delivered and program is deploying sensors across the state and operationalizing the metrics solution.	High

It was found that the delivery of DSIs required continuous engagement with various stakeholder groups, both formal and informal. Table 4-3 provides a summary of stakeholders along with the number of governance meetings that were used to inform the data for this research. The Stakeholder Complexity categories are based on qualitative assessment by the delivery team using the number of internal and external stakeholders, single vs multiple division involvement and the number of governance meetings.

Table 4-3
Stakeholder Summary of Transport DSIs

Program	No of Stakeholders ^(a)						Governance Meetings Participated ^(b)	Working Group Meetings Participated ^(c)	Stakeholder Complexity
	Total	Management	Governance	Support	Technical	Business			
Vanguard	67	7	15	5	32	8	36	702	Very High
Ferry	62	5	13	5	21	18	15	310	Medium
CTABS	17	7	0	5	5	0	5	35	Low
PTIPS Analytics	34	7	5	5	12	5	4	152	Low
Light Rail Priority	83	16	6	5	49	7	13	418	High
MPR	68	4	29	6	21	4	24	427	Very High
Active Transport	77	20	10	7	20	20	8	640	Very High

(a) Stakeholders included Program Steering Committee, Operational Systems Management, Business Users and Delivery & Support Team.

(b) Working Groups included Scrum Team, Architecture, Change, Data Management and Program/Project Working Groups.

(c) Governance Meetings included Portfolio & Program Board and Project Control Group Meetings.

Looking back at our first and seventh (final) DSIs, risks around the delivery processes reduced between Vanguard and Active Transport as the organisation and delivery teams better understood the “how-to” of DSIs and moved from the Exploration to the Exploitation stages. However, what did not change was the uncertainty around datasets due to the number of datasets being ingested, their source, their inherent quality and the transformation complexity involved. This resulted in very high data complexity, as shown in Table 4-4. The uncertainty was lower when the datasets were internal (i.e., sourced from internal systems) or known (i.e., in the public domain) but in majority of DSIs, the datasets were external (i.e., sourced from external organisations/systems), and the team did not know about them until we started ingesting them. The Data Complexity category was based on the number of datasets, their source and their transformation complexity.

Table 4-4
Source Data Summary of Transport DSIs

Program	No of Data Sources	Count by Type			Count by Source		Count by Transformation Complexity ^(a)				Data Complexity ^(b)
		Reference	Master	Transaction	Internal	External	Very High	High	Medium	Low	
Vanguard	14	8	0	6	8	6	3	3	0	8	Very High
Ferry	10	5	1	4	6	4	0	3	2	5	Low
CTABS	16	2	6	8	0	16	0	0	16	0	Medium
PTIPS Analytics	7	5	1	1	5	2	1	0	1	5	Very High
Light Rail Priority	7	5	1	1	5	2	1	0	1	5	High
MPR	20	5	3	12	8	12	3	6	6	5	Very High
Active Transport	18	1	1	16	4	14	2	8	4	4	Very High

(a) Transformation Complexity was determined by application of business rules.

(b) Data Complexity was based on the number of data sources, the source and the transformation complexity.

Similar technology was used across the DSIs except for Vanguard, when cloud-based solutions were emerging and had not been implemented within Transport. The Technology Complexity was based on qualitative assessments by the delivery team using a number of (new) technology components and whether a solution architecture existed.

Table 4-5
Technology Summary of Transport DSIs

Program	Hardware	Software	Database	Solution Architecture	Technology Complexity
Vanguard	Private Cloud - NSW Gov Data Centre	Tableau, SAP Business Objects, TIBCO	Oracle	Architecture developed & productionised for first time	Very High
Ferry	Public Cloud - Microsoft Azure	PowerBI, DAX, Microsoft Azure	Microsoft SQL Server	Added functionality to existing architecture	Medium
CTABS	Public Cloud - Microsoft Azure	PowerBI, Microsoft Azure	Microsoft SQL Server	Analytics solution piloted on new architecture	Medium
PTIPS Analytics	Public Cloud - Microsoft Azure	PowerBI, DAX, Microsoft Azure	Microsoft SQL Server	Analytics solution productionised on new architecture for first time	High
Light Rail Priority	Public Cloud - Amazon Web Services	PowerBI, DAX, Microsoft Azure	Microsoft SQL Server	Added functionality to existing architecture	Medium
MPR	Public Cloud - Microsoft Azure	PowerBI, DAX, Microsoft Azure	Microsoft SQL Server	Added functionality to existing architecture	Medium
Active Transport	Public Cloud - Microsoft Azure	PowerBI, DAX, Microsoft Azure, Salesforce CRM, Salesforce Interaction Studio, Salesforce Grants	Microsoft SQL Server	Added functionality to existing architecture	Very High

A review of the risks, dependencies and constraints of the DSIs in Table 4-6 shows that the majority of the ratings fell into the High and Very High categories, thus contributing to higher complexity and uncertainty in delivery. These ratings were allocated according to the Transport Enterprise Risk Management (TERM) framework and were based on a monthly review by the delivery team of risks, dependencies and constraints.

Table 4-6
Risks, Dependencies and Constraints of Transport DSIs

Program	Risks	Dependencies	Constraints	Complexity
Vanguard	Very High	Very High	High	Very High
Ferry	Medium	Medium	Medium	Medium
CTABS	Medium	Medium	Medium	Medium
PTIPS Analytics	High	Medium	High	Medium
Light Rail Priority	High	High	High	High
MPR	Very High	Very High	High	Very High
Active Transport	Very High	Very High	High	Very High

Our delivery of seven DSIs showed that the skills specific to each of the program management, change management, scaled agile, data management and data science domains were required, as shown in Table 4-7. Depending upon the type of DSI, the skill level and amount of time required for each of the role varied, but they were all essential to successful delivery. It can be argued that some of the roles in Data Management and Data Science domain are also required in other ICT initiatives. As highlighted in Section 2.5, this research is at frontier of portfolio and program management and the expectation is that over time the gap between DSI and ICT initiatives will reduce and more of these roles will become essential in delivering ICT initiatives.

Table 4-7
Roles and Responsibilities across DSI Domains

Domain	Role	Responsibility	DSI Essential	ICT Essential
Program Management	Business Owner	Have the primary business and technical responsibility for governance, compliance, and return on investment (ROI) for a	Yes	Yes
Program Management	Product Manager	Owns Vision and Roadmap and defines features and releases.	Yes	Yes
Program Management	Program Manager	Responsible for overseeing the achievement of larger organizational goals by coordinating efforts between different projects	Yes	Yes
Change Management	Customer	Buyers of a solution. Internal customers are part of the enterprise whereas external customers are outside the enterprise and can be Business-to-Business (B2B), Business-to-Professional (B2P) or Business-to-Consumer (B2C).	Yes	Yes
Change Management	Change Manager	Lead change management and adoption of product and processes.	Yes	Yes
Change Management	Change Analyst	Support Change Manager by deep-diving into stakeholder identification, impact assessment and taking them through the	Yes	Yes
Scaled Agile	Domain Architect	A specialist with deep knowledge within a particular domain of their expertise. A domain could be 'Data Services,' 'Process Design,' 'Integration Services', 'Domain Expert for SAP', etc.	Yes	Yes
Scaled Agile	Product Owner	Responsible for defining and prioritising stories to streamline the execution of program priorities while maintaining the conceptual and technical integrity of the Features or components.	Yes	Yes
Scaled Agile	Scrum Master	Servant leaders and coaches for an agile team who help remove impediments and foster an environment for high-performing team dynamics, continuous flow, and relentless improvement.	Yes	Yes
Scaled Agile	Business Analyst	Guide businesses in improving processes, products, services and software through data analysis. They act as an interface between IT and the business to help bridge the gap and improve efficiency.	Yes	Yes
Scaled Agile	Tester	Responsible for the quality of software development and deployment and performing automated and manual tests to ensure	Yes	Yes
Scaled Agile	Cloud Engineer	Responsible for duties associated with cloud computing, including design, planning, management, maintenance and support. Can encompass a few different roles such as cloud architect, cloud software engineer, etc.	Yes	No
Scaled Agile	Security Engineer	Identify threats and vulnerabilities in systems and software, develop and implement solutions to defend against hacking, malware and ransomware, insider threats and all types of cybercrime.	Yes	No
Data Management	Data Owner	Has the authority and accountability for the information assets. Decides who has the right to access and edit data and how it's used and be responsible for overseeing and protecting a data domain.	Yes	No
Data Management	Data Custodian	Responsible for the safe custody, transport, storage of the data and implementation of business rules and is generally a technical role. Can also be called Database Administrator (DBA), Data Modeller, ETL Developer. Acts as the proxy for the Data Owner where Data Owner is external.	Yes	No
Data Management	Data Steward	Responsible for data content, context, and associated business rules and is generally a business role.	Yes	No
Data Science	Information Architect	Implements information structure, features, functionality, UI and focuses on structural design and implementation of an	Yes	No
Data Science	Data Architect	Responsible for data architecture and data integration and may work at the enterprise level or functional level. Work on the structural design of an infrastructure specific to collecting data, pulling it through a lifecycle and pushing it into other meaningful systems.	Yes	No
Data Science	Data Scientist	Analytical data experts who have the technical skills to solve complex problems and curiosity to explore what problems need to be solved. They're part mathematician, part computer scientist and part trend-spotter straddling both the business and IT worlds.	Yes	No
Data Science	Machine Learning (ML) Engineer	Focuses on researching, building and designing self-running artificial intelligence (AI) systems to automate predictive models and deploy into production environment.	Yes	No
Data Science	Data Engineer	Works with multiple databases to capture and process live, streaming, and distributed data. Designs and develops data collection, management, and search-and-retrieval systems in order to support the collection, processing, exploitation, analysis and dissemination of complex datasets.	Yes	No
Data Science	Data Analyst	Responsible for providing descriptive statistics, probability models, and other quantitative assessments of raw, processed, and generated data. Employs a combination of traditional statistical and machine learning/artificial intelligence techniques to analyse complex data sets in support of analytic, collection, and managerial activities.	Yes	No
Data Science	Data Communications Specialist	Responsible for communicating and presenting summaries of structured and unstructured data in visual, text-based and interactive formats. Requires strong technical knowledge for implementing data visualizations using technologies such as PowerBI, Tableau, etc.	Yes	No

Comparing the skills requirement of the various DSIs (as shown in Table 4-7) with the availability of skills at the time, a qualitative assessment of the skills gap (as listed in Table 4-8) shows that CTABS had greatest gap. This led to the project being terminated.

Table 4-8
Skills Summary of Transport DSIs

Program	Skills Required	Skills Available	Skills Gap
Vanguard	Very High	High	Medium
Ferry	High	High	Low
CTABS	Medium	Low	High
PTIPS Analytics	High	High	Low
Light Rail Priority	High	High	Low
MPR	Very High	Very High	Low
Active Transport	Very High	Very High	Low

Table 4-9 and *Figure 4-2* present an analysis of DSI delivery complexity across multiple lenses, highlighting the interplay between various factors such as skills gap, organisational maturity and benefits realisation. While the data originates from assessments by the delivery team, my contribution lies in synthesising these findings, identifying overarching trends, and drawing implications for DSI management and governance. Specifically, the qualitative analysis of the spider chart underscores the pivotal role of skills development in determining DSI outcomes, offering valuable insights for organisational capacity-building initiatives.

Table 4-9
DSIs Delivery Complexity

Program	Stakeholder Complexity	Technology Complexity	Data Complexity	Skills Gap	Budget Size	Benefits Complexity	Requirements Uncertainty	Risks, Dependencies & Constraints
Vanguard	Very High	Very High	Very High	Medium	Large	Very High	Very High	Very High
Ferry	Medium	Medium	Low	Low	Large	High	Low	Medium
CTABS	Low	Medium	Medium	High	Small	Medium	Low	Medium
PTIPS Analytics	Low	High	Very High	Low	Small	High	Medium	Medium
Light Rail Priority	High	Medium	High	Low	Medium	Medium	Low	High
MPR	Very High	Medium	Very High	Low	Large	High	Low	Very High
Active Transport	Very High	Very High	Very High	Low	Medium	High	High	Very High

A qualitative analysis of the spider chart by the delivery team using values from Table 4-9 revealed that, regardless of the complexity of DSIs, it was the skills gap that determined the outcomes delivered by the DSIs. For example, Vanguard had a medium skills gap and did not deliver all the business case benefits and CTABS, with its high skills gap could deliver no business case benefits and had to be terminated.

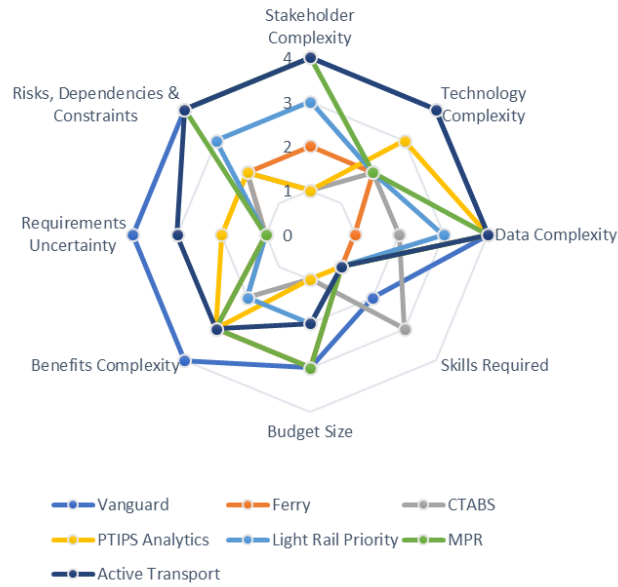


Figure 4-2. DSI Delivery Complexity

4.2 CONCLUSION

In this section, I reviewed the data for seven Transport DSIs with a focus on identifying patterns and trends relevant to the research objectives. While the data presented originates from the organisation and the delivery team's assessments, my role involved synthesising and interpreting this information to elucidate key insights and implications for managing DSIs effectively. By contextualising the findings within the broader research framework, this chapter contributes to a comprehensive understanding of the challenges and opportunities associated with DSI implementation in organisational settings. In my analysis, the skills gap was an indication of the maturity of both the organisation and the delivery team. It also reflected the organisational journey from exploration to exploitation of DSIs. The higher the skills gap, lower is the maturity and lower are the chances of benefits being delivered by the DSI.

5. Unique Characteristics of DSIs

This chapter is based on my published paper “Unique characteristics of Data Science Initiatives: Implications for Program Management” (Mathur et al., 2024).

5.1 ABSTRACT

Data is increasingly ubiquitous in organisational life with many investments using data at the ideation stage to inform a business case through to delivery of the core scope elements to benefits realisation. While DSIs have emerged as a popular mechanism for extracting value from data, their track record has drawn substantial criticism from sponsors. For example, the success rate of delivering DSIs is not perceived as high, with Gartner estimating that 85% of projects fail (Asay, 2017). This chapter argues that one crucial reason for this failure is that DSIs have unique characteristics that make traditional practices for conceptualising and managing ICT-enabled programs ineffective. I build this argument by drawing on DSI case studies delivered over five years at the statutory body responsible for transport in the state of New South Wales in Australia, Transport for NSW (Transport). I conclude by explaining how managing DSIs as “exploratory projects” may improve the success rate of implementations.

5.2 INTRODUCTION

DSIs⁷ have unique challenges that make the application of traditional program management techniques problematic. These challenges arise primarily due to the uncertainties inherent in the data ingested, which has a cascading impact on the scope, the schedule and, ultimately, the value creation.

⁷ Data science initiatives (DSIs) are defined as related projects or programs that involve the application of data science techniques and methods to address complex problems, generate insights or create new products or services.

In all DSIs, data from a range of sources, known and unknown, are ingested into a data store and transformed to generate insights. At the commencement of any DSI, the quality and structure of the data being ingested is relatively unknown. This suggests that there are occasions when DSIs need to be managed as exploratory projects due to limited “information-before-action”. Lenfle (2008) describes exploratory projects as those for which neither technologies nor customer requirements are known at the start of the project. This uncertainty makes it difficult to manage DSIs as Exploitative Projects, which focus on optimising the triple constraints of cost-quality-time to deliver new products and services (Lenfle, 2008). The fundamental tension between the exploitation of old certainties and the exploration of new possibilities identified by March (1991) is particularly relevant to DSIs.

Furthermore, the sequential and predefined waterfall approaches to program management adopted by professional project and program management bodies to deliver DSIs set up structural tensions particularly in business case development, program design, delivery and benefits realisations to undermine coherent governance across the investment life cycle. Whereas a known scope-and-benefits profile and old certainties allow an exploitative project to move sequentially and seamlessly from one gate to another, the uncertainties in DSIs hamper the sequential software development approach of analysis, design, development, testing and deployment. This makes DSIs more aligned to iterative, innovative and exploratory projects.

In this chapter, I identify unique characteristics of DSIs that distinguish them from typical ICT-enabled programs to help scholars and practitioners better understand when and why waterfall approaches are likely to fail and what alternative might enable them to deliver more successful business outcomes. I therefore draw on evidence from in-depth case studies of six DSIs delivered over five years at Transport. The external validity of these findings was probed using semi-structured interviews with practitioners from diverse organisations involved in the delivery of the DSIs. I conclude by arguing that practitioners need to understand the exploratory characteristics of DSIs when planning and delivering them and move away from traditional approaches that fail to accommodate the uncertainty and ambiguity that currently shape DSI delivery.

5.3 LITERATURE REVIEW

The focus of this research was to improve DSI delivery practices, with the target audience being program managers delivering ICT initiatives, portfolio managers and policy makers approving business cases and establishing governance mechanisms and academics teaching program management. Some of the key issues motivating this research were:

- Why 85% of big-data projects fail (Asay, 2017) when 73% projects (overall) meet original goals (Project Management Institute, 2021a),
- What makes DSIs unique compared to other ICT initiatives,
- How uncertainty in scope, schedule and benefits is managed in ICT projects, and
- How DSIs can be delivered effectively.

Significant literature exists about various domains of ICT initiatives delivery. However, a gap exists in relation to DSIs, especially the lack of acknowledgement of their uncertainty, complexity and exploratory nature, which has a negative impact on DSI management and governance.

5.3.1 Exploratory and Exploitative Projects

Exploratory projects can be characterised as projects for which, from the outset, neither the goals nor the means of attaining them are clearly defined (Lenfle, 2008). While Turner and Cochrane (1993) did not use the term “exploratory” per se, the Type-4 projects in their goals and methods matrix are those where neither the goals nor the methods are well defined and point to the exploratory nature of certain types of projects. Delivery of such projects cannot be effected using standard project management methodology, which largely focuses on delivery of a defined scope, cost and schedule. The inclusion of Agile Practice Guide in PMBOK Guide – Sixth Edition (Project Management Institute, 2021b) in our interpretation is an acknowledgment by Project Management Institute that waterfall methods do not allow effective risk management of ICT-enabled projects thus causing high failure in delivery. As a result, Agile adoption by software development teams has increased significantly from 37% in 2020 to 86% in 2021 (digital.ai, 2021). CH Loch et al. (2011) highlight concerns using standard risk management methods for novel projects with unforeseeable uncertainties, complexities and unknown unknowns. They expand

further by saying that a project plan for novel projects is nothing but a starting point, an illusion or a simple sketch and consequently a conventional project risk management approach cannot be followed. The governance of exploratory and innovative projects is another area requiring consideration, as supervising managers tend to tighten rather than loosen control when projects become more uncertain (Christoph Loch & Sommer, 2019). Christoph Loch and Sommer (2019) discuss the need for flexible execution methods for exploratory projects by managers and offer proposals on how to make exploratory projects more acceptable to top management attempting to resolve a very fundamental managerial dilemma.

Exploitative or development projects, on the other hand, deliver a unique product or service and thus benefits are relatively more easily understood by business managers, especially those applying a financial lens to the investment.

5.3.2 Data Science Initiatives

My case studies showed the need for domains such as program management, change management, data management, data science and development-operations (DevOps) to deliver DSIs. I start with definitions of big data, business intelligence and data science to introduce them as foundational elements of DSIs, with the purpose of identifying gaps in both the literature and in practice and leading up to the research question later. The term “big data” has become ubiquitous and several authors have attempted to standardise its definition (De Mauro et al., 2016; Ward & Barker, 2013). “Business intelligence” is an umbrella term which includes applications, tools, infrastructure and practices to enable access and analysis of information to optimise performance and decision-making (Gartner, 2023; Halper, 2015; Larson & Chang, 2016). “Data science” can be defined as the collection, preparation, analysis, visualisation, management and preservation of large collections of information to generate actionable insights (Donoho, 2017; Saltz & Stanton, 2017). DSIs often also include implementation of AI and ML.

For the DSI case studies in this research, it was never about bringing in data from one source and closing the project; rather, I saw the DSIs being typically implemented as a program on a continuous spectrum rather than a single one-off project.

Data management is the development, execution and supervision of plans, programs and practices that deliver, control, protect and enhance the value of data and information assets

throughout their lifecycles (Earley, 2017). I see 11 knowledge areas of data management (Earley, 2017) as foundational elements for a DSI delivery team, unlike other ICT-enabled programs.

Viaene and Van den Bunder (2011) have attempted to define the skills that differentiate successful business analytics project managers from traditional project managers. They recommend more adaptive experimentation-based approaches, emphasising that less specific early planning, good-enough requirements and experimental and evolutionary design with significant ongoing learning and change are better suited to analytics projects. Through a survey of 237 practitioners, Franková et al. (2016) found that agile and iterative delivery was more appropriate to big data projects. In summary, growing research points to using Agile methods to deliver DSIs.

A common requirement for large projects with globally distributed teams needing to collaborate and coordinate has led to the popularity of scaled agile frameworks such as Scaled Agile Framework (SAFe), Large-Scale Scrum (LeSS) and LeanSAFE. I see the relevance of scaling is similarly high in the context of DSIs, as often multiple geographically-spread teams within an organisation are involved in delivering data science outcomes.

The delivery of DSIs requires practitioners and management to have a good understanding of data mining and data science delivery processes. Some of the widely-used frameworks in the delivery of DSIs are the KDD model (Fayyad et al., 1996), the Cross-Industry Standard Process for Data Mining (CRISP-DM), the SEMMA process (SAS Institute, 2009), the Obtain, Scrub, Explore, Model and Interpret (OSEMN) process (Mason & Wiggins, 2010), the TDSP (Severtson et al., 2017) and the Foundational Methodology for Data Science (FMDS) (Rollins, 2015). At Transport, CRISP-DM was and continues to be used by teams delivering DSIs.

I thus argue that DSIs are typically exploratory projects due to the uncertainty around their benefits and the delivery itself. Other than the setup of the foundational infrastructure, the delivery work packages of DSIs cannot be clearly defined at the outset and thus the schedule cannot be prepared in detail. DSIs cannot conform to the rational waterfall approach of projects in terms of delivering a unique product, service or result within a specified period, defined budget and quality requirements. While iterative delivery has been proposed for software projects for some time (Mathur, 2005), DSIs tend to be delivered in a progressive elaboration using Agile methods.

5.3.3 Summary and Implications

A review of the related literature shows that a gap exists in how DSIs are defined in portfolios and how they are executed as programs. This research addressed that gap through the following research question:

“What unique characteristics cause DSIs to face challenges delivering envisaged value when using traditional processes for managing ICT-enabled programs?”

Program management for ICT-enabled programs has a rich literature and proven delivery frameworks, which have matured since 1990 (Axelos, 2022; Project Management Institute, 2016, 2017b, 2019, 2021b). However, the failure rate in the delivery of DSIs points to a gap in the literature to address the exploratory and innovative nature of DSIs. This research delivers a significant contribution to the body of knowledge for program management relevant to both researchers and practitioners of the emerging data science domain. Without the proposed body of work, there will be more failed programs, more dissatisfied sponsors and delay in much-needed investment in this emerging domain, as well as delay in the benefits that will flow from harnessing the data and the nuggets in it.

5.4 RESEARCH SETTING AND METHODS

Applying a practice lens to the delivery of DSIs guided me to focus on the full life cycle of DSIs. Such a focus required deep engagement in the field, observing and interacting with decision-makers, business stakeholders, program managers and delivery team members. As a result, I chose to study the delivery of DSIs within a single organisation (Transport) where I am employed full-time and continue to deliver DSIs. This allowed me good access to data to conduct the case studies.

5.4.1 Research Methods

A multi-methods approach (Hunter & Brewer, 2015; Straits & Singleton, 2018) combining case studies and semi-structured interviews with practitioners was used. To obtain granularity of program life cycle as well as variation for analytical comparisons, an embedded case design was selected (Yin, 2018) to analyse six DSIs in Transport, each of which provided a unique scope and opportunity to understand characteristics of DSIs. My interest was to understand the characteristics of DSIs as experienced by the organisation’s participants themselves and identify the unique improvements that this class of initiatives could bring to the organisation. The six DSIs chosen as

case studies reflect the chronology and the maturity of Transport in delivering DSIs (*Figure 4-1*). I saw specific patterns emerging as we progressed through DSIs and by the time we reached the sixth, I saw stability in patterns and consistency in characteristics. I continue to deliver more DSIs, further validating the findings, but for the purposes of this chapter, I stopped at the sixth DSI. The number of cases selected followed the recommendations made by Small (2009) and Yin (2018) in setting up a multiple case study design, with the extreme cases expected to contribute to theoretical replication (predicting contrasting results) and the semi-structured interviews with practitioners providing literal replication (finding similarities).

Using an interpretive research tradition associated with case studies, ontological and epistemological assumptions on DSI characteristics emerged; these were externally validated with practitioners from five other organisations delivering DSIs by using semi-structured interviews. Informed consent was sought from interviewees by carefully explaining the study and its aims, as well as explaining their ethical rights during interviews. Ethics approval was granted by UTS prior to commencing the research (Application ID ETH20-5468, dated 28th January 2021). The interviews used open-ended questions to gain the lived experienced of interviewees. An interpretive approach (Sandberg, 2005) to justify the knowledge produced was adopted by analysing interview transcripts, leading to coherent interpretations of DSI characteristics. Each of the characteristics was either confirmed or altered according to the responses to the questions. A new characteristic emerged after the second interview and was further validated by the remaining interviewees. Follow-up questions were asked, especially when interviewees described the DSIs they had delivered and the associated ambiguity and complexity of some of the characteristics.

Iterating the in-depth analysis of each case, comparisons across cases and connections to the literature (Eisenhardt, 1989), I reviewed diverse stakeholder groups as well as formal and informal interactions (as shown in Table 4-3) and how they influenced the delivery of DSIs, which in turn led to further analysis and theorising (Agar, 1986). For the majority of our stakeholders in Transport, this was their first experience working on or interacting with a DSI. It was a steep learning curve for them and our interactions with them not only educated but also informed us of the nuances of DSIs. I also used supplementary evidence, such as documents and participant observations from each DSI. The evidence included program management artefacts, as listed in Table 5-1. The observations took place throughout the delivery of six DSIs, especially in the Agile

Ceremonies listed in Table 5-1, which were a good reflection of what the team felt at the end of each two-weekly sprint.

Table 5-1
Program Management Artefacts and Ceremonies

Program Management Documents
Program Management Plan
Risk Register
Schedule
Communication
Minutes of Governance and Working Group Meetings
Stories and Burndown Charts in Jira and Confluence
Agile Ceremonies
Daily Standup
Iteration Planning
PI Planning
Backlog Grooming
Iteration retrospectives
PI retrospectives

The research question emerged over time by reviewing the challenges faced by the delivery teams backed by evidence from the case studies.

I was the Program Manager of the six DSIs chosen as case studies which were delivered between January 2017 and December 2020 and could thus bring in-depth insights about the program life cycle.

The participants for the semi-structured interviews represented five Australian-based organisations that delivered DSIs in Australia, New Zealand and the USA. While my sample size was small, the five interviews helped me to validate the trustworthiness and reliability of the DSI characteristics by probing whether geographical or organisational boundaries might induce variation in the findings.

5.4.2 Research Setting

My research was situated within the Operational Systems division of Transport for NSW (Transport), a state government enterprise that leads the development of safe, integrated and efficient transport systems for the people of New South Wales in Australia. Transport's functions

include transport planning, strategy, policy, procurement and other non-service delivery functions across all modes of transport – roads, rail, ferries, light rail, metro and point-to-point. At the time of my research, the organisation was in the early stages of executing AU\$72.1 billion worth of Future Transport 2056 Services and Infrastructure Plans (SIPs). These included more than 300 initiatives to be delivered in the first 10 years of the 40-year vision, underpinned by a data-driven technology roadmap (Transport for NSW, 2021). Most initiatives had a significant technology component, including data and data analytics.

Transport generates significant amount of data from bus, ferry, light rail, metro, heavy rail and active transport (walking and cycling) modes every day. The real-time data includes timetable information, the position of every transport vehicle and the predicted time of arrival at the next transit stop, all of which is shared on all passenger apps. It also includes event information (such as door opening and closing information) from sensors, temperature, speed and number of passengers; ticketing information such as Opal cards (smartcard tickets) being tapped on and off at the gates and checking of tickets; monitoring the use of cycleways and pedestrian walkways, and to crime and incident information throughout the transport network.

While Transport was always aware that the data could be used to improve customer service in the multi-modal journeys and to manage the performance of the transport operators, the lack of technology, skills and investment prevented the organisation from mining and capturing the value from the data it was generating. For example, Transport continued to rely on operators to advise management through their monthly reports if the services they delivered met contractual performance requirements. The awareness of this shortfall led to the Future Transport Strategy 2056 being released in 2018, including a data ecosystem within Transport to provide continuous improvements on its asset performance and improved customer and operational information. My journey mirrors that of Transport as an organisation. When I joined Transport in late 2016 as a program manager, I had significant experience in delivering ICT transformation programs using both Waterfall and Agile methods but none in delivering data-related programs. This chapter tracks the delivery of the DSIs I managed, the challenges I faced in understanding some of their unique characteristics, my ability to influence near real-time data-driven decision-making in Transport and the growth in the ability of Transport to harness the value of data, underpinned by my own ability to deliver DSIs.

In summary, the research method uses a participant-observation technique in multiple case studies over their full program life cycle and covering a period of five years collecting DSI data. The research used documentation, archival records, direct observations, participant observation and physical artefacts as sources of data.

5.5 DATA COLLECTION AND ANALYSIS

The six DSIs chosen as case studies (the first six DSIs listed in Table 4-1) represent contemporary phenomena in-depth and within their real-world context (Yin, 2018) at Transport that was particularly useful for the research question because the organisation needs to better understand the unique characteristics of DSIs.

The chronology of the six DSIs was bracketed into three stages that Transport went through as the six DSIs were delivered, namely: Exploration, Transition and Exploitation. When I commenced my first DSI (Vanguard) as a traditional ICT program, I faced challenges managing the schedule and the planned milestones were not met. In hindsight, the organisation was not aware of the exploratory nature of DSIs. Initially, the key stakeholders of the Vanguard program had the expectation that I would be able to document the scope, time and costs as part of a program management plan and the delivery team would deliver the program as per the schedule. That initial expectation could not be met. However, as we continued with Vanguard, we started acknowledging the unique characteristics and making changes to the delivery processes. At macro level, I mapped this initial stage to Transport's "exploration" stage. The "transition" stage was mapped to the organisation accepting the uniqueness of DSIs and adapting the delivery processes and building skills to deliver the DSI successfully. The "exploitation" stage refers to a mature state where the organisation accepts that datasets come with uncertainty, where agile methods are practiced and management accepts that DSI business cases may have no measurable benefits.

In the context of Transport, the DSI mapped the "exploration" stage to Vanguard and CTABS, the "transition" stage to Ferry, Light Rail Priority and PTIPS Analytics and the "exploitation" stage to MPR. *Figure 4-1* shows the timeline, complexity, project stages and highlights of the six DSIs, indicating the researcher's journey from uncertainty and frustration at not being able to deliver program outcomes as per the schedule to accepting the exploratory nature of DSIs, planning for uncertainty and engaging stakeholders effectively. While each of the six

DSIs supported different parts of Transport with different requirements, this chapter focuses on the first (Vanguard) and the sixth (MPR), as they represented boundary conditions of the story presented here. In other words, I present details of the initial “exploration” stage and close with that of the final “exploitation” stage.

The delivery of DSIs required continuous engagement with various stakeholder groups through formal and informal interactions. Table 4-3 provides a summary of stakeholders and with governance meetings that were used to inform the data for this chapter.

The Transport case studies, as shown in Table 4-1 and Figure 4-1, demonstrated our inability to baseline the scope and schedule. As shown in *Figure 2-19*, each DSI went through the steps of business understanding, data understanding, data preparation, modelling, evaluation and deployment, using one or more datasets. Vanguard, which was the first DSI, ingested only six out of a possible 21 datasets. The ingestion of two key datasets, ORPA (tickets scanned) and ePIN (fines issued), was dependent on a third-party vendor and it took more than 18 months for the contracts to be signed and scheduled and the datasets to be delivered. Even when the initial data arrived, the team found integrity issues in it. For example, there was fines data without the mandatory tickets scanned data which required going back to the vendor to remediate, continuously impacting the delivery schedule. As I progressed the delivery of our first DSI Vanguard to provision for uncertainty, not only did we adopt agile ways of working but created a high-level flexible schedule which acknowledged the complexity of the data. The shift in our approach to the scheduling was evident in the way the waterfall-agile hybrid timeline in the Vanguard business case (Figure 5-1) changed to a data-centric timeline in January 2018 (Figure 5-2). Figure 5-1 shows two sequential “Write Agile Stories” activities, when the requirements were captured, and the “Build + System Testing Iterations”, when the system was built, the overall schedule taking 16 months to deliver. In reality, Vanguard took 27 months to deliver (Figure 4-1) and Figure 5-2 shows a portion of the overall schedule (January 2018 – October 2018), reflecting how the schedule was broken down by various data sets (ORPA, STA, SRS, ePIN, FCS) to show the data-centric approach.

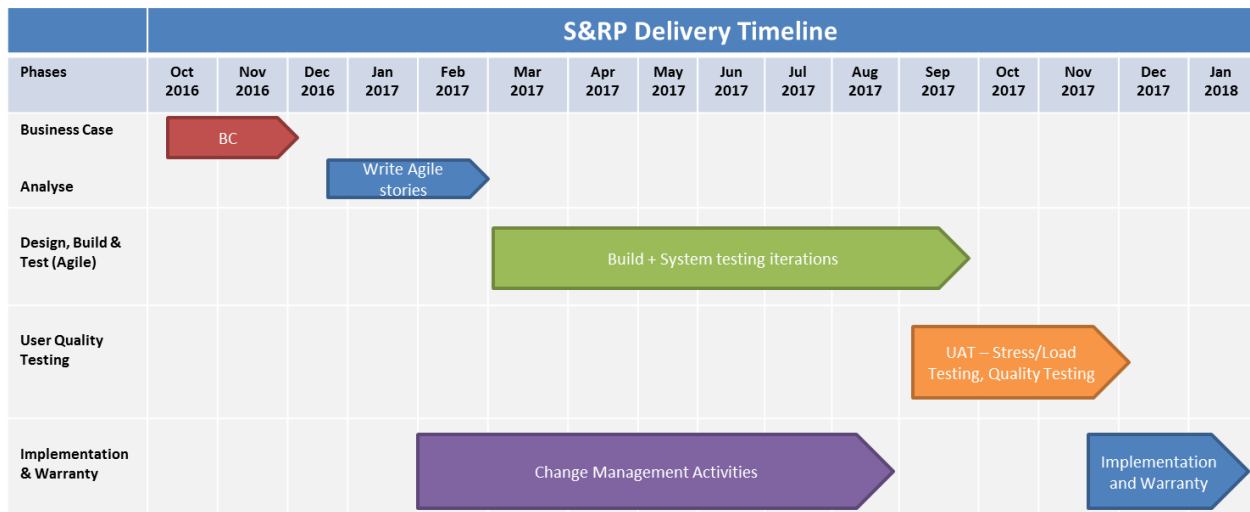
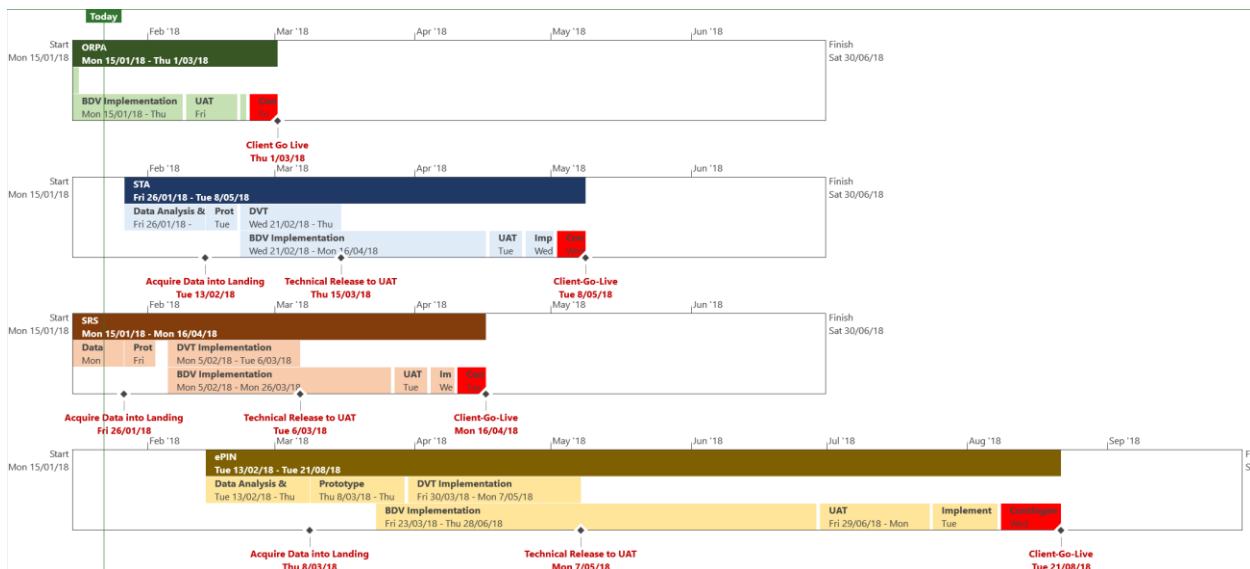


Figure 5-1. Vanguard Delivery Timeline in the Business Case (Oct 2016)



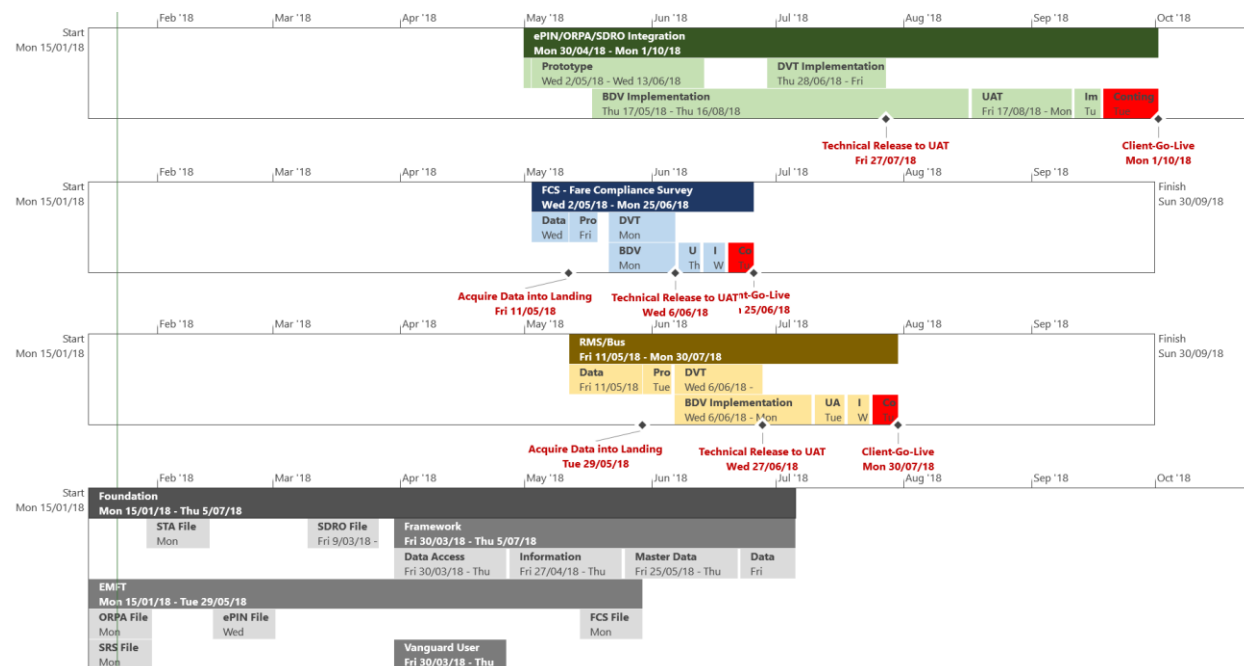


Figure 5-2. Vanguard Delivery Timeline using a data-centric approach (Jan 2018)

The uncertainty was further highlighted by comparing the planned vs delivered dates of various datasets, some of which were almost a year late based on the two-and-a-half year Vanguard program shown in Table 5-2. Specifically, the ORPA dataset was delivered six months late and ePIN more than a year late. This had a cascading impact on fines collection data from Revenue NSW, which was delayed by a further seven months.

Table 5-2
Vanguard – Planned vs Actual Delivery of datasets

Data-Set	Description	Planned	Delivered	Variance (Days)
BOCSAR	Crime data on Transport network	1/05/2018	17/11/2017	-165
STA Incidents	Incident data from State Transit Authority	16/10/2018	18/07/2018	-90
SRS Incidents	Incident data from State Rail Services	16/10/2018	27/03/2019	162
ORPA	Tickets checked data from Opal Revenue Protection App	31/10/2017	18/05/2018	199
ePIN	Electronic penalty infringement notice data	31/10/2017	8/11/2018	373
RNSW	Fines collection data from Revenue NSW	6/02/2018	26/09/2018	232

Our sixth DSI, the Multi-Modal Performance Reporting (MPR) program, commenced in January 2020 to “ensure data management and architectural consistency of Operational Data Lake across multiple performance reporting business cases”. The initial scope included bus (metro), ferry, light rail and bus (regional) performance reporting and the scope was extended in June 2020 to include Sydney Metro performance reporting as well as data ingestion and self-service projects.

Our organisation and teams had transitioned to the exploitation stage by the time the MPR delivery took place. Learnings from the previous DSIs were applied to identify the core roles required for successful DSI delivery. These included cloud architect, information architect, data architect, scrum master and change manager roles at the program level to bring in governance, architectural and change management consistency and each project had product owner, business analyst, PowerBI and ETL developers and tester roles to bring in delivery consistency.

The MPR program had six projects being executed in parallel and was the most mature DSI delivered at Transport. While uncertainty around data and scheduling were still there, the team could manage risks better. Daily 15-minute stand-up meetings ensured that any blockers were addressed promptly. Fortnightly Agile Ceremonies such as retrospectives, Planning and Reviews took place without fail. Regular backlog grooming⁸ allowed us to plan the next sprint. Monthly program risk review meetings were scheduled with the full team to update the RAID register. The communication with the stakeholders worked well, and frequent showcases with product owners enabled them to provide feedback. All the lessons from earlier implementations were applied to the MPR and the program was able to retain key resources from the other five DSIs. The schedule was managed using quarterly iterations and Harvey balls to show the progress of each deliverable, as shown in Figure 5-3. I no longer did detailed scheduling at individual dashboard and dataset level; instead, I identified user stories for each sprint till the dashboard and associated datasets went through user acceptance testing and were deployed in the production environment.

⁸ Backlog grooming is an agile event where backlog items are discussed, reviewed and prioritised by product managers, product owners and the rest of the team.

		Iteration 1 (Jan – Mar)	Iteration 2 (Apr – Jun)	Iteration 3 (Jul – Sep)	Iteration 4 (Oct – Dec)
MPR Program		Technical Assurance Plan Approval by OTASC	- MPR Solution Architecture (Initial) Approval by TDA & OTASC - System Classification - Cloud Risk Assessment - Business Impact Assessment	- Masterdata HUB Solution - Service Design (Application Layer) - Information Security Management Plan	- Service Design (Platform) & Transition - MPR Solution Architecture (Final) Approval - Penetration Testing & Cyber Security Report - Business Glossary
	Ferry	- Patronage / Opal data Analysis - Data Modelling for GTFS feed	On-Time Running Report	- Service Availability Dashboard - Service Interruption Dashboard - Service Investigator Dashboard - Unscheduled Dashboard	- Executive Dashboards - Customer Feedback (Salesforce) - Customer Satisfaction Index - Customer Complaints - Patronage Dashboard - Training Videos - Productionise Dashboards
	Light Rail	- Journey Time & Frequency Report - Service Availability Report	- Service Availability Report - Headway Performance - Service Adjustments - Service Interruption	- Productionise four Dashboards - Executive Dashboards (incl. 9:00am Minister Report) - Traffic Signal Delay (TSD) Dashboard	- Training Videos - Productionise Dashboards
	TCB	Training for PTIPS Analytics Reports	On-Time Running Dashboard	- Productionise OTR & Passenger Load Dashboard - Passenger Load Dashboard	- Training Videos - Passenger Load Dashboard - Cancelled & Incomplete Trips
	Sydney Metro			- Journey Time & Frequency Report - Headway Performance - Service Availability Report	- Service Interruption - Patronage Report - Police Reported Crime (BOCSAR) - Asset Failures - Asset Management - Safety Incidents - Training Videos - Productionise Dashboards
	Data Ingestion		- Ingest Timetable & Trip Data - All modes - OpenData Self-Service - Ferry data	- Productionise Timetable & Trip Data Ingestion - OpenData Self-Service – API Access to ODL	- OpenData Self-Service – Direct Access to ODL - Training Videos
ODL Foundation		- Non Prod Environment Design & Setup (Initial) - Data Acquisition & Processing Contracts - ODL Foundation Contract	Non Prod Environment Design & Setup (Final)	- MPR MVP Solution Architecture - MPR Prod Environment Build & Migration - Initial ODL Platform Solution Review with Microsoft & Databricks	- ODL Platform Solution Architecture (Final) Approval by TDA & OTASC - Prod Environment Build - External Review (Microsoft++) Platform

Figure 5-3. MPR Delivery Roadmap

Transport as an organisation matured in its delivery of these six DSIs. When Harris, the Test Lead on the Ferry Performance Reporting Project, raised risk MPR-R-008 “Risk that business requirements are not delivered”, Sophie, our Portfolio Manager, requested in our regular program risk review meeting that we accept the risk with a note “... requirements will evolve with data discovery and development of dashboards.”

Our data-related processes matured as Harvey, our developer, commented at our 20th April 2020 retrospectives meeting, “Some nice design in data loading to standardise the data loading.” He added, “Consider data management is part of instead of addition to the implementation. One goal is to be able to hand over to somebody else to support as well as share the best practice and design. The value is beyond the project scope.” Mo, our information architect, commented in the same retrospective “CSELR Automation have gone well. The data being refreshed on a daily

basis.” At the 4th May 2020 retrospectives, Mo commented, “TCB Data ingestion went quite good.”

Deep engagement with business is essential for Agile delivery. The team acknowledged this requirement several times during the retrospectives: “*Work closely with business stakeholders*” (5th April 2020); “*Stay nimble and agile in how we deliver.*” (13th July 2020); “*Being proactive in gathering AND finalising requirements – at least one sprint prior*”, “*Engage with stakeholders at an early stage*” and “*Stakeholder engagement during daily stand-up*”, “*More frequent checking of report with certain stakeholders or points of contact, especially when requirements are not clear*” (27th July 2020); and “*Involve business stakeholders in dashboard testing*” (10th September 2020).

Showcases with stakeholders and team velocity were acknowledged as key success factors. Abhra, our scrum master, lauded “*Team pace*” (5th April 2020); “*Improved performance. Appreciate the entire team*” (20th April 2020); “*Inclined growth on Productivity*” (4th May 2020); “*Maintain Sprint Velocity*” (1st June 2020); and “*Good pace of overall team velocity*” (13th July 2020) Retrospectives. “*Continue Show cases with business*” (5th April 2020); “*Showcase MPR Program to (non-Program) OS and CST Stakeholders*” (4th May 2020) and “*Showcase (sell) MPR outcomes to wider Transport Community.*” (15th June 2020) were some of the other acknowledgments in retrospectives.

A significant change took place when sponsors accepted that (1) uncertainty existed in the delivery of the right quality of data with appropriate business rules and visualisations and (2) that the MPR program team would deliver the best possible outcome in the shortest amount of time. This was a marked shift from the waterfall-oriented fixed time-cost-scope mindset.

The MPR program closed in June 2021 with component projects closing between October 2020 and May 2021. A scalable big-data platform was delivered to store, process and service the analytics needs of contract managers, operators and data analysts. The Agile delivery allowed business value to be delivered consistently and incrementally and our ability to deliver value allowed us to add more projects to the program.

Terry, the light rail performance reporting business sponsor, nominated the project team for internal operational systems rewards and recognition. The MPR program was a finalist in the “Excellence in Transport Data Award” category at the ITS Australia Awards in 2020. The MPR

Success Story published by our delivery partners ([Agile Analytics](#) and [Data Driven](#)) showcased the strength of partnership and using technology to deliver data-driven business outcomes.

Looking back at our first and sixth DSIs, the degree of risk around the delivery processes between Vanguard and MPR reduced as the organisation and delivery teams better understood the “how-to” of DSIs and moved from the Exploration to the Exploitation stages. However, what did not change was the uncertainty around datasets due to the number of datasets being ingested, the best location to source them, the inherent quality of their integrity and completeness and the complexity of their transformation, all of which resulted in very high data complexity, as shown in Table 4-4. The uncertainty was lower when the datasets were internal or known, but in the majority of DSIs, the datasets were external and the team did not know about them until we started ingesting them.

A review of the risks, dependencies and constraints of the six DSIs in Table 4-6 shows that a majority of the ratings fell into the “high” and “very high” categories, thus contributing to higher complexity and uncertainty in delivery. The ratings followed the TERM framework and were based on a monthly review of risks by the DSI delivery team.

Overall, this analysis led us to define our first characteristic of DSIs, namely, that they carry a high degree of uncertainty from initiation right through to the closing phases.

A comparison of the Business Case Benefits and Benefits Delivered from the six DSIs, as in Table 4-1 *Summary of Transport DSIs*, shows that the benefits to Transport were enablers for decision-making.

Table 5-3
Vanguard Business Case Quantified Benefits

Benefit Type	Current	Target	Benefit Measure	Benefiit Owner
Improved Fare Compliance	6.40%	5%	Non-Compliance Rate	Security and Revenue Protection
Improved fine payment	32%	45% (in Y7)	% paid on time	Sydney Trains, TfNSW.

Although the first DSI had “*Improved fare compliance*” as a business case benefit, as shown in Table 5-3, the delivery team was unable to establish the direct benefit contribution. It did provide analytics on fare evasion to Transport’s Security and Revenue Protection team, to transport officers and to the NSW Police but was not responsible for what was done to reduce the fare evasion,

including changing the behaviour of the travelling public. The team subsequently identified deeper socio-economic factors which lead to fare evasion in certain parts of Sydney and New South Wales, to the extent of finding repeat offenders. In the Vanguard Success Story video, Tony, who became the sponsor in 2018, said, “[The] *Vanguard program provided an innovative opportunity to consolidate transport crime statistics, security incidents and fare compliance data into one system. The challenge was that data sets were not centralised and held by several government agencies and key transport operators. File structures differed and data collation validation systems were not automated. There were some inefficiencies within the Security and Revenue Protection teams where time was lost collecting, cleaning and collating data rather than conducting analysis and developing strategies and operational outcomes in the Transport Cluster. So, we looked to find a better solution.*” Tony concluded, “*Vanguard has consolidated key datasets. It has greatly enhanced our capability to analyse trends, identify hot spots and share these insights with our partners.*” Vanguard laid the foundation for data management and DSI delivery and became a showcase for people, process and technology outcomes. It was nominated for Operational Systems Rewards and Recognition, 2018 Transport (Cluster) Awards under the Safety Category and was a finalist in the 2019 PMI’s “Innovation in Project Management” Awards category. In the nomination for the awards, I listed the direct and indirect outcomes that Vanguard delivered, as shown in *Figure 5-4*:

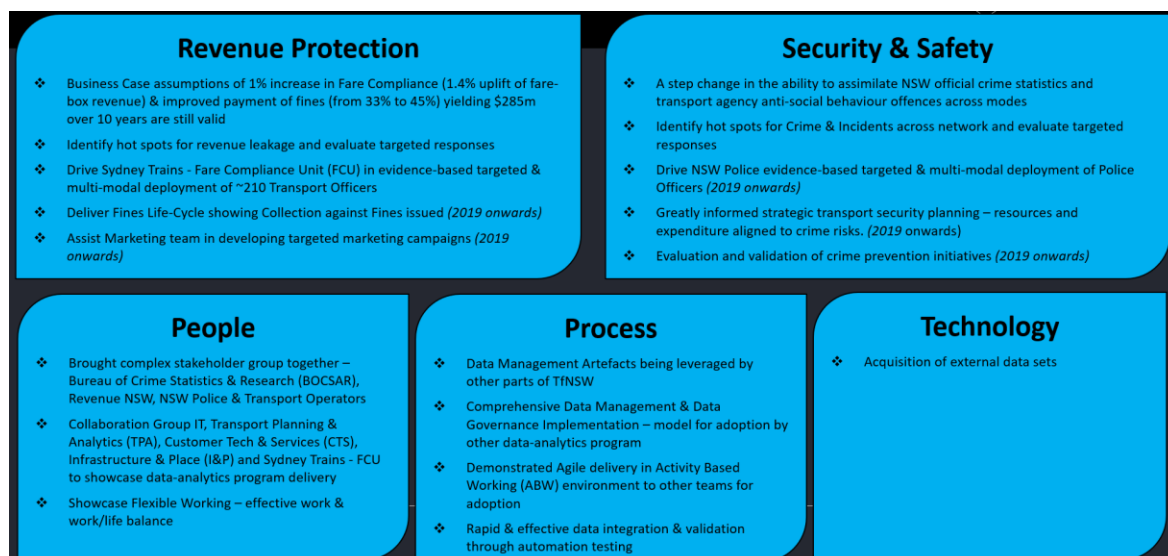


Figure 5-4. Outcomes Delivered by Vanguard

All the subsequent business cases showed benefits as enablers. The sixth DSI, MPR, gave us insights into bus, ferry, light rail and Sydney Metro performance; however, at best they allowed the contract management teams to identify and resolve only operational issues. I was unable to quantify any hard benefits, such as performance penalties, from the insights we gained in the sixth DSI, MPR. Examples of the benefits outlined in the Light Rail Performance Reporting Business case are:

- obtained evidence-based and data-driven contract management for KPIs and performance management,
- avoided costs of stand-alone data repositories and analytical services, and
- reduced system complexity because of standardised data exchange.

This was different from ICT business cases where there is typically a link to measure increases in revenue, customer satisfaction, efficiency or a reduction in costs. This led us to our second characteristic, that DSIs are enablers for decision-making but may not have a direct benefit contribution.

A review of the business case benefits in Table 4-1 *Summary of Transport DSIs* and my assessment of benefits complexity and requirements uncertainty in Table 4-9 showed a general lack of clarity in what we wanted to achieve out of these investments and how we wanted to achieve them. The ratings were based on the judgement of the team and reflect the relative complexity and uncertainty of the six DSIs.

Whether it was Vanguard or MPR, we had no certainty that with the datasets identified during the business case development, we would achieve the identified outcomes. This led us to our third characteristic, that neither the goals for a DSI nor the means of attaining them can be clearly defined from the outset.

Our case study of the six DSIs in this research showed the dependency they had on each other. The first two DSIs built our foundational knowledge and the capability to deliver them. From the third DSI onwards, more services, datasets and DSI delivery processes were added to and improved our technology platform. The Benefits Mapping and Dependencies of the six DSIs are shown in *Figure 5-5*.

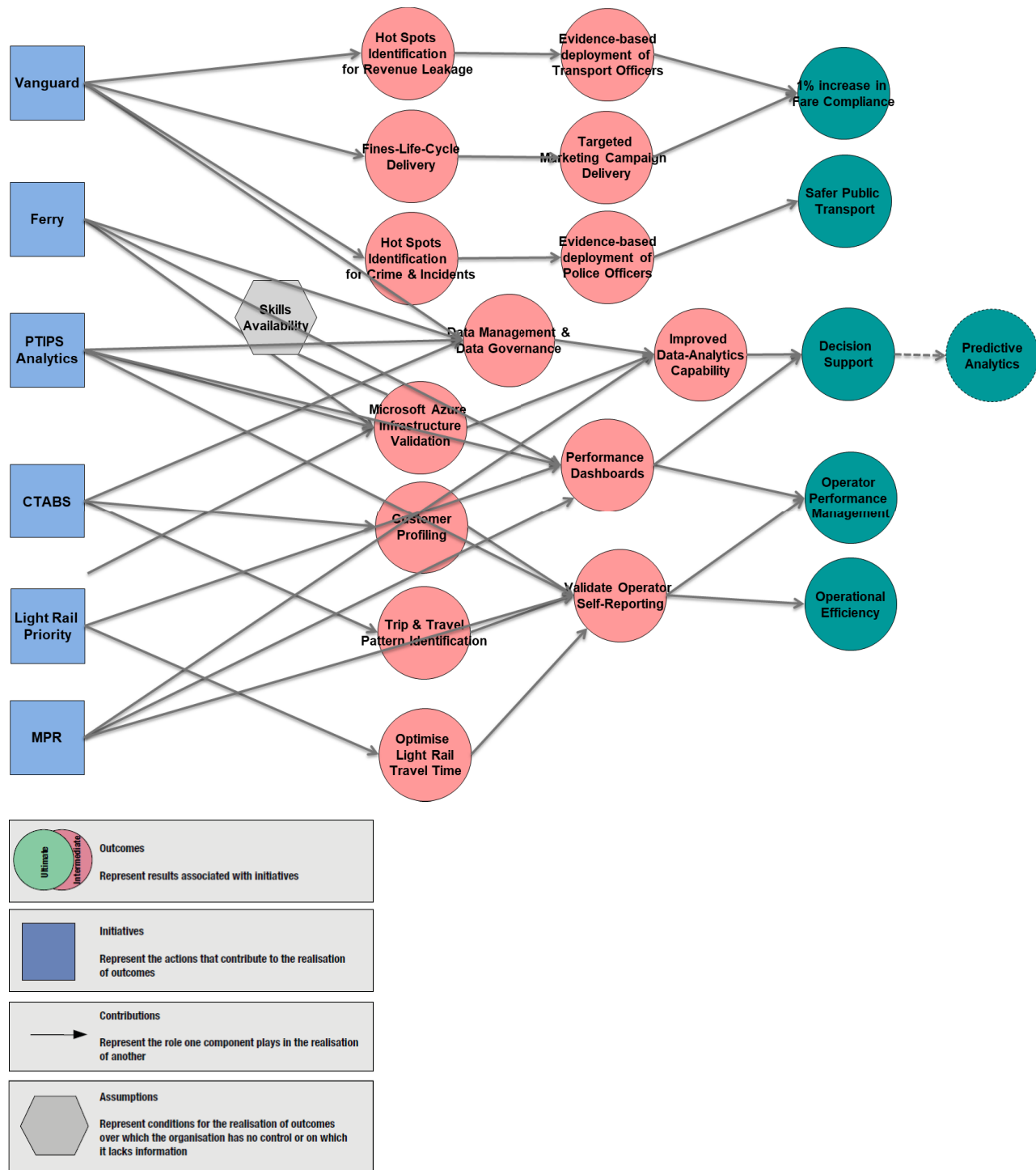


Figure 5-5. Benefits Mapping and Dependencies of the six DSIs

I saw that the sixth DSI both used the foundational layers delivered by the previous DSIs and reused the datasets ingested by them. While there was often no hard dependency on datasets

previously ingested, the insights got richer in each subsequent DSI. This led us to our fourth characteristic that DSIs are not independent of each other and act as an enabler for the next one.

Our delivery of the six DSIs showed that specific skills in each of program management, change management, scaled agile, data management and data science domains are often required, as shown in Table 4-7. Depending upon the type of DSI, the skill level and amount of time required for each of the roles may vary but they are all essential to successful delivery. I also noted that an ICT program using agile methods will have some of the roles defined in Table 4-7 but there are skills in the data management and data science domains that are specific to DSIs. The last two columns in Table 4-7 compare which roles are essential in DSI programs compared to ICT programs. This led us to our fifth characteristic, that the skills required to deliver a DSI are different to those required for a typical ICT program.

All six DSIs are now closed and this has allowed us to carry out both real-time and retrospective data collection. While the scale of the DSIs is different, together they have allowed us to identify characteristics of DSIs which brought uncertainty into their management and governance. Table 5-4 shows the gaps and issues I observed across the three program phases, namely, pre-, during and post-delivery of the DSIs.

Table 5-4
Program Life Cycle Deliverables and DSI Gaps and Issues

Key Phase Deliverables	Gaps and Issues for DSIs
Program Definition Phase	
Key deliverables of this phase were Business Case, Program Charter and Program Management Plan.	- For DSIs, risks associated with both costs and benefits are high. Considering the time it takes to develop and get a business case approved in both the public and private sectors, the accuracy of the documents is questionable. - Unless the Program Management Plan stays at a high level, the accuracy of scope and schedule is low. The delivery mechanism will evolve as the components are identified and executed.
Program Delivery Phase	
In this phase, individual components are	For DSIs, identification of all components

initiated, planned, executed, transitioned and closed while benefits are delivered, transitioned and sustained in accordance with the Program Management Plan.

upfront is difficult at the time the Program Management Plan is developed and hence only limited planning can be carried out due to the high degree of uncertainty.

- The benefits are discovered as the components are planned and executed, again due to the high degree of uncertainty.

Program Closure Phase

In this phase, the Program Benefits are transitioned to the sustaining organisation and the program is closed.

- While sponsors and stakeholders continuously communicate and are kept informed about both the costs and benefits delivered, for an uninitiated stakeholder, the value delivered by the program may be questionable. The outcomes are often enablers to organisational decision-making capability rather than absolute financial and non-financial metrics.

The six DSIs gave us a good initial understanding of their unique characteristics. These were further validated during the semi-structured interviews with practitioners from five organisations. Table 5-5 provides a background of the semi-structured interview participants, all of whom were practitioners delivering DSIs.

Table 5-5
About Semi-Structured Interview Participants

Name	Years of Experience	DSI Delivery Experience	Interview Date
Abhijit	13	13	22/05/2021
Mo	27	27	28/05/2021
Iman	20	18	31/05/2021
Kale	7	7	23/05/2021
Rodney	23	5	1/07/2021

A semi-structured questionnaire consisting of 17 questions was prepared for the qualitative data collection. Open questions were used and all participants were asked same set of questions during the one-hour interviews. The objective of the questionnaire was to gain an appreciation from practitioners of the challenges they faced in delivering DSIs and to compare the

characteristics identified through our case studies at Transport with their cross-industry experience. The first section of the questionnaire was intended to find more about the participants, their organisation and typical projects they had delivered. The second section was structured to reveal the participants' views of DSIs, how they compared with other ICT programs and the frameworks/methodologies they used. The third section was used to validate each of the five characteristics of DSIs that had emerged from my case study of the six Transport DSIs. In the last section, which allowed any other comments or thoughts to be expressed, the participants were asked if they would like to add any other characteristics of DSIs. Based on their responses, a sixth characteristic emerged. The questions across four sections are listed in Table 5-6.

Table 5-6
Interview Questions

Section 1 – Interviewee Background	
1	Could you please tell me about your background? How long have you been working in ICT sector? What type of industries have you worked in?
2	What is a typical week in your role look like? What are the challenges you face?
Section 2 – Comments on DSIs	
1	What is DSI to you? Please share your experiences of a DSI program you have worked on. What was your role and what was the size of Team/Budget/Duration?
2	How long have you been managing DSIs? What has changed in terms of people, process and technology?
3	How do you compare an ICT program with DSI? What similarities do you see? What differences do you see? Has this changed over time? If so, why?
4	What Delivery Frameworks/Methodologies have you used for DSIs. Why did you choose them? Is there a criterion you use when choosing the Delivery Framework / Methodology for a DSI?
5	Were the DSIs you delivered considered successful by Sponsor/Client. What made them successful? When were they considered not to be successful?
6	What key risks and issues have you seen in DSIs?
7	Anything you could have done differently to deliver DSIs?
Section 3 – Validation of DSI Characteristics	
1	DSIs carry high degree of uncertainty right from initiation through to closing phases.
2	DSIs are enablers for decision-making and may not have a direct benefit contribution.
3	Neither the goals nor the means of attaining them are clearly defined from the outset for a DSI.
4	DSIs are not independent of each other and act as an enabler to next one.

- 5 Skills required to deliver a DSI are different to typical ICT program.
- 6 DSIs do not end and after initial delivery convert into continuous business improvement initiative.

Section 4 – Closing

- 1 Is there any other part of DSI which you would like to discuss?
- 2 Can you think of other people I should talk to about DSIs?

In the semi-structured interviews, I wanted to probe the themes that consistently emerged in the six Transport case studies. When asked about uncertainty, four out of five respondents agreed with this characteristic and shared what they were doing to address it. However, one respondent highlighted the increasing maturity associated with DSIs and the availability of “out-of-the-box” models, which can reduce uncertainty in specific cases. Iman commented on the existence of unknowns in DSIs, saying, *“There is [a] cone of uncertainty and for DSIs the variation from beginning [to end] could be plus minus 99%.”* Iman justified using Agile methods *“to reduce level of uncertainty by developing, showing and getting feedback in agile and iterative way”*. Kale acknowledged the inherent risks in most DSIs and mentioned that his consulting organisation had a high bar when selecting client DSIs for delivery. His organisation did not accept DSIs for delivery if they carried a high risk; this was to maintain high customer satisfaction and as a result, *“We have got most initiatives to probably a 98% success rate and very low failure rates and adoption rates from projects at the moment.”* He emphasised the risk-averse approach of, *“If we don't think we can do it, we won't take it.”* An obvious downside of this approach was that Kale was walking away from some business opportunities in a market where the demand was high and the supply of skilled resources comparatively low. Rodney agreed that within an organisation, the degree of uncertainty started to drop once the organisation had done a few DSIs, validating their exploratory nature with the comment, *“You learn lot of hard lessons on the first one.”* However, Abhijit pointed out that with the availability of *“plug-and-play use cases”* the DSI industry was maturing. The increasing availability of such solutions allowed system integrators to reuse models: *“It's become a 60-40 mix, 60% of it is pre-built stuff that people are just coming and plugging in for you now.”* Abhijit acknowledged, *“When people are doing exploration, your assumption is 100% right.”*

This led me to conclude that when the data was from a well-defined and structured source, such as Salesforce, SAP or Peoplesoft applications, the degree of uncertainty was reduced and

organisations could choose to use pre-built solutions. Hence this characteristic needed to be qualified by the type and complexity of the data source and the model being built. At the same time, the responses validated our hypothesis of the emerging shift from exploration to exploitation as the organisation and industry matured in DSI delivery. This allowed us to finalise our first characteristic that *“DSIs carry a high degree of uncertainty right from initiation through to the closing phases, except for when the data is from a well-defined and structured source.”*

While the characteristic of DSIs as enablers was true for the Transport case studies, a different story emerged from the interviews. Mo acknowledged the characteristic but also highlighted an example from work he did with traders at Energy NZ, saying, *“There can be situations where you can actually allocate benefits.”* Iman agreed that while DSIs often did not deliver a tangible outcome, they did validate a few things: *“At the end we have learning and knowledge and that knowledge is the outcome of DSI project.”* Kale’s experience was like Mo’s. He said, *“There are cases when they can have a direct benefit contribution”* and cited examples where a business derived five to 10 times its investment through a DSI. Rodney categorised DSIs. His first category was a proof of value (PoV) or proof of concept (PoC) and his second category was operationalising a solution, in relation to which he commented, *“It wouldn’t be fair to compare the operationalisation of the product to the PoV.”* His view was that in PoV and PoC an idea or theory is tested on whether to move forward or not. At the same time, however, he said that he had *“... seen examples where by implementing a DSI, businesses were able to actually save dollars.”* Abhijit echoed these sentiments, saying, *“I think it’s bit of half and half.”*

Hence, I can say that the validity of this enabling characteristic is based on the context in which a DSI is delivered. This allowed us to finalise our second characteristic, that *“DSIs are often enablers for decision-making and may not have a direct benefit contribution.”*

The “lack of clarity on goals and means of attaining them” characteristic was generally ratified by the interview responses. Mo agreed with the characteristic that you iterate and adjust what you are delivering: *“... need to go through the discovery process and understand what your final goal is.”* Iman mentioned that, unlike an engineering design, there are lot moving parts and *“... in DSIs you don’t know that it is a 4-bedroom, 2-bathroom and a backyard house.”* On managing expectations, he added, *“You don’t expect any set delivery from the beginning, even as you get to know better about technology possibility.”* Kale highlighted the fluidity in starting DSIs,

saying, *“What you end up with is not the same thing and so what that means is that very few organisations know the proper requirements of what they're looking to achieve when they approach an ML initiative.”* He emphasised that, unlike building software where you are lot more definitive around ways and outcomes, you are not never sure where you will end up with DSIs. Rodney agreed with the characteristic, with a caveat that as the market matures the goals will become clear. He said *“The cloud vendors are packaging these services up and making you pay by the time you use it.”* Abhijit said that with the emergence of pre-built solutions for standard data sources, the *“how-to”* was becoming clearer. However, he felt that the definition of success for a DSI engagement was one of the biggest challenges today, saying, *“The customer sometimes would not know what the end game is they're looking for.”*

The exploratory nature of DSIs was highlighted in this characteristic and ratified by interviews, with the caveat that as the market matures, the emergence of pre-built solutions will reduce the uncertainty. This allowed us to map DSIs to Type-4 projects (Turner & Cochrane, 1993) and finalise our third characteristic, that *“Neither the goals nor the means of attaining them are clearly defined from the outset for a DSI with the caveat that as the market matures, the emergence of pre-built solutions will reduce the uncertainty.”*

The interdependency of DSIs was emphasised throughout the interview responses. Mo agreed, *“The sequencing is very important as it gives the foundation to the subsequent ones.”* Iman further elaborated that this was a knowledge-building process where you are adding layers: *“You build the first knowledge set and based on that you determine the next level on top of that.”* Kale commented on both technology and people; on the technology aspect he flagged breaking DSIs into comparable streams of similar use cases: *“You can use a template from the first one to solve the second one, assuming you're trying to solve similar use cases like customer churn, cross sell.”* On the people side, he commented, *“They get more knowledge every single time they do it, so there are incremental gains.”* Rodney agreed that there was a layered approach to DSIs where you keep iterating: *“It's kind of the agile continuous improvement idea to a delivery–fail fast and move on as opposed to fail slow.”* Abhijit agreed that DSIs can help drive practitioners towards a business outcome and often are quite related. *“You move from one use case to another use case, which is essentially taking you from where you are to your mission and vision of where you want to be in two or three years.”*

In summary, the characteristic was validated and allowed us to finalise our fourth characteristic that *“DSIs are not independent of each other and act as an enabler for the next one.”*

The skills requirement characteristic was apparent in all the interview responses, reflecting the evolving nature of the ICT sector and especially data science. Iman mentioned skillsets in the delivery and implementation domains and the need for cross-functional teams that understand the end-to-end delivery process: *“The same person may play different role at different stages of DSI depending on the complexity and size of project.”* Kale also mentioned different skill sets with reference to software engineering versus data science in terms of programming, programming languages, techniques and even mindset. He commented, *“When you talk about DSIs, it is hard problem-solving using data techniques and I'm not saying that ICT doesn't have to do that, but you are architecting and re-engineering business processes at the same time, which could be a very different thing to - I'm going to manage security or networking or infrastructure.”* Rodney elaborated on that, saying that five years previously some of these roles did not even exist. Abhijit went further, saying that besides technical skills, *“I think skill varies significantly in understanding and appreciating the business context.”*

In summary, this characteristic was validated quite emphatically in all responses and allowed us to finalise the fifth characteristic, that *“Skills required to deliver a DSI are different to those required for a typical ICT program.”*

Our sixth DSI characteristic came up when the interviewees were asked if they had seen any other characteristic. Iman, who first raised this, saw a pattern that once you show value and what is possible, DSIs can go from one phase to another as the business wishes to go further or expand it to other functional areas: *“It has to be ongoing and a mindset of not making it finite in terms of budget and delivery.”* Kale agreed that the business context changes overtime and what that means is that once a DSI is delivered, the product needs to be monitored and improved over time. *“I've got a saying that we don't do projects, we build products and it's exactly that point, which is, once you've built something, a model will change overtime.”* Rodney concurred. *“That's the nature of software now. [...] the software is never finished; it's just continually changing.”* Abhijit mentioned his experience in the pharmaceutical device manufacturing industry where *“... they have grown and scaled that model out.”*

While this characteristic was thought of later, it was still validated and shows how DSIs deliver products with data and models that have both evolved over time. In relation to the six Transport DSI case studies in this research, five products are still being used and continuously improved, thereby validating this characteristic. This allowed us to finalise our sixth characteristic that “*DSIs do not end and after initial delivery convert into managing the product, model and data.*”

5.6 CONCLUSION AND RECOMMENDATIONS

In this section, I review our research question: “What unique characteristics cause DSIs to face challenges delivering envisaged value when using traditional processes for managing ICT-enabled programs?” and summarise my conclusion. I first start with the conclusions from this chapter, discuss the limitations of my research and the implications for program managers, portfolio managers/policy makers and academics in understanding characteristics of DSIs and end with the recommendations.

5.6.1 Conclusion

I conclude that the current literature does not adequately cover the unique characteristics of DSIs and that business managers and practitioners need to be informed about the differences between DSIs and ICT-enabled programs so that they can adopt methods to improve the chance of successful business outcomes.

Program Management for ICT-enabled programs has rich literature and proven delivery frameworks which have matured over the past three decades (Axelos, 2022; Project Management Institute, 2016, 2017b). This chapter makes a significant contribution to the practice of the emerging field of data science and program management.

The current program management literature does not adequately support delivery of innovative and exploratory DSIs and instead focuses on risk elimination and rapid delivery of business outcomes for exploitative initiatives. In this chapter I have identified six unique characteristics of DSIs to be used in the delivery of DSIs and complementary domains, as identified in the literature review and in practice. These are the PMI’s *The Standard for Program Management* (Project Management Institute, 2017b) for program management; the Proscii Framework (Hiatt, 2006) for people change management; Scaled Agile (SAFe) (Scaled Agile,

2023) for solution delivery, DAMA's DMBOK (Earley, 2017) for data management and CRISP-DM (Chapman et al., 2000) for data science processes. Practitioners should consider integrating these domains into any DSI delivery framework they are developing to ensure that the envisaged value is delivered and to address some of the challenges that I faced over five years of delivering DSIs. While each of the highlighted domains is rich in information and mature, the lack of integration will cause the continued failure of DSIs.

I also conclude that the organisation and its teams go through the stages of exploration, transition and exploitation in DSI delivery and the DSI delivery process becomes more efficient as more are delivered. With the exception of data from a well-defined and structured source, every new dataset carries uncertainty in scope and quality. This brings in the need to frame the underpinning exploration component of a DSI combined with a shift to exploitation as the organisation and teams mature.

From a theoretical contribution, I have expanded the exploration and uncertainty in projects (Alvarez, Afuah, & Gibson, 2018; Lenfle, 2008; CH Loch et al., 2011) into the realm of data science and shown how they contribute to the failure (or success) of DSI delivery. There is a growing concern that applying an incorrect approach to management of projects in the face of uncertainty is both common and detrimental to their performance. Atkinson, Crawford, and Ward (2006) raised concerns that common project management practice does not address fundamental sources of uncertainty, especially in "soft" projects where flexibility and tolerance of vagueness are necessary. They attributed uncertainty to lack of information, ambiguity, characteristics of project parties, trade-offs between trust and control mechanisms and varying agendas in different stages of the project life cycle. Innovative and exploratory projects face challenges when organisations attempt a standardised and routinised approach to manage uncertainty instead of applying an adaptive model and treating innovative projects as "voyages of discovery" (Davies et al., 2018, p. 969).

5.6.2 Limitations and Implications of Research

This research used six DSIs from one public sector organisation in Australia as case studies to identify unique characteristics and then validated with semi-structured interviews with practitioners from five organisations. I undertook development of a DSI delivery framework that incorporated program management, change management, agile delivery, data management and

data science domains to assist practitioners. Sandberg (2005) notes that truth is always something unfinished within the interpretive tradition and researchers cannot generate absolute truth claims. I believe that the DSI characteristics I have identified do not present an exhaustive and universal set and more may emerge as the field of data science advances. Future research can include validating the characteristics with other public and private sector organisations delivering DSIs in other countries. Another aspect is that DSIs are a more recent phenomenon and sit in a rapidly evolving technology and delivery space. This has an impact on the currency of the research work being done, as some of the characteristics will change as the maturity of DSIs changes from being exploratory to exploitative.

Limited availability of methods and standards in DSI delivery has caused business managers and practitioners to chart their own path and thus there is inconsistency in how DSIs are treated and delivered in different organisations. With the emergence of research such as this, it is expected that the standardisation on DSIs will increase and help guide practitioners in the efficient delivery of DSIs.

5.6.3 Recommendations

I started with five unique characteristics and validated them using semi-structured interviews with five practitioners. As a result, the characteristics have been updated and a sixth one has emerged. The six characteristics are summarised in Table 5-7:

Table 5-7
Unique Characteristics of DSIs

No	Description
(i)	DSIs carry a high degree of uncertainty from initiation through to their closing phases except for when the data is from a well-defined and structured source.
(ii)	DSIs are often enablers for decision-making and may not have a direct benefit contribution.
(iii)	Neither the goals nor the means of attaining them are clearly defined from the outset for a DSI, with the caveat that as the market matures, the emergence of pre-built solutions will reduce the uncertainty.
(iv)	DSIs are not independent of each other and can act as enablers to the next one.
(v)	The skills required to deliver a DSI are different from those required for a typical ICT program.
(vi)	DSIs do not end and after the initial delivery but convert into management of the product, the model and the data.

The characteristics described earlier make DSIs different from ICT-enabled programs. While there are exceptions on both sides, Table 5-8 highlights some of the key differences between DSIs and ICT-enabled programs.

Table 5-8
Key differences between DSIs and ICT-enabled Programs

Area	DSIs	ICT-enabled Programs
Class	Primarily exploratory	Primarily exploitative
Uncertainty	Medium to high degree	Low to medium degree
Benefits Identification	Often have no direct benefits	Direct benefits can be identified
Experimentation	High degree	Low to medium degree
Enablers	Benefit from previous implementations	May or may not benefit from previous implementations
Skill set	Specific data management and data science skills are mandatory	Do not require specific data management and data science skills
Product Management	Product, model and data need to be maintained	Product lifecycle is not always required

In addition, I recommend exploiting the concept of ambidextrous organisation. Organisations should aim to balance exploration and exploitation activities effectively. This entails simultaneously pursuing new opportunities (exploration) while also optimising existing processes (exploitation). By fostering ambidexterity, organisations can adapt to the uncertain and exploratory nature of DSIs while still ensuring efficiency and effectiveness in their operations. Leaders should encourage a culture that values innovation and experimentation while also maintaining a focus on delivering tangible outcomes. Implementing ambidextrous practices can enhance an organisation's ability to navigate the complexities of DSI delivery and ultimately improve its success rate.

I suggest additional research to validate these characteristics with other public and private sector organisations delivering DSIs. As the field is evolving rapidly, I believe that the six characteristics identified in this chapter will also evolve. It is possible that new characteristics may emerge and some identified here may no longer be treated as unique as DSIs go from being exploratory to exploitative. With the size of investment under way in DSIs and the need for data-driven decision-making in organisations, additional research is essential in the field of DSIs. The

characteristics identified in this chapter will deliver a small but significant contribution to the body of knowledge for program management relevant to both literature and practitioners. Without this understanding, there will be more failed programs and dissatisfied sponsors and delay much-needed investment in this emerging field. It will also delay the benefits that will flow from harnessing the data and improving data-driven decisioning capability.

6. Framework to Manage DSIs

This chapter is based on my published paper “A Framework to Manage Data Science Initiatives” (Mathur, Sankaran, et al., 2021).

6.1 ABSTRACT

The success rate of delivering DSIs is not perceived to be high, with Gartner estimating that 85% of projects fail (Asay, 2017). DSIs have unique characteristics and pose challenges in delivering the envisaged value when using traditional processes for managing ICT-enabled programs. There are occasions when DSIs should be managed as exploratory projects. In this chapter, I review the related delivery frameworks and propose a framework that synthesises program management, change management, scaled agile, data management and data science to manage DSIs effectively. The framework covers people and processes and, to keep the framework generic, specifically excludes rapidly evolving products and technologies. The framework may enable consistency in how practitioners plan and execute DSIs, potentially leading to an improvement in the success rate of DSI implementations.

6.2 INTRODUCTION

DSIs have unique characteristics and pose challenges delivering the envisaged value when using traditional processes for managing ICT-enabled programs. Due to the uncertainties involved in their data, scope and schedule, DSIs often present themselves as candidates to be managed as exploratory projects. Furthermore, the waterfall approaches to program management adopted by peak bodies such as Project Management Institute, Axelos, International Project Management Association, etc. set up structural tensions between business case development, program design, delivery and benefits realisations that decouple value creation from capture and thus undermine coherent governance across the investment life cycle. In this chapter, I expand on existing frameworks proposed in the program management, change management, scaled agile, data management and data science domains and propose a synthesised framework to deliver DSIs as exploratory projects. While existing frameworks are expected to develop as the data science field

matures, they are currently inadequate to deal with such projects. Therefore, it was timely to undertake an investigation to address the gap identified in the literature and in practice to deal with the uncertainty and complexity experienced in delivering DSIs.

Balancing exploration and exploitation is key to organisational success and survival (March, 1991). Projects often act as the organisational vehicle for delivering change (Brady & Davies, 2004). Exploitative projects focus on optimising the triple constraints of cost-quality-time to deliver new products and services whereas exploratory projects are projects where neither the goals nor the means of attaining them are clearly defined from the outset (Lenfle, 2008). This framing brings to the surface the fact that the traditional view of project management as the accomplishment of a clearly defined goal with cost-quality-time triple constraints does not fit neatly with the logic of innovation that is first and foremost characterised by discovery (Van de Ven, Polley, Garud, & Venkataraman, 1999), unforeseeable uncertainty (CH Loch et al., 2011) and expansion (Hatchuel, 2001).

In this chapter, I draw on in-depth case studies of six DSIs delivered between 2017 and 2022 at Transport for NSW (Transport). Transport is a state government agency responsible for delivering safe, integrated and efficient transport systems to the people of New South Wales. At the time of this research, Transport was in the early stages of executing AU\$41.5 billion worth of projects for its Future Transport 2056 Strategy. More than 300 initiatives were identified for delivery in the first 10 years of the strategy, underpinned by data-driven technology roadmap. The majority of the initiatives had a significant technology component, including data and data analytics. Transport generates significant amount of data from its bus, ferry, light rail, metro and heavy rail modes every day. As big-data technology has matured in the past few years, it has given Transport an ability to store, transform and use the data generated from internet of things (IoT) devices, something it was unable to do effectively earlier. *Figure 4-1* provides an overview of the six DSIs used as case studies, including their delivery timeline.

As per Unique Characteristics of DSIs, the case study of the six DSIs highlighted that they have unique characteristics:

- (i) **Degree of Uncertainty:** Like mining and oil exploratory projects, DSIs carry a high degree of uncertainty. It is hard to know what nuggets you will get from the data. You do not know how much time it will take. The quality of the data is relatively unknown as you are relying

on other operational systems to provide them. Tracking a DSI using a detailed schedule is almost impossible.

- (ii) **Enablers for Decision-Making:** Unless you are a provider of data or in the data services business, DSIs deliver a data product or a data service which will be used by the sponsoring organisation for decision-making. While decisions may have a real outcome for the organisation, the DSI itself acts only as an enabler. Hence the business case for DSIs can be difficult for as decision-makers to justify in terms of hard costs and quantified benefits.
- (iii) **Unclear Goals:** While high-level outcomes are often known, the path to reach them is not. There is high degree of experimentation.
- (iv) **Interdependency:** Each DSI potentially leverages the previous one for data management and data governance. These include shared reference/master data, security and governance requirements.
- (v) **Skills Requirement:** For a DSI to be successful, specific skills are required of the team. Some of these include product owner, data architect, data analyst, data scientist, UX (user experience) designer, visualisation, cloud engineer skills etc. A standard ICT program team will not deliver a successful outcome.

I aim to contribute to project and program management research by proposing a synthesised framework explicating the different logic (Lenfle, 2016) required to deliver DSIs as exploratory projects and incorporating program management, change management, agile delivery, data management and data science domains. This contribution will help fill an important gap in the literature on exploratory projects concerning the fine-grained details of their management (Christoph Loch & Sommer, 2019). It also highlights the important contribution project management scholars can make to the understanding of how value gets created through DSIs.

6.3 LITERATURE REVIEW

Creation of a comprehensive program delivery framework for DSIs requires understanding of several domains and this section attempts to cover them, as shown in *Figure 6-1*. I start with the program management literature, which has matured since 1990. I then move to the people side of initiatives and discuss change management. I argue that Agile software development methods are

more suitable for addressing the exploratory nature of DSIs and commence with a review of Scaled Agile frameworks for large-scale agile delivery. I then introduce data to this chapter by exploring data management and conclude by discussing data mining and data science project management and processes.

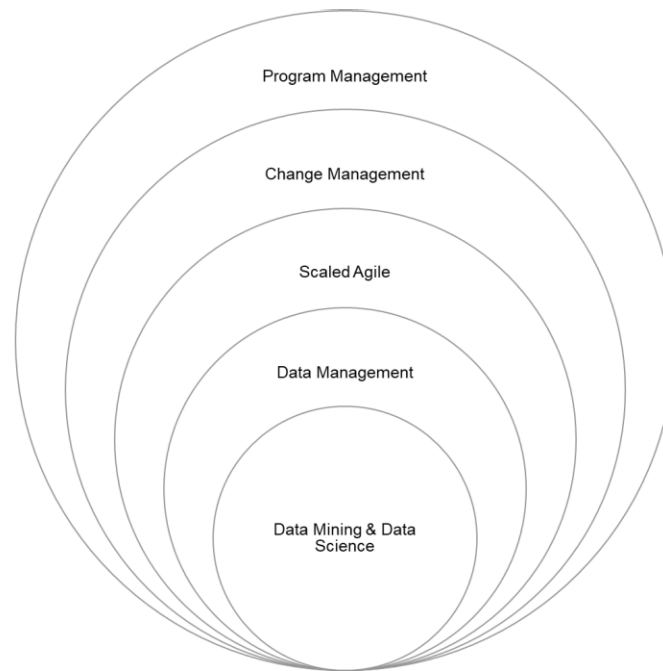


Figure 6-1. Understanding the Knowledge Gap for DSIs

6.3.1 Program Management

I see DSIs being typically implemented as a program on a continuous spectrum rather than a single one-off project and will focus on program management rather than project management processes for delivery. “A program is defined as related projects, subsidiary programs and program activities managed in a coordinated manner to obtain benefits not available from managing them individually” (Project Management Institute, 2017b, p. 3). “Program management is defined as the application of knowledge, skills and principles to a program to achieve the program objectives and to obtain benefits and control not available by managing program components individually” (Project Management Institute, 2017b, p. 8). A review of the program life cycle has identified gaps in its use for DSI delivery. A typical program life cycle according to the PMI standard contains three phases: Program Definition, Program Delivery and Program Closure, as shown in *Figure 1-1*.

MSP is another program management framework whereby large complex change can be broken down into manageable, interrelated projects (Axelos, 2022). The MSP framework is based on three core concepts: MSP Principles; MSP Governance Themes; and MSP Transformational Flow.

Both the PMI and Axelos standards are widely accepted and used in the industry, with the PMI standard being principle-based and Axelos providing detailed guidance.

6.3.2 Change Management

Value realisation for any program occurs when the product and service created are adopted by users. Change management is a systematic approach that includes dealing with the transition or transformation of organisational goals, core values, processes or technologies. Kotter's Change Management Model is one of the most popular in the world (J. Kotter, 2007). It has eight stages, each focused on employees' response to change: (i) Create a sense of urgency; (ii) Build a guiding coalition; (iii) Form strategic vision and initiatives; (iv) Enlist volunteer army; (v) Enable action by removing barriers; (vi) Generate short-term wins; (vii) Sustain acceleration; and (viii) Institute change.

McKinsey's 7-S Change Management Model is one of the longest-lasting change management models serving professionals (Lorenzi & Waterman, 1985). It consists of seven crucial categories: strategy, structure, systems, shared values, style, staff and skills that companies should be aware of when implementing change.

The ADKAR change management model is used by change managers to find out various gaps in the process so that effective training can be offered to employees (Hiatt, 2006). Even though the ADKAR model focuses on business-oriented goals, it can be very useful in supporting employees to go through the process of awareness, desire, knowledge, ability and reinforcement more easily.

The Kübler-Ross Five Stage Change Management Model helps employers understand their employees better and empathise with them (Kübler-Ross, 2009). This model can be applied to other life situations, such as loss of job, changes in work and less serious health conditions. It consists of five stages, namely denial, anger, bargaining, depression and acceptance, that employees may undergo during organisational changes.

A comparison of these four change management models is shown in Table 6-1, highlighting their benefits and limitations.

Table 6-1
Comparison of Change Management Models

Model	Description	Benefits	Limitations
Kotter's Model	Steps to encourage new behaviours for successful organisational change	Provides an eight-step actionable model	Lack of measurement processes and time consuming
McKinsey's 7-S Model	Seven categories structural model that focuses on a holistic approach to organisational change	Provides guidance and focuses on the whole organisation	Very complex model
ADKAR Model	Five step process: Awareness, Desire, Knowledge, Ability and Reinforcement	Rewards individual change in organisational change process	Cumbersome process for large organisations
Kubler Ross Five Stage Model	Model based on emotional journey – five stages of grief	Most change frameworks address these stages	No clear guidance for operational change

Of these four models, ADKAR and Kotter continue to be referenced and widely used by practitioners when delivering ICT change programs. In organisational change, such as restructuring, mergers and acquisitions, the Kübler-Ross model is often mentioned and used. The three models are popular because they are simple to understand and apply.

6.3.3 Scaled Agile

Large, often globally distributed projects have proliferated and their widespread teams need to collaborate and coordinate nationally and internationally. This has led to the popularity of scaled agile frameworks such as Scaled Agile Framework (SAFe), Large-Scale Scrum (LeSS) and LeanSAFE (Ebert & Paasivaara, 2017; Leffingwell, 2007). In the context of DSIs, I see the relevance of scaling in delivering data science outcomes to be as high as any other project involving multiple geographically-spread teams within an organisation.

SAFe is a set of organisation and workflow patterns for implementing agile practices at enterprise scale (Scaled Agile, 2023). The SAFe framework is a body of knowledge that includes

structured guidance on roles and responsibilities, how to plan and manage the work and values to uphold to achieve business agility while using Lean, Agile and DevOps.

Large Scale Scrum (LeSS) is a framework for scaling Scrum to multiple teams who work together on a single product. LeSS starts with a foundation of one Scrum team that then applies it to multiple teams who work together on the one product (Larman & Vodde, 2016). LeSS Huge is a second LeSS framework suitable for LeSS adoptions of more than eight teams (Larman & Vodde, 2016). Conceptually, it is LeSS scaled up further by having multiple (smaller) LeSS frameworks stacked on top of each other.

The Spotify model is a people-driven, autonomous framework for scaling agile (Kniberg & Ivarsson, 2012). This model emphasises the importance of culture and networking. It uses Squads, Tribes, Chapters and Guilds, but the foundation of the model is the squad, which acts as the Scrum team.

The Disciplined Agile 2.0 process decision framework provides light-weight guidance to help organisations streamline their information technology (IT) processes in a context-sensitive manner (Ambler & Lines, 2016). It does this by showing how various activities, such as solution delivery, operations, enterprise architecture, portfolio management and many others, work together in a cohesive whole. The framework also describes what these activities should address, provides a range of options for doing so and describes the trade-offs associated with each option. It has a risk-value delivery lifecycle, is goal-driven, is enterprise aware and is scalable.

Scrum@Scale is a framework in which a network of teams operating consistently with the Scrum Guide to address complex adaptive problems while creatively delivering products of the highest possible value (Sutherland, 2019). These “products” may be physical, digital, complex integrated systems, processes or services. The Scrum Guide describes the minimal set of components needed to create a team environment that drives innovation, customer satisfaction, performance and happiness.

A comparison of the five scaled agile frameworks in *Figure 2-18* shows their strengths across 16 criteria (Atlassian, 2020).

6.3.4 Data Management

Data management is the development, execution and supervision of plans, programs and practices that deliver, control, protect and enhance the value of data and information assets throughout their lifecycles (Earley, 2017). *Figure 2-17* defines the 11 data management knowledge areas with data governance at the centre of wheel. The other knowledge areas, while necessary, can be implemented at different times depending upon the requirements of the organisation.

6.3.5 Data Science Processes

A methodology is a general strategy that guides the processes and activities within a given domain. A methodology does not depend on technologies or tools, nor is it a set of techniques or recipes. Rather, it provides the data scientist with a framework with which to proceed with whatever methods, processes and heuristics will be used to obtain answers or results (Rollins, 2015). Development of a DSI delivery framework requires good understanding of the data mining and data science delivery processes.

The KDD model, first published in 1996, is one of the early iterative and interactive models to extract useful knowledge from data (Fayyad et al., 1996). It recognises the fact that the value of storing volumes of data depends on our ability to extract useful reports, spot interesting events and trends, support decisions and policy based on statistical analysis and inference and exploit the data to achieve business, operational or scientific goals.

Cross-Industry Standard Process for Data Mining (CRISP-DM) was launched in late 1996 by Daimler Chrysler (then Daimler-Benz), SPSS (then ISL) and NCR and is a widely-used analytics model methodology with six phases: business understanding, data understanding, data preparation, modelling, evaluation and deployment (Chapman et al., 2000).

The SEMMA model was developed by the SAS Institute (SAS Institute, 2009). This model has five sequential steps and guides the implementation of data mining applications. Although SEMMA is often considered to be a general data mining methodology, SAS claims that it is a logical organisation of the functional toolset of SAS Enterprise Miner for carrying out the core tasks of data mining.

Mason and Wiggins (2010) proposed a linear aperiodic Obtain, Scrub, Explore, Model and Interpret (OSEMN) model for data science processes.

TDSP, launched in 2016, is an agile, iterative data science methodology designed to deliver predictive analytics solutions and intelligent applications efficiently (Severtson et al., 2017). TDSP helps improve team collaboration and learning by suggesting how team roles work best together. It includes best practices and structures from Microsoft and other industry leaders to help successful implementation of DSIs and its goal is to help companies fully realise the benefits of their analytics program.

The FMDS methodology launched in 2015 has some similarities and many features of the KDD Process and CRISP-DM. In addition, it provides a number of new practices, such as use of extremely large data volumes, text and image analytics, deep learning, AI and language processing (Rollins, 2015). The FMDS's 10 steps illustrate the iterative nature of problem-solving when utilising data to discover security insights.

Larson and Chang (2016) acknowledge that business intelligence delivery with Agile methods has matured and have proposed dependent but distinct frameworks for business intelligence and data science delivery.

Azevedo and Santos (2008) compared the KDD, SEMMA and CRISP-DM processes and concluded that that SEMMA and CRISP-DM can be viewed as an implementation of the KDD process, as shown in *Figure 6-2*.

KDD	SEMMA	CRISP-DM
Pre KDD	-----	Business understanding
Selection	Sample	Data Understanding
Pre processing	Explore	
Transformation	Modify	Data preparation
Data mining	Model	Modeling
Interpretation/Evaluation	Assessment	Evaluation
Post KDD	-----	Deployment

Figure 6-2. Comparison of KDD, SEMMA and CRISP-DM (Azevedo & Santos, 2008, p. 5)

Foroughi and Luksch (2018) compared the KDD, CRISP-DM, FMDS and TDSP processes against four common iteratives stages of problem definition / formulation; data gathering; data modelling and data production, as shown in *Figure 2-19*. They found that the KDD process did not cover the business understanding and deployment phases of the CRISP-DM methodology. However, CRISP-DM did not have the analytic approach, identification of suitable data collection strategy and data resources and the feedback phases of FMDS. While FMDS and TDSP were very

similar, the detailed stages of FMDS could be more useful for a wide range of projects but TDSP used a specific set of Microsoft tools and infrastructure to deliver intelligent applications by deploying ML or AI models.

6.3.6 Summary and Implications

In this section, I reviewed various domains essential to developing a program delivery framework for DSIs. Our framework at Transport required end-to-end program management from business case through to benefits realisation. I reviewed change management to ensure that investment in DSI would deliver value to the organisation. I reviewed several scaled agile frameworks to assess their suitability for both smaller and large-scale DSI implementations. Since an understanding of data management is essential to all DSIs, I reviewed DAMA's DMBOK. I then reviewed several data mining and data science methods used since about 1995 to find that FMDS and TDSP from IBM and Microsoft respectively were built upon the solid foundations of KDD and CRISP-DM.

This view of the literature motivated us to ask the following research question: "What design principles should be incorporated in a DSI delivery framework so that program managers can adopt a predictive path to realise value from such investments?"

6.4 RESEARCH SETTING AND METHODS

6.4.1 Research Methods

Our research insights emerged from my desire to deliver DSIs effectively underpinned by Transport's Future Transport 2056 Strategy to embed technologies such as big data, the IoT, ML and AI to deliver and improve customer journeys. Taking a practice lens on the delivery of DSIs guided us to focus on the full life cycle of DSIs. This required deep engagement in the field, observing and interacting with decision-makers, business stakeholders, program managers and delivery team members. As a result, I chose to study DSI delivery within a single organisation (Transport), where I am employed full-time, and to set up a DSI delivery capability while delivering those DSIs. This gave me access to data to conduct the case studies. To obtain granularity of the program life cycle as well as variation for analytical comparisons, I used an embedded case design (Yin, 2018) to track the unfolding of six DSIs in Transport, each of which provided a unique scope and opportunity to build DSI delivery capability. The six DSIs provided

me with an opportunity to use an embedded case study research method covering all three purposes – exploratory, descriptive and explanatory (Scholz & Tietje, 2002; Yin, 2018). My interest was to understand DSI delivery as experienced by the organisational participants themselves and to identify uniqueness in this portfolio of initiatives to bring about improvements within the organisation.

Iterating in-depth analysis of each case, comparisons across cases and connections to the literature (Eisenhardt, 1989), I reviewed the entire ecosystem of stakeholders and how they influenced the delivery of DSIs, all of which led me to further analysis and theorising (Agar, 1986). I used a variety of evidence, including documents, artefacts and participant observations from each DSI. Although the research design was aimed initially at understanding the unique characteristics of DSIs, the perceived success and failure patterns of DSIs led us to review the business cases themselves. Consistent with inductive research approaches, my research question emerged over time, as I engaged iteratively with the evidence from the field and with the extant research, which helped me make sense of what I found.

I was the program manager of the six DSIs chosen as case studies and these were delivered between January 2017 and June 2021, bringing with them in-depth insights into the program life cycle.

6.4.2 Research Setting

My research was situated within Operational Systems division of Transport, a state government enterprise that leads the development of safe, integrated and efficient transport systems for the people of New South Wales in Australia. Transport's functions include transport planning, strategy, policy, procurement and other non-service delivery functions across all modes of transport: roads, rail, ferries, light rail, metro and point-to-point. At the time of this research, Transport was in the early stages of executing \$41.5 billion worth of Future Transport 2056 SIPs, which set out more than 300 initiatives to be delivered in the first 10 years of the 40-year vision, all underpinned by a data-driven technology roadmap (Transport for NSW, 2021). The majority of the initiatives had (and have) a significant technology component, including data and data analytics. As I was a Program Manager at Transport, I had full access to the artefacts of the DSIs used as a basis for the research.

Transport generates significant amount of data every day from its bus, ferry, light rail, metro and heavy rail modes. The real-time data includes timetable information, the position of every transport vehicle and the predicted time of arrival at the next transit stop, which is shared on all passenger apps. The data also includes event information from sensors, such as doors opening and closing information, temperature, speed and number of passengers; ticketing information such as Opal cards being tapped on and off at the gates and checking of tickets through to crime and incident information on the transport network. While the realisation that the data for improving customer service and managing performance of transport operators was always there, the lack of technology, skills and investment prevented Transport from mining and monetising the data it was generating. This realisation led to the Future Transport Strategy 2056 driving a data ecosystem within Transport to provide continuous improvements on its asset performance and improved customer and operational information. My journey mirrors that of Transport as an organisation. When I joined Transport in late 2016 as a program manager, I had significant experience in delivering ICT transformation programs using both Waterfall and Agile methods but none in delivering data-related programs. This chapter therefore tracks the delivery of the DSIs, the challenges I faced in understanding some of their unique characteristics, my ability to influence near real-time data-driven decision-making in Transport and the growth in maturity of Transport to harness the value of data underpinned by my own ability to deliver DSIs.

The research method used the participant-observation technique and multiple case studies over the full program life cycle covering a period of four years. Out of the six potential sources of data, namely, documentation, archival records, interviews, direct observations, participant-observation and physical artefacts, all except interviews were used.

The six DSIs chosen as case studies represent contemporary phenomenon at Transport (Yin, 2018) that was particularly useful for our research question because the organisation needs to better understand the business cases and delivery of DSIs and to deliver them consistently. Table 4-1 is a summary of the six DSIs and their characteristics.

This chapter organises the combined case studies into three stages: exploration, takeoff and maturity. These are the stages that Transport itself went through in the delivery of the six DSIs. The exploration stage is mapped to the Vanguard and CTABS case studies; the take-off stage is mapped to the Ferry, Light Rail Priority and PTIPS Analytics case studies, and the maturity stage

is mapped to the MPR case study. *Figure 4-1* shows the timeline and highlights of the six DSIs, indicating my journey from uncertainty and frustration at not being able to deliver the program outcomes as per the schedule to accepting the exploratory nature of DSIs through to planning for the uncertainty and engaging the stakeholders effectively. While each of the six DSIs was unique, this chapter focuses on the first (Vanguard) and the sixth (MPR) as they represent the boundary conditions of the story presented here. In other words, I present the details of one case at the exploration stage of the DSI experience and one at the maturity stage of the DSI experience.

6.5 DATA COLLECTION AND ANALYSIS

All six of the DSIs at Transport that I managed were used to collect data. Once they were all delivered they made both real-time and retrospective data collection possible. While the scale of the DSIs was different, together they paint a good picture of unique characteristics, how their governance was structured and how their business cases were written. Table 4-1 provides a summary of the six DSIs and Table 5-4 shows the gaps and issues identified across the three program phases.

6.6 DSI DELIVERY FRAMEWORK

In this section I review the design principles used to build the framework and its processes.

6.6.1 The Design Principles

Considering the exploratory nature of DSIs, the framework conformed to the following design principles:

- End-to-end delivery of solution and value realisation,
- Core and non-core domains identification,
- Use of agile methods instead of waterfall to support the exploratory nature of DSIs,
- Support both single team and scaled agile delivery of DSIs,
- Specify people, process and deliverables, and
- Agnostic regarding tools and technologies.

6.6.2 The Framework

The proposed framework had, and being still active, continues to have, five core domains and integrates the PMI's Standard for Program Management (Project Management Institute, 2017b) for program management, the Proscii Framework (Hiatt, 2006) for people change management, Scaled Agile (SAFe) (Scaled Agile, 2023) for solution delivery, DAMA's DMBOK (Earley, 2017) for data management and CRISP-DM (Chapman et al., 2000) for data science processes, as shown in *Figure 6-3*.

As the domains are modular, they allow organisations to replace methods. For example, in the program management domain, the PMI methods (Project Management Institute, 2017b) could be replaced with MSP (Axelos, 2022). Further, the framework is flexible to allow integration with other organisational domains, such as risk management, procurement management and asset management.

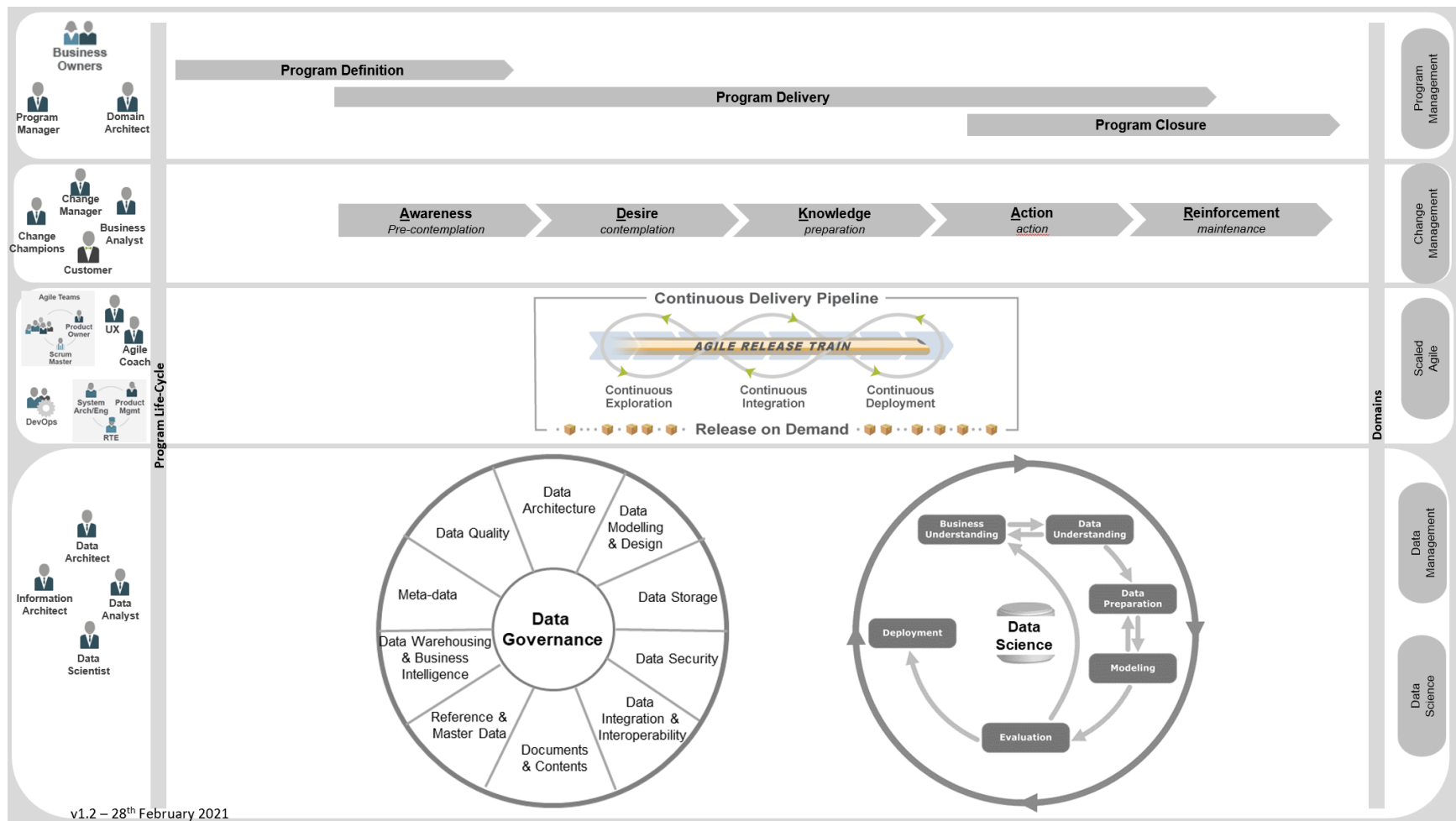


Figure 6-3. DSI Delivery Framework

6.6.3 The Processes

The first domain of the DSI delivery framework is program management. As stated previously, I used processes in the PMI's Standard for Program Management (Project Management Institute, 2017b) as the basis of this domain. An overview of the three phases is given in Table 6-2:

Table 6-2
DSI Delivery Framework – Program Management Phases

Phase	Description
1. Program Definition	This phase consists of program activities conducted to authorise the DSI program and develop the program roadmap required to achieve the expected results. As part of the program definition, the program business case and program charter are formulated. Once approved, the program management plan is prepared.
2. Program Delivery	Program delivery comprises the program activities performed to produce the intended results of each component in accordance with the program management plan. Throughout this phase, individual DSI components are initiated, planned, executed, transitioned and closed, while benefits are delivered, transitioned and sustained. The initial components deliver the foundational capability of the big-data platform and bring in core reference and master data sets. The later components build on existing datasets to deliver advanced capability.
3. Program Closure	This phase includes the program activities necessary to transition the program benefits to the sustaining organisation and formally close the program in a controlled manner. During program closure, the program is transitioned and closed or terminated early, or work is transitioned to another program.

The second domain of the DSI delivery framework is change management. Successful change management helps in the effective use of the capability delivered by the DSI and thus in value realisation of the DSI investment. I used processes in the ADKAR change management model (Hiatt, 2006) as the basis of this domain. The ADKAR model is prescriptive and goal-oriented and each milestone must be achieved to define success. An overview of the five steps of ADKAR is in Table 6-3:

Table 6-3
DSI Delivery Framework – Change Management Steps

Steps	Description
1. <u>A</u> wareness	Awareness of the need for change
2. <u>D</u> esire	Desire to participate and support in the change
3. <u>K</u> nowledge	Knowledge of what to do during and after the change
4. <u>A</u> bility	Ability to realise or implement the change as required
5. <u>R</u> einforcement	Reinforcement to ensure the results of a change continue

The third domain of the DSI delivery framework is Scaled Agile. As per the literature review, I saw there were several methods that could potentially be used for DSI delivery. I opted for SAFe (Scaled Agile, 2023) due its prescriptiveness and ability to scale in large geographically dispersed organisations. SAFe had also been chosen by Transport as the framework to deliver business agility. This chapter covers some of the key SAFe processes relevant to the program delivery phase and these are defined in Table 6-4:

Table 6-4
DSI Delivery Framework – Scaled Agile Processes

Process	Description
1. Planning Interval (PI) Planning	Planning Interval (PI) Planning is a cadence-based event that serves as the heartbeat of the Agile Release Train (ART), aligning all the teams on the ART to a shared mission and vision. Held over two days, PI Planning has a standard agenda that includes a presentation of business context and vision, followed by team planning breakouts—where the teams create their iteration plans and objectives for the upcoming PI. Facilitated by the Release Train Engineer (RTE), this event includes all members of the ART and occurs within the Innovation and Planning (IP) iteration.
2. Continuous Exploration	Continuous Exploration (CE) is the process that drives innovation and fosters alignment on what should be built by continually exploring market and customer needs and defining a vision, roadmap and set of features for a solution that addresses those needs. CE is the first aspect in the four-part Continuous Delivery Pipeline of Continuous Exploration (CE), Continuous Integration (CI), Continuous Deployment and Release on Demand (<i>Figure 6-3</i>). CE is integral to that process and focuses on applying design thinking to understand and build alignment on new opportunities and what needs to be built, while recognising that all such ideas are hypotheses that need to be

validated. CE replaces traditional waterfall approaches of upfront, rigid definitions of requirements with a process that generates a consistent flow of work that is ready for the ART to implement. This flow of work is captured as features and stories, defined in small batches that can travel easily through the remaining aspects of the continuous delivery pipeline to the customer. Feedback is built into the process, enabling teams to adjust to market needs.

3. Continuous
Integration

Continuous Integration (CI) is the process of taking features from the Program Backlog and developing, testing, integrating and validating them in a staging environment where they are ready for deployment and release. CI is the second aspect in the four-part Continuous Delivery Pipeline of Continuous Exploration (CE), Continuous Integration (CI), Continuous Deployment and Release on Demand (*Figure 6-3*).

CI is a critical technical practice for each ART. It improves quality, reduces risk and establishes a fast, reliable and sustainable development pace. With CI, the “system always runs,” meaning it’s potentially deployable, even during development. CI is most easily applied to software solutions where small, tested vertical threads can deliver value independently. In larger, multiplatform software systems, the challenge is harder. Each platform has technical constructs and the platforms must be continuously integrated to prove new functionality. In complex systems comprised of software, hardware and components and services provided by suppliers, CI is harder still. But the fact remains: integrating and testing components together frequently is the only practical way to fully validate a solution.

4. Continuous
Deployment

Continuous Deployment (CD) is the process that takes validated features in a staging environment and deploys them into the production environment, where they are readied for release. CD is the third aspect in the four-part Continuous Delivery Pipeline of Continuous Exploration (CE), Continuous Integration (CI), Continuous Deployment and Release on Demand (*Figure 6-3*).

The ability to release on demand is a critical competency for each ART and Solution Train. It allows businesses to respond to market opportunities with the highest-value solutions in the shortest sustainable lead times and at a rate that permits customers to absorb the new functionality. To support a business that wants to release on demand, features must be waiting and verified in production before the business needs them. Therefore, it is optimal to separate the deployment process from the release process so that deployed changes move into the production environment in a manner that does not affect the behaviour of the current system. This gives the teams the ability to make

smaller, incremental changes, which can be deployed to production continually but are not released to end users until the time is right.

- | | |
|----------------------|--|
| 5. Release on Demand | <p>Release on Demand is the process that deploys new functionality into production and releases it immediately or incrementally to customers based on demand. Release on Demand is the final aspect in the four-part Continuous Delivery Pipeline of Continuous Exploration (CE), Continuous Integration (CI), Continuous Deployment and Release on Demand.</p> <p>The three aspects that precede Release on Demand help ensure that new functionality is continuously readied and verified in the production environment. But since tangible development value only occurs when end users are operating the solution in their environment, releasing that value at the right time is critical for the enterprise to gain the real benefits of agility. The decision as to when and what to release is a crucial economic driver that requires careful consideration. For many, continuous delivery is the desired end state, allowing new functionality to be released as soon as it is developed. But more often the release is a decoupled, on-demand activity, occurring for specific users, timed for when they need it, or when it makes the most economic sense for the enterprise.</p> |
| 6. System Demo | <p>The system demo is a significant event that provides an integrated view of new features for the most recent iteration delivered by all the teams in the ART. Each demo gives ART stakeholders an objective measure of progress during a PI. A system demo is a critical event. It's the method for assessing the solution's current state and gathering immediate, ART-level feedback from the people doing the work, as well as critical feedback from Business Owners, sponsors, stakeholders and customers. The demo is the one real measure of value, velocity and progress of the fully integrated work across all the teams.</p> |
-

The fourth domain of the DSI delivery framework is data management. Every DSI will set up and build on an organisation's data management capability. Our data management process used the 11 knowledge areas from DAMA's DMBok (Earley, 2017), as shown in *Figure 2-17*. An overview of the 11 knowledge areas is shown in Table 6-5:

Table 6-5
DSI Delivery Framework – Data Management Processes

Knowledge Area	Description
1. Data Governance	The exercise of authority, control and shared decision-making (planning, monitoring and enforcement) over the management of data assets.
2. Data Architecture	Identifying the data needs of the enterprise (regardless of structure) and designing and maintaining the master blueprints to meet those needs. Using master blueprints to guide data integration, control data assets and align data investments with business strategy.
3. Data Modelling & Design	Data modelling is the process of discovering, analysing and scoping data requirements and then representing and communicating these data requirements in a precise form called data model. This process is iterative and may include a conceptual, logical and physical model.
4. Data Storage & Operations	The design, implementation and support of stored data to maximise its value.
5. Data Security	Definition, planning, development and execution of security policies and procedures to provide proper authentication, authorisation, access and auditing of data and information assets.
6. Data Integration & Interoperability	Managing the movement and consolidation of data within and between applications and organisations.
7. Document & Content Management	Planning, implementation and control activities for lifecycle management of data and information found in any form or medium.
8. Reference & Master Data	Managing shared data to meet organisational goals, reduce risks associated with data redundancy, ensure higher quality and reduce the costs of data integration.
9. Data Warehousing & Business Intelligence	Planning, implementation and control processes to provide decision support data and support knowledge workers engaged in reporting, query and analysis.
10. Metadata Management	Planning, implementation and control activities to enable access to high quality integrated metadata.
11. Data Quality	The planning, implementation and control of activities that apply quality management techniques to data to assure it is fit for consumption and meets the needs of data consumers.

At the core of DSI delivery framework is the data science processes which is the fifth domain. The data science process uses the six steps as shown in *Figure 2-19* from CRISP-DM (Chapman et al., 2000) and has been updated from original data mining purpose to suit DSIs. While originally conceived in 1996, CRISP-DM model continues to be used and adapted by several organisations due to its simplicity and relevance. One of the significant features of this model is being iterative and steps two to six are continuously repeated as part of the product life cycle. An overview of the six steps is in Table 6-6:

Table 6-6
DSI Delivery Framework – Data Science Processes

Process	Description
Program Definition Phase	
1. Business Understanding	This initial phase focuses on understanding the DSI objectives and requirements from a business perspective, then converting this knowledge into a problem definition and a preliminary plan designed to achieve the objectives.
Program Delivery Phase	
2. Data Understanding	The data understanding phase starts with an initial data collection and proceeds with activities to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.
3. Data Preparation	The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modelling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modelling tools.
4. Modelling	This phase concentrates on predictive or descriptive model development and by using the first version of the prepared dataset as a training set (historical data). The modelling process is extremely iterative as it provides intermediate insights and reputable refinement of data preparation and model specification. It is significantly helpful to try several algorithms with specific parameters to find the ideal model.
5. Evaluation	During model development and before deployment, the model needs to be evaluated to understand its quality and ensure that it properly and fully addresses the business problem. Model evaluation entails computing various diagnostic measures and other outputs such as tables and graphs, enabling interpretation of the model's quality and its efficacy in solving the problem. For

a predictive model, a testing set is used, which is independent of the training set but follows the same probability distribution and has a known outcome. The testing set is used to evaluate the model so it can be refined as needed.

6. Deployment This phase productionises the solution such that “live” models are available for organisation’s decision-making processes and the source data is continuously ingested into production environment. The deployment phase can be as simple as generating a report or as complex as implementing a repeatable data science process across the enterprise.

A summary of the DSI delivery processes across the program phases is shown in Table 6-7.

Table 6-7

DSI Delivery Framework – Summary of key processes across five domains and program phases

Domain	Program Phases		
	Definition	Delivery	Closure
1. Program Management	Program Planning		Program Closure
2. Change Management	Change Management Planning	System Demo (Showcase)	
3. Scaled Agile		- PI Planning - Continuous Exploration - Continuous Integration - Continuous Deployment - Release on Demand	
4. Data Management	Data Management Planning	- Implement Data Governance - Establish Data Architecture - Data Modelling & Design - Manage Data Storage & Operations - Implement Data Security - Implement Data Integration & Interoperability - Implement Document & Content Management - Publish Reference & Master	

			Data
			• Implement Data Warehousing & Business Intelligence
			• Implement Metadata Management
			• Implement Data Quality
5. Data Science	• Business Understanding		• Data Understanding • Data Preparation • Modelling • Evaluation • Deployment

6.7 CONCLUSION AND RECOMMENDATION

6.7.1 Conclusion

In this section, I review my research question: “*What design principles should be incorporated in a data science initiative (DSI) delivery framework so that program managers can adopt a predictive path to realise value from such investments?*” and summarise our conclusion. I first start with the limitations of our research and then discuss the implications of the proposed DSI delivery framework for business and program managers. I conclude that the current literature does not cover the delivery of DSIs adequately and that the proposed framework is a step in the right direction to assist practitioners.

6.7.2 Limitations and Implications of the Research

This research used six DSIs from one public sector organisation as a case study to develop a DSI delivery framework using five domains. Future research can include validating the framework with other public and private sector organisations delivering DSIs. Another aspect is that DSIs are a more recent phenomenon and sit in a rapidly evolving technology and delivery space. This has an impact on the currency of my research.

The limited availability of methods and standards in DSI delivery has caused business managers and program managers to chart their own path and thus introduce inconsistency in how

DSIs are treated and delivered in different organisations. With the emergence of research such as this, it can be expected that standardisation of DSIs will increase and provide guidance to practitioners in terms of more efficient delivery.

6.7.3 Recommendations

Program management for ICT-enabled programs has rich literature and proven delivery frameworks, which have matured over the past three decades (Axelos, 2022; Project Management Institute, 2016, 2017b). This chapter makes a significant contribution to the theory and practice of the emerging field of data science.

The current program management literature does not adequately support DSI delivery, focusing instead on risk elimination and rapid delivery of business outcomes. I propose a DSI delivery framework that has five core domains and integrates the following: the PMI Standard for Program Management (Project Management Institute, 2017b) for program management; the Prosci framework (Hiatt, 2006) for people change management; Scaled Agile (SAFe) (Scaled Agile, 2023) for solution delivery, DAMA's DMBOK (Earley, 2017) for data management and CRISP-DM (Chapman et al., 2000) for data science processes, as shown in *Figure 6-3*.

I suggest additional research to fine-tune the proposed DSI delivery framework which currently has been used for only one public sector organisation (Transport). I validated the reliability of the framework through monitoring its use at Transport as well as through semi-structured interviews with Transport stakeholders. The framework proposed in this research will deliver a significant contribution to the body of knowledge for program management relevant to both literature and practitioners. If a well-designed framework is not used, it could lead to more failed programs, dissatisfied sponsors and delay much-needed investment in this emerging field, as well as delay the benefits that will flow from harnessing the data and the nuggets in it.

7. Minimum Viable Governance for DSIs

This chapter is based on my published paper “Minimum Viable Governance for Data Science Initiatives - A Transport for NSW Case Study” (Mathur et al., 2022).

7.1 ABSTRACT

Too much governance in organisations can stifle innovation. Too little governance can waste precious resources. Business agility demands empowerment of people to take decisions on initiatives designed to deliver innovative products and services. Traditional monthly and quarterly governance forums such as steering committees and program boards for decision-making potentially impede the flow of work when the delivery of a program or project is done using agile methods in two-weekly sprints and decisions are required at a different and more frequent speed. DSIs, which are exploratory and innovative in nature, follow agile delivery methods to enable faster delivery.

This chapter describes an exploratory study of implementing governance for DSIs based on a single case study. It investigates agile governance for DSIs at project, program and portfolio level and suggests eight guiding principles focusing on product and portfolio governance. It is targeted at practitioners to guide them in setting the minimum viable level of governance to ensure value is realised from their DSIs and at academics to advance research in the governance of DSIs.

7.2 INTRODUCTION

The year 2020 saw one of the biggest disruptions experienced by mankind in recent history due to the impacts of the Covid-19 pandemic on businesses and individuals globally. Businesses that were able to adapt to a new world order survived while those who failed perished. The largest companies in the world, such as Google, Amazon and Facebook, which are enterprises fuelled by data science, used data to improve customer satisfaction and maximise profits and continued to thrive despite the pandemic. Individuals who traditionally worked in the offices of their

organisations had to adapt working from home. However, the need for business agility and data-driven decisioning did not slow down.

DSIs⁹ have emerged as a popular mechanism in several organisations for extracting value from data. Being exploratory in nature, DSIs follow agile delivery methods. They have unique characteristics which separate them from the way in which a typical ICT-enabled program is conceptualised and managed. However, the track record of these initiatives has drawn substantial criticism from sponsors. For example, the success rate of DSIs is not perceived to be high, with Gartner estimating that 85% of big data projects fail (Asay, 2017), A very high percentage (87%) of DSIs never make it to production (VentureBeat, 2019) and through 2022, only 20% of analytic insights delivered business outcomes (Gartner, 2019). Becker (2017) identified 17 areas causing the high rate of failure in DSIs, nine of them attributed to project management or organisational issues, such as wrong/inadequate skills, incorrect business objectives, insufficient ROI/business case, improper scope, management and cultural resistance, inadequate management and governance, incorrect project structure, poor communication and problem avoidance.

In most organisations portfolio planning and budget cycles are annual; however, this does not allow an organisation to adapt to market changes and digital disruption that require a more dynamic approach. The problem is compounded when agile delivery methods are used, especially for DSIs, where scope and design decisions are required in a typical two-week sprint cadence. Following the Goldilocks principle of being just right, governance that is neither too light nor too overbearing is recommended (Knapp et al., 2016; Mullaly, 2009; Mullaly, 2010; Williams, 2017).

This chapter explores ways to set up the minimum level of viable governance to ensure that the expected value is realised from DSIs. It investigates the governance of DSIs at portfolio, program and project level and draws upon experiences from Transport for NSW (Transport) as a case study. I work in the Active Transport branch of Transport and have implemented the Scaled Agile Framework (SAFe) (Scaled Agile, 2023) to manage the \$950 million Active Transport projects portfolio. This portfolio is transforming from a project mindset to a product mindset for

⁹ Data Science Initiatives (DSIs) are defined as related projects or programs that involve the application of data science techniques and methods to address complex problems, generate insights, or create new products or services.

both digital and engineering initiatives in order to improve the flow of work (Kersten, 2018). This entails funding value streams, not projects, and thus keeping teams permanent.

This chapter describes research that I believe can be useful to academics for further studies into the governance of agile projects of an exploratory nature and to practitioners of portfolio, program and project management who are involved in decision-making about and delivery of such initiatives.

Building on the case study findings, I argue that practitioners need to understand the portfolio and product governance requirements when planning and delivering DSIs and move away from traditional approaches which fail to account for the nuances this class of programs brings to the field, especially in handling the uncertainty and ambiguity that continues during a DSI life cycle. I also introduce the term “minimum viable governance” or MVG for DSIs to ensure that they have a balanced governance strategy from ideation through to delivery and benefits realisation to support their exploratory nature.

The research presented in this paper is informed by my firsthand experiences being actively involved in the governance and management of data science initiatives (DSIs) within Transport. I held Program Manager and Director/Sponsor roles to spearhead efforts to enhance the effectiveness and efficiency of DSIs through strategic governance interventions.

The chapter is organised as follows: in the next section, I discuss the literature on portfolio, program and project management and product versus project governance. I then discuss the research setting and methods used for this investigation. Following this, as part of data collection and analysis, I introduce an Active Transport case study that covers the challenges, solutions implemented and business outcomes delivered. The final section highlights the findings, the implications for practice and recommendations.

7.3 LITERATURE REVIEW

The focus of this research was to explore governance for DSIs with the target audience being program managers of technology initiatives, portfolio managers/policy makers approving business cases and establishing governance mechanisms and academics researching program management and governance. Some of the key issues motivating this research were:

- Why 85% of big-data projects fail (Asay, 2017) when 73% projects (overall) meet their original goals (Project Management Institute, 2021a),
- How DSIs can be delivered effectively, and
- An appropriate level of governance for DSIs that allows exploration and does not stifle innovation.

Any investment should follow established the governance structures of an organisation, governance principles that usually flow down through the organisation to influence the governance objectives at portfolio, program, project and operational levels and help create the value system of an organisation (Project Management Institute, 2017a). Governance activities ensure that management activities are defined, planned and implemented within portfolio, program and project management (Project Management Institute, 2016). Further, management activities are more operational and tactical than governance activities, which are more strategic and focus on oversight and guidance. Significant literature already exists covering the corporate, portfolio, program and product governance hierarchy. However, there is a gap in the literature in relation to the governance of DSIs, especially the lack of acknowledgment of the uncertainty, complexity and exploratory nature of DSIs which so influences their management and governance.

I start the review with a discussion on corporate, portfolio and program governance. As DSIs often follow agile methods for delivery, I review agile governance. As I see DSIs as a continuous project delivery, I apply program management and program governance to them and specifically exclude project governance from this review. All DSI-related technology, tools and software delivery processes are also excluded from this review. Finally, I discuss the knowledge gap that exists in the literature in relation to DSIs, which in turn influences their governance and management.

7.3.1 Corporate, Portfolio and Program Governance

Corporate governance deals with the mechanisms by which stakeholders of a corporation exercise control over corporate insiders and management such that their interests are protected (John & Senbet, 1998). These authors define stakeholders of a corporation as equity holders, creditors and other claimants who supply capital, employees, consumers, suppliers and the government. Claessens (2006) defines corporate governance as the rules under which firms

operate, with the rules coming from such sources as the legal system, financial markets and factor (labour) markets and including corporate social responsibility and sustainability. Müller, Turner, Andersen, Shao, and Kvalnes (2016) categorise corporate governance into (1) control systems for directing and controlling organisations, (2) processes to allow corporations to be responsive to their stakeholders and (3) relationships among internal and external participants of the firm. Tricker and Tricker (2015) specify 10 core principles of corporate governance and state that good corporate governance should be integrated into the company's business strategy and not viewed as simply a compliance obligation. These principles are:

1. Lay solid foundations for management and oversight,
2. Structure the board to add value,
3. Promote ethical and responsible decision-making,
4. Safeguard integrity in financial reporting,
5. Make timely and balanced disclosure,
6. Respect the rights of shareholders,
7. Recognise and manage risk,
8. Encourage enhanced performance,
9. Remunerate fairly and responsibly and
10. Recognise the legitimate interests of stakeholders.

I expect all DSIs to operate within the boundaries of corporate governance, especially in terms of financial delegation of authority.

Portfolio management deals with the coordination and control of multiple projects pursuing the same strategic goals and competing for the same resources, such that managers prioritise among projects to achieve strategic benefits (Cooper, Edgett, & Kleinschmidt, 1997). Martinsuo (2013) views portfolio management as negotiation, bargaining and structural reconfiguration, in addition to rational decision processes to respond to uncertainties and complexities in business environments. The Project Management Institute (2017a) describes portfolio management as centralised management of one or more portfolios to achieve strategic objectives. Portfolio

management is also a dynamic activity through which an organisation invests its resources to achieve its strategic objectives by identifying, categorising, monitoring, evaluating, integrating, selecting, prioritising, optimising, balancing, authorising, controlling and terminating portfolio components. Jenner and Kilford (2011) define portfolio management as a coordinated collection of strategic processes and decisions that together enable the most effective balance of organisational change and BAU. They describe the portfolio management cycles as a series of broadly sequential practices (understand, categorise, prioritise, balance and plan) and portfolio delivery containing practices (management control, benefits management, financial management, risk management, stakeholder engagement, organisational governance and resource management) undertaken broadly simultaneously. Portfolio governance is a set of practices, functions and processes within a framework based on a set of principles that are the fundamental norms, rules or values that guide portfolio management activities in order to optimise investments and meet organisational strategic and operational goals (Project Management Institute, 2017a). Effective portfolio governance provides clarity about what decisions are made, where and by whom and what criteria are used in researching these decisions and provide alignment to the wider organisational governance structure (Jenner & Kilford, 2011). I see portfolio governance as a mechanism for doing the right DSIs and rendering higher value for money to the organisation rather than just doing the DSIs correctly.

Program management concerns the harmonised management of a number of projects and other actions that will generate competitive advantage (Thiry, 2016). Artto, Martinsuo, Gemünden, and Murtoaro (2009) define program management as the coordinated organisation, direction and implementation of a portfolio of projects and activities that together achieve outcomes and realise benefits that are of strategic importance. Similarly, program management is defined as the application of knowledge, skills and principles to a program to achieve the program objectives and to obtain benefits and control not available by managing program components individually (Project Management Institute, 2017b). MSP breaks down large, complex change into manageable, interrelated projects based on three core concepts: MSP Principles, MSP Governance Themes and MSP Transformational Flow (Axelos, 2022). Program governance comprises the framework, functions and processes by which a program is monitored, managed and supported in order to meet organisational strategic and operational goals. Its focus is on delivering the program benefits by establishing the systems and methods by which that program and its strategy are defined,

authorised, monitored and supported by its sponsoring organisation (Project Management Institute, 2017b). In a traditional view of the governance hierarchy, component projects follow program governance and the portfolio provides the governance policies, oversight, control, integration and decision-making functions and processes to programs within the portfolio. Key program governance roles are program sponsor, program steering committee, program management office, program manager, project manager(s) and other stakeholders. Axelos (2022) defines nine governance themes that need to be considered when delivering a program of transformational change – program organisation, vision, leadership and stakeholder engagement, benefits management, blueprint design and delivery, planning and control, business case, risk and issue management and quality and assurance management. I discuss program governance but not project governance, as DSIs are not delivered as a one-off project and generally are part of a larger initiative.

Müller (2009, 2016) identified four models of organisational project governance – process models, governance and governmentality-based models, nested models and layered models, which cover agency, institutional, stakeholder and transaction cost theories. He then described the model of four-governance paradigms linking corporate governance with organisational project governance through an overlay of two dimensions – corporate governance orientation from shareholder to stakeholder and the control structures in a project's parent organisation from a behaviour-to-outcome perspective. The conformist paradigm assumes efficiency is maximised by trusting process over the capabilities of individuals; the Flexible Economist paradigm aims for a flexible application of the most effective project management methods, tools, techniques and management approaches often through a PMO; the Versatile Artist paradigm is characterised by organisations which employ senior and experienced project managers, who develop their methodologies and work practices in accordance with project needs and the Agile Pragmatist paradigm emphasises process compliance with a time-phased delivery of functionalities.

7.3.2 Agile Governance

Businesses require flexible and customisable technology environments and effective and responsive governance in order to deliver value faster, better and more cheaply (Luna, Kruchten, Pedrosa, Neto, & de Moura, 2014). The demand for agility imposes challenges on portfolio management and governance, as best-practice frameworks for portfolio management (Jenner &

Kilford, 2011; Project Management Institute, 2017a) or IT governance (Agutter, 2020; ISACA, 2018) are more suited to stable environments where traditional command-and-control settings (Horlach et al., 2019) work well. Traditional project governance focuses on successful delivery of the iron triangle – time, cost and quality (Pollack et al., 2018) – that is used in measuring success. In the case of innovation and exploratory projects that generally follow agile delivery methods, the scope and the goals are difficult to determine upfront and rather evolve in a continuous and iterative process, also referred to as “progressive elaboration” (Collyer & Warren, 2009). This is because the requirements of such projects tend to be ambiguous and change due to their uncertainty, which in turn leads to agile teams using four different categories of iteration objectives – functionality, schedule, quality and team satisfaction (Drury-Grogan, 2014). Lappi et al. (2018) compared the traditional and agile project governance practices, as shown in Table 7-1, to show the practices that exist in traditional governance that can be transferred to Agile, either as they are or with some modification. It can be seen from their suggestion that customer-centricity and team empowerment makes Agile governance more suited to the changing demands faced by the digital economy.

Table 7-1

Comparison of Traditional and Agile Project Governance Practices (Lappi et al., 2018, p. 55)

Traditional Practices That Can Be Transferred to Agile as Such	Traditional Practices That Can Be Modified to Agile	Emerging Agile-Specific Practices with No Regard to Traditional	Remaining Gaps Between Agile and Traditional Project Governance
Distinguished front-end phase for goal setting and business case	Scope setting is visionary and iterative process, starting with product vision and user stories	Inclusion of customer within the project team for decision making and validation efficiency	Linking and aligning the agile project objectives with organizations' strategy
Long-term post-project benefit realization	Short-term value realization through utilization of deliverables	Using validated—tested and accepted—deliverables as project performance and incentives basis	Connection between formal and flexible budgeting
Clear organizational and ownership structure	Flexible, value-based project budgeting and contracting	Change management is embedded into iterative planning and delivery	Impact of collaborative project models to agile project governance
External audits by third parties	Motivation through peer and team performance-related incentives	Agile philosophy and principles as project culture foundation	Conflict escalation and resolution within agile project
Project stakeholder management	Visual, real-time project monitoring through agile specific and traditional measures	Project team empowerment for autonomy, flexibility, and decision-making power	Application and impact of formalized project governance structures such as steering groups, portfolios, and programs in agile context
Project risk management	Product and sprint backlogs are the main planning and communication means	Project team composed of multitalented, self-organized, and cross-functional individuals, both experts and beginners	Project human resource and capacity management in agile
Necessary documentation for external validation and communication	Onsite, informal, and real-time communication (internal)	Agile mindset and technical capabilities as competence requirements	
Project capability management on organizational level through HR and PMOs	Project management daily activities are shared on team level	Project manager role changes to administrative and facilitative	
	Sprint retrospectives used as interproject learning method	Agile coach and scrum master manage agile project team performance	

In agile projects, the focus of coordination and communication is to provide sufficient real-time information about the project's status to the empowered project team and customer representatives visually and in real time. To enable this, agile projects prefer product vision and backlogs instead of a formal, rigid project plan (Sheffield & Lemétayer, 2013).

Both the Project Management Institute (2021b) and the Association of Project Management (2016) provide 12 principles on agile governance (as shown in Table 7-2) and a foundation for DSI governance.

Table 7-2
Guiding Principles for Agile Governance

Principles of Project Management (Project Management Institute, 2021b)	Principles For Governance Of Agile Change (Association of Project Management, 2016)
1. Be a diligent, respectful and caring steward	1. Focus on the business need
2. Create a collaborative project team environment	2. Value driven
3. Effectively engage with stakeholders	3. Incremental delivery
4. Focus on value	4. Timebox delivery

5. Recognise, evaluate and respond to system interactions	5. Empowered teams and decision-making
6. Demonstrate leadership behaviours	6. Collaboration
7. Tailor based on context	7. Enhanced communication
8. Build quality into processes and deliverables	8. Just enough definition
9. Navigate complexity	9. Constant striving for improvement
10. Optimise risk responses	10. 'Learn forward'
11. Embrace adaptability and resiliency	11. Demonstrate control
12. Enable change to achieve the envisioned future state	12. Change control

Scaled Agile frameworks that use a lean-agile approach have been found to be more suited than traditional approaches to portfolio management in a global economy due to the impact of digital disruption. A lean-agile approach recommends organising people around value streams, funding value streams, having lean budgets and guardrails, adjusting value stream budgets dynamically, decentralising the intake of work by value streams, constructing lean business cases with minimum viable product and having milestones based on working solutions (Scaled Agile, 2023). This is compared to the more traditional approach of organising people in functional silos, funding projects, using project-cost accounting, doing big upfront/top-down planning, centralising work intake, creating detailed business cases based on speculative ROI and governing projects by phase gates and waterfall milestones.

SAFe (Scaled Agile, 2023) is based on 10 underlying lean-agile principles which shape the governance and practices of SAFe. These principles are:

- take an economic view,
- apply systems thinking,
- assume variability, preserve options,
- build incrementally with fast, integrated learning cycles,
- base milestones on objective evaluation of working systems,
- visualise and limit WIP, reduce batch sizes and manage queue lengths,
- apply cadence, synchronise with cross-domain planning,

- unlock intrinsic motivation of knowledge workers,
- decentralise decision-making, and
- organise around value.

In SAFe, the portfolio aligns strategy with execution via a collection of development value streams which focus on building the right things with the appropriate level of investments in solutions for the portfolio to meet its strategic objectives. Key concepts used in SAFe portfolio management are:

- Value streams – Every value stream must fund the people and resources necessary to build Solutions that deliver value to the internal or external customer,
- Lean budgets – Lean budgeting allows fast and empowered decision-making, with appropriate financial control and accountability through guardrails,
- Portfolio Kanban – The portfolio Kanban system makes the WIP limits to help assure that demand is matched to the actual capacity of value streams and ARTs,
- Portfolio Vision – The portfolio vision is a description of the future state of a portfolio's value streams and solutions and describes how they will cooperate to achieve the portfolio's objectives and the broader aim of the enterprise, and
- Portfolio Canvas – The portfolio canvas provides critical inputs to the portfolio vision, portfolio backlog and lean budgets.

SAFe establishes guardrails to ensure that the mix of investments addresses both near-term opportunities and long-term strategy, that investments in technology, infrastructure and maintenance are not regularly ignored and that large investments are approved appropriately. SAFe lean portfolio management relies on three significant events—strategic portfolio review, portfolio sync and participatory budgeting—to ensure the portfolio is well managed. SAFe has program and solution backlogs as repositories for all the upcoming work that affect the behaviour of the solution. The backlogs are a short-term holding area for features and capabilities that have gone through their respective Kanban systems and have been approved for implementation. SAFe has adopted the WSJF using the cost-of-delay (CoD) concept (Reinertsen, 2009) as a method of prioritising. CoD is calculated as the sum of user-business value, time criticality and risk reduction

and/or opportunity enablement. WSJF makes a relative estimate by concentrating on one item at a time, evaluating the item with the smallest value as 1 and assigning the remainder to the relative value. Relative values are the Fibonacci numbers used in planning poker. Finally, the WSJF value can calculate the CoD as the job size (job duration) and the item with the highest WSJF value has the highest priority. Active Transport is using WSJF as one of the lenses to prioritise initiatives covering engineering and DSIs.

For performance monitoring, SAFe has three measurement domains – outcomes, flow and competency. These can be applied at all levels of a SAFe enterprise to measure performance – SAFe Portfolio, Solution Train, Agile Release Train and Agile Team. SAFe further proposes minimum metrics, as in Table 7-3, to measure portfolio performance against strategy implementation and spending alignment with the agreed boundaries and continuous improvement.

Table 7-3

Lean Portfolio Metrics (Scaled Agile, 2023)

Goals	Measures	Desired Benefits
Employee Engagement	Employee surveys; Agile People Operations data	Improved employee satisfaction, higher employee engagement, and better business outcomes
Customer Satisfaction	Net Promoter Score (NPS)	Improved customer experiences and loyalty
Partner Health	Partner and vendor surveys	Improved ecosystem relationships
Business Agility	Business Agility self-assessment	Improved ability to respond to market changes, emerging opportunities and threats
Portfolio Performance	Objectives and Key Results (OKRs); LPM Self-Assessment	Better alignment and improved progress toward the portfolio's strategic themes, and portfolio performance
Value Stream Performance	Value Stream KPIs	Improved performance against the forecasted business outcomes
Program Predictability	Program predictability measure	Improved actual business value achieved
Time-to-Market	Number of releases; feature cycle time	More frequent releases; faster feature delivery
Relentless improvement	Self-assessments for each level of the Framework	Relentless improvement in team, program and portfolio performance
Quality	Defect count and cycle time; support call volume; escaped defects	Improved customer satisfaction; reduced support call volume, and lower product development costs

© Scaled Agile, Inc.

7.3.3 Summary of Literature Review

In the previous section, I investigated governance at all levels of an organisation, starting with corporate (Claessens, 2006; John & Senbet, 1998; Tricker & Tricker, 2015), then portfolio (Jenner & Kilford, 2011; Project Management Institute, 2017a) and program governance (Axelos, 2022; Project Management Institute, 2017b). I then explored agile governance (Agutter, 2020; Horlach et al., 2019; ISACA, 2018; Lappi et al., 2018; Pollack et al., 2018), which is relevant to our fast-changing world and expected by digital economy. I presented a detailed review of SAFe (Scaled Agile, 2023) as it has been implemented in the Active Transport branch of Transport in both engineering programs and DSIs. I identified a shift towards lean portfolio management as it creates a trade-off between the different value streams in the portfolio, a continuous investment flow replacing the annual planning and budget cycle, decentralised decision-making in execution, outcome-based governance and replacement of project funding by value stream funding.

In summary, significant literature exists covering the corporate, portfolio, program and product governance hierarchies. I note however, that a gap exists in the literature in relation to DSI governance, especially in the lack of acknowledgment of the uncertainty, complexity and exploratory nature which influences the management and governance of DSIs. This view of the literature and our need to balance over- and under-governance motivated us to ask the following research question:

“What is the minimal viable governance required for Data Science Initiatives to ensure they deliver envisaged value efficiently and effectively?”.

7.4 RESEARCH SETTING AND METHODS

7.4.1 Research Methods

The research focused on the governance of DSIs and how they interact with the organisation and the wider projects portfolio of which they are a part. Such a focus requires deep engagement in the field, observing and interacting with decision-makers, business stakeholders, program managers and delivery team members. As a result, I chose a multi-methods research methodology which is appropriate for exploratory research when the aim is to gain familiarity with a problem or to generate new insights for future research (Eisenhardt, 1989). For this chapter, I chose a single case study method of a large portfolio as it provided the opportunity to enhance understanding and

simultaneously enable a generalisation of the findings (Straits & Singleton, 2018; Yin, 2018). The DSIs chosen for this case study were delivered between July 2021 to December 2022, bringing in-depth insights of the program life cycle.

I studied DSI governance within a single organisation (Transport) where I am employed full-time and continue to deliver DSIs. This gave me good access to data to conduct the case study. My interest was to understand DSI governance of DSIs as experienced by the organisation's participants themselves and identify the uniqueness of this class of initiatives compared to other ICT and engineering initiatives. While I led the delivery of several DSIs outside the Active Transport portfolio, for the purposes of this chapter, I used only Active Transport DSIs to inform and validate the findings. This was because the Active Transport portfolio has benefited from the lessons learnt from the delivery of earlier DSIs.

Using an interpretive research tradition associated with case studies, ontological and epistemological assumptions on DSI governance emerged. As evidence I used data from SAP, Jira Align, Jira and Confluence software and lean portfolio management artefacts based on SAFe (Scaled Agile, 2023). The observations took place over a period of 18 months between July 2021 and December 2022 during the delivery of DSIs, especially during agile-based ceremonies such as Active Transport portfolio board meetings, portfolio sync, portfolio review, participatory budgeting, daily standup, iteration planning, planning interval (PI) planning, backlog grooming and iteration review, all of which reflected well what governance levels were appropriate.

7.4.2 Research Setting

As stated above, our research was situated within the Active Transport branch of Transport for NSW (Transport), a state government enterprise that leads the development of safe, integrated and efficient transport systems for the people of New South Wales in Australia. The organisation is in early stages of delivering \$72.2 billion worth of investment in transport infrastructure as outlined in its *Future Transport 2056 Strategy* (Transport for NSW, 2018). Most of the initiatives have a significant technology component, including in DSIs, which are enabling Transport to become a data-driven organisation in managing traffic congestion and delivery of services to customers (Transport for NSW, 2021).

The Active Transport portfolio consists of engineering projects delivering walking and cycling infrastructure across New South Wales as well as DSIs for benefits tracking and supporting

investment decisions. The portfolio aligns with the cycling strategy and action plan of the New South Wales Government to move towards a more sustainable Sydney (City of Sydney, 2018). Active Transport is a sub-portfolio of a larger Transport portfolio and has adopted lean governance for its definition and delivery (Jenner & Kilford, 2011; Scaled Agile, 2023). Its governance is influenced by the organisation's strategic themes, which then drive the portfolio vision, portfolio and team backlogs, budgets and guardrails.

When I commenced data collection in July 2021, Active Transport was functionally part of the Greater Sydney division and responsible for the whole state of New South Wales, including Greater Sydney and the Regional and Outer Metropolitan regions. In December 2021, considering the economic, health and sustainability benefits Active Transport brings to the state, a new minister of Active Transport was announced, resulting in a separate Cities and Active Transport division within Transport. With this increased focus on Active Transport, the outcome of this research assumed greater significance, not least because the DSI governance would evolve to reflect organisational governance.

The active transport portfolio consists of AU\$1.2 billion worth of infrastructure to be delivered between 2014 and 2027 across the state. The delivery of investments is either through 128 councils for smaller projects or the Infrastructure and Place Division of Transport for major projects. In the Get NSW Active (Walking and Cycling) program alone, Active Transport issues grants worth between AU\$50 million and AU\$75 million annually to councils to deliver walking and cycling infrastructure on local roads. The role of the Active Transport branch is program management of this major and politically visible stream of over 200 investments. The Parramatta Road Urban Amenity Improvement Program (PRUAIP) is another significant (AU\$198 million) investment to improve Parramatta Road, a state road that runs through multiple councils. The role of Active Transport is program management of the 32 projects involved, including coordinating all approvals within Transport. Major projects such as Parramatta Light Rail, Sydney Metro West and the 16 Regional Cities program deliver walking and cycling as part of their scope and the role of Active Transport branch is to track the delivery of scope and conformance to walking and cycling policies and procedures.

7.5 DATA COLLECTION AND ANALYSIS

Our data collection took place between July 2021 and December 2022. I used data from Excel spreadsheets, SAP, Jira Align, Jira and Confluence software and lean portfolio management artefacts based on SAFe.

7.5.1 The Challenge

While Transport has been delivering Active Transport (Walking and Cycling) projects since 2010, considering the significance of walking and cycling on the economy, health, sustainability and contribution to strategy (Transport for NSW, 2018), a separate branch was established in July 2021 to manage the Active Transport portfolio. This contrasted with other modes of transport—bus, ferry and light rail—which are part of operations and do not exist as a separate branch. At the time of its establishment, the Active Transport portfolio had three challenges:

- Manage the Active Transport portfolio efficiently and effectively,
- Use data for benefits tracking and investment decisioning, and
- Establish ways of working with the new team.

The Active Transport and wider Greater Sydney leadership teams came from traditional engineering backgrounds. The portfolio management was largely focused on reporting and to a lesser extent on “doing the right thing”. The governance was based on a gated process suitable for large infrastructure projects and did not account for innovation and the exploratory nature of DSIs. Large Excel spreadsheets with 30+ columns were used to manage the portfolio and management reporting was through PowerPoint slides. There was no mechanism to capture portfolio-level risks and issues. The portfolio budget was constructed manually by reviewing multiple sources, signed off by the Deputy Secretary and shared with the Minister of Transport and other stakeholders through House File Notes, a mechanism for communicating with the minister’s office. The monthly data collection of project financials for 400-plus in-progress projects was a challenge to compile due to the lack of technology, process and resources. This meant that confidence in the accuracy of portfolio reporting was low.

The data and analytics function of Active Transport was established to track benefits and help the branch understand where the next infrastructure should be built. In the case of the bus, ferry, light rail, Sydney trains and metro modes, usage could be derived by Opal Card tap-ons and

tap-offs. However, quantitative assessment of walking and cycling trips was challenging as the start and end of a trip triggered no event that could be measured. Piezo-electric and pneumatic tube counters were deployed on some cycleways to count cycle usage. Out of 100-plus cycleways and walkways delivered since 2017, the usage of only 14 permanent cycleways and nine pop-up cycleways was tracked daily during July 2021. In addition, some of the counters were broken and could not be fixed as their maintenance contracts had expired. While the Covid-19 pandemic gave Active Transport a once-in-a-generation impetus, the ability to measure the walking and cycling volumes for Greater Sydney and the Regional and Outer Metropolitan areas did not exist. To assess demand for the walking and cycling infrastructure across the state was similarly challenging in the absence of any sensors and any surveys that were used were annual or at best biannual and thus did not provide any real-time pulse of our walking and cycling customers. This meant projects were being approved in the absence of underlying data for both usage and the voice of the customer.

The Active Transport team brought together two distinct streams of work. The first stream of work was the management of civil engineering projects, working with local councils to build new active transport infrastructure. The second stream consisted of technology functions that support civil engineering projects through functionality such as the grants portal to help councils apply for Active Transport infrastructure funding and the collection and integration of data from embedded sensors to monitor usage and effectiveness and assist with program planning and rollout. The team blended a wide variety of skills from civil engineering through project and program management to technology and data specialists. These skill sets had their own ways of working which sat uncomfortably together and made planning and working as a team difficult. There was a common view across the team that planning and other upfront activities needed to be 100 per cent complete and signed off before work could start; this made the team slow and unresponsive to new or changed work. Planning and review cycles were annual which also limited their ability to respond in a timely manner. Establishing consistent ways of working was therefore paramount for the Active Transport leadership team.

7.5.2 The Solution

To address the existing challenges and to ensure that the Active Transport portfolio delivered its planned outcomes, any solution had to cover people, process and technology. I explored the following options:

Establish Minimum Viable Governance

Considering the size of the portfolio and the mix of civil engineering and DSIs, we implemented minimum portfolio and product governance but sufficient to deliver strategic outcomes and support business agility and transparency. This was aligned to the SAFe Lean-Agile.

The first activity the team undertook was to establish the vision aligned to the wider corporate vision of “*More people walking and cycling in NSW*” and the purpose “*The creation of viable transport modes as an integral part of the NSW transport system to encourage mode shift by giving customers greater choice*”. While powerful, the vision was simple enough for everyone to understand what Active Transport was trying to do.

The next step was to identify the portfolio operational and development value streams. Eight value streams emerged: walking, cycling, end-of-trip facilities, micro-mobility, schools, change and communications, maintenance, portfolio management and data and analytics. Metrics were established against each one. SAFe uses guardrails to implement portfolio governance. For the budget guardrails, we allocated the portfolio funding across three horizons – evaluating, emerging and investing/extracting but none in the retiring horizon, as shown in *Figure 7-1*.

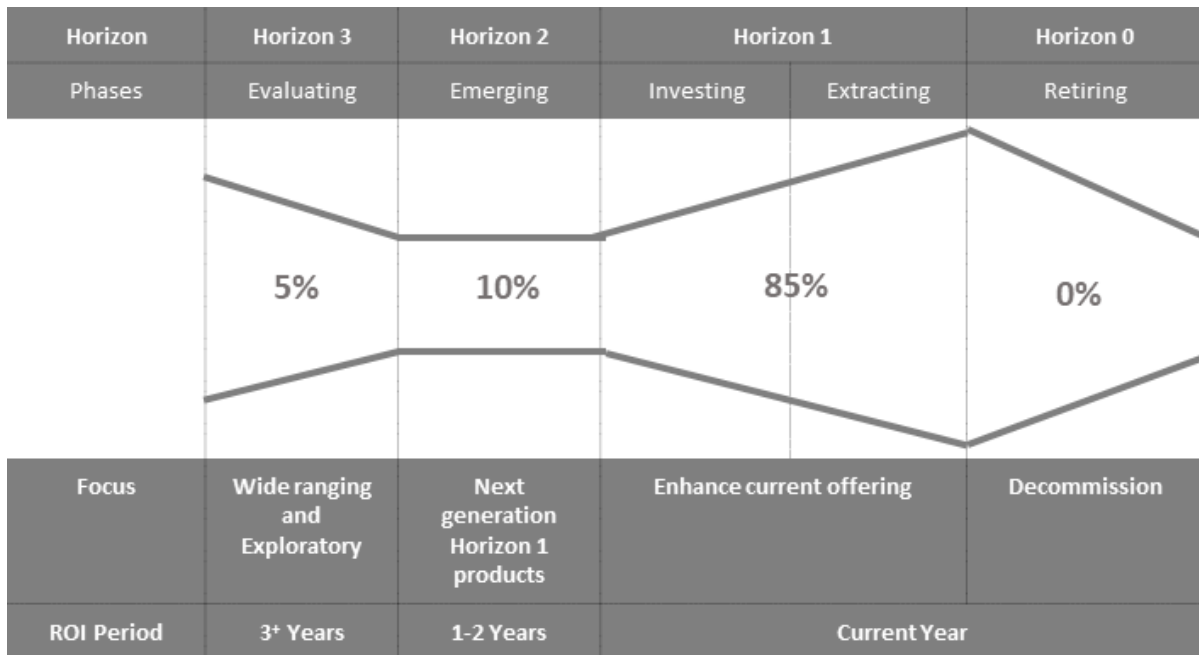


Figure 7-1. Active Transport investment horizon budget guardrail

As the Active Transport portfolio supports the Greater Sydney (GS) and Regional and Outer Metropolitan (ROM) divisions, a portfolio board was established with dotted reporting to both GS and ROM governance. The purpose of the board was to:

- oversee the Active Transport programs and projects and ensure they were continually aligned with the Future Transport 2056 (Transport for NSW, 2018),
- Oversee the identification and prioritisation of Active Transport needs and initiatives for investment,
- Oversee development of the Active Transport pipeline and ensure it was integrated into wider planning, development and delivery sustainable of travel choices, and
- Oversee the Active Transport portfolio and ensure scrutiny was in place for projects and programs.

The product manager and product owner collectively provided the product governance of artefacts such as features, stories, non-functional requirements and design artefacts. The product manager owned the vision and roadmap and provided a list of outcomes for the planning iteration (PI) which spanned three months (or six two-weekly iterations) prior to each PI event. The product owner defined the scope of iterations and stories and worked with agile teams to deliver.

Define DSI roadmap

The Active Transport leadership engaged a consultancy in May 2021 to develop a data and analytics roadmap for FY22–FY24. This was finalised in July 2021 and covered both walking and cycling, as shown in *Figure 7-2*.

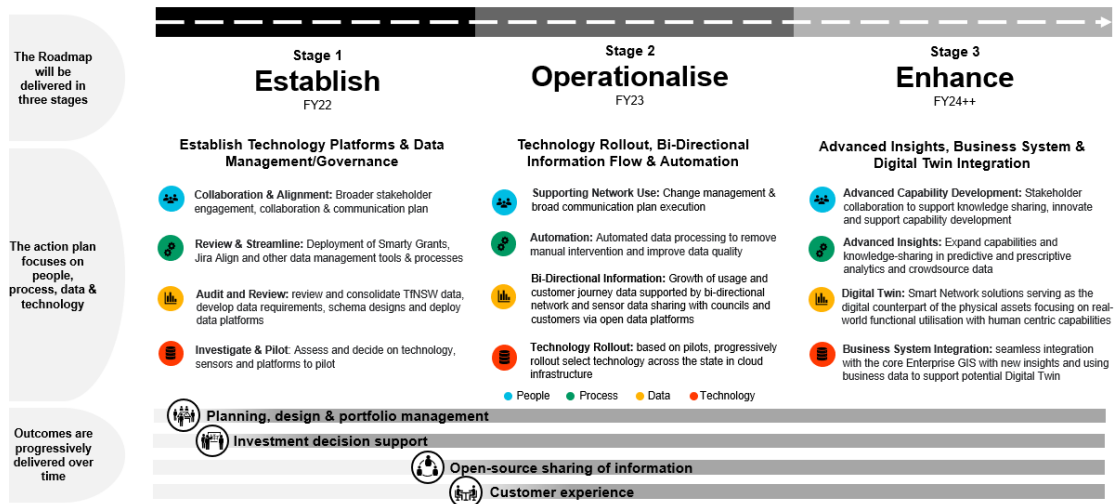


Figure 7-2. Active Transport Data and Analytics Roadmap

Eight teams were established to deliver the data and analytics program, which was one of Active Transport’s development value streams. The Planning Interval (PI) business outcomes were provided to the agile teams to shape what the teams would deliver in each of the Pis.

Empower people

The Active Transport leadership had already considered SAFe as a basis for new ways of working. An initial investigation confirmed that choice. For the team, SAFe provided:

- A consistent framework for ways of working,
- A means of coordinating and planning at the portfolio level through PI planning,
- A common language to describe the work and how the work flowed,
- Visibility through iteration and PI showcases, and
- A mechanism for team-led continuous improvement.

Although all the teams were new to Agile, a decision was made to launch quickly with a PI planning event as the benefits of having a coordinated plan were seen to outweigh any advantages

that might be gained by establishing Agile Team practices first. The key benefit was in the coordination, regardless of practices at the team level.

Launching quickly with PI planning also helped overcome the “won’t work for us” view that parts of the team had developed around agile practices. The rapid delivery of an integrated plan helped show that these techniques were in fact highly relevant and useful for the work they were doing.

The leadership team and key team members attended a two-day Leading SAFe course to establish a common understanding of the process and language before running the PI event. The first PI was set for only two iterations in duration to accelerate the learning of the teams and to synchronise the teams into a quarterly cycle. This was considered preferable to delaying the launch by four or five weeks. While it would mean two PI events close together, it would very quickly give the team additional experience in PI planning and let them see the first PI planning as a trial run for the main event two iterations later. This allowed the teams to relax and took some of the initial launch pressure off. Because of the short first PI, the PI event was also shortened to two half days.

The Portfolio, Data and Analytics team was split into four:

- Data Analytics (Data Amigos),
- Agile Portfolio Management Office (BEAT),
- Get NSW Active (Walking and Cycling) program (Active LAMS), and
- Parramatta Road Urban Amenity Improvement Program (Pruaipers)

Despite the limited time for the teams to prepare, they were able to develop a well aligned plan for the first, shortened PI and a draft plan for the first half of the second PI. A decision was made by leadership not to establish transformation metrics for the SAFe implementation initially. The view was that the transformation was for the teams, not to prove anything to anyone else. As the portfolio director said, “We will measure what we need to help the teams but proof that we are doing better will come from delivering more value, not in optimising some numbers.” A coach was embedded to help the teams develop their own agile practices as the PI progressed.

Innovate continuously

The Active Transport leadership team had already considered SAFe as a basis for new ways. We implemented tools such as Jira Align, Jira, Sharepoint, PowerBI in addition to SAP to capture portfolio epics, features, stories and other artefacts. Our iteration and PI retrospectives provided the feedback on what was working and what was not working to pivot if required.

Visualise everything

Information visualisation allows information to be conveyed efficiently and improves users' comprehension and ability to remember (Obie et al., 2019). We commenced by developing our intranet for use as a channel to communicate with the Active Transport team. We then created visualisations for Portfolio Analytics using data from SAP and Jira Align to show various breakdown by council, budget versus actuals, length of cycleways delivered in kilometres and open-to-traffic milestones. Next were visualisations of network analytics using PowerBI, with the data coming from piezo and pneumatic tube counters and cameras deployed across New South Wales, collected every five seconds in some cases. *Figure 7-3* shows the average walking and cycling counts per site per day for one year. While Active Transport's focus was on walking and cycling volumes only, the ML model had been trained to identify other vehicle types such as car, truck, bus and motorcycle, which made this solution scalable to the whole of Transport. The dashboards showed near real-time movement of all vehicles within the previous 24-hour period at one intersection in Sydney.

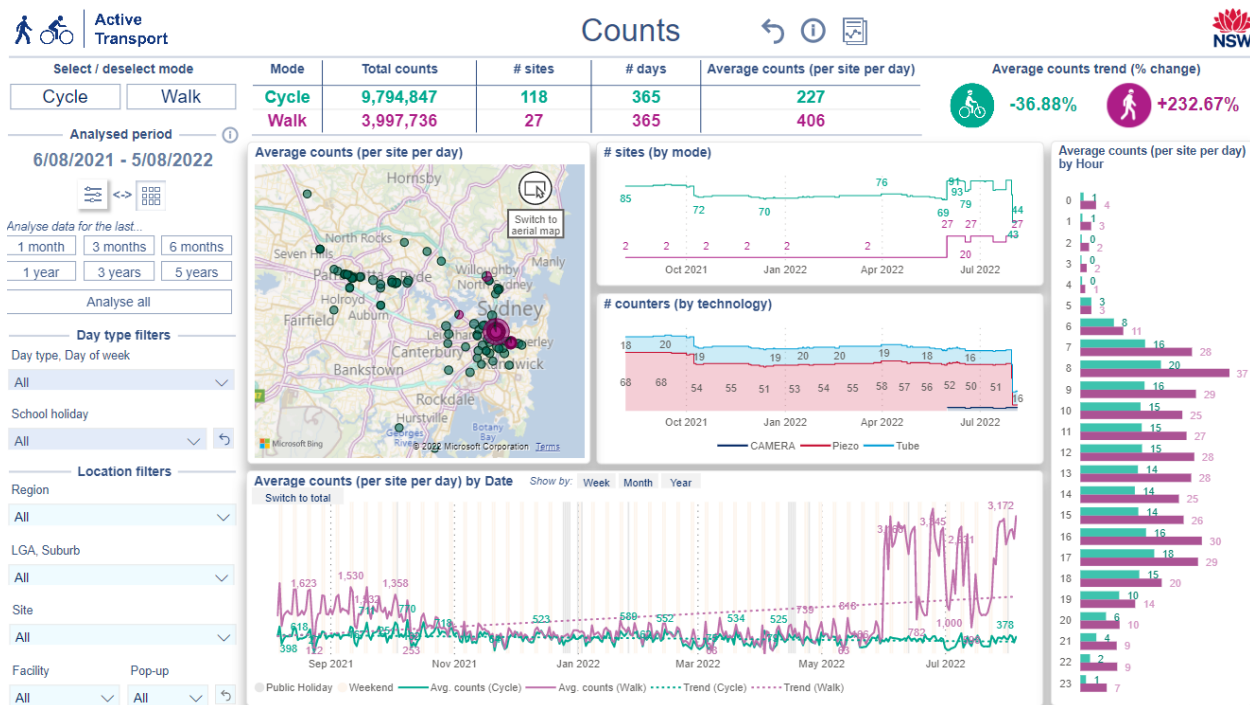


Figure 7-3. Average Walking and Cycling Counts in NSW

Understanding the voice of the customer became a priority for Transport. One of the solutions captured customer sentiments expressed in social media and other digital channels in near real time and were expressed as Walking Sentiment, Cycling Sentiment and Active Transport Commuter Sentiment scores. In a Transport first, this allowed us to better understand the impact of Active Transport on commuters and stakeholders throughout the infrastructure project life cycle.

Visualisation was a journey and we discovered more relevant datasets and new requirements emerged. Our portfolio and network analytics provided us with a great vehicle for storytelling.

Measure performance

Active Transport established key performance indicators (KPIs) across the value streams, as shown in Table 7-4.

Table 7-4

Active Transport Metrics

Objectives	KPI
------------	-----

Enable safe, reliable and connected walking and bike riding environments	• Kms of bicycle paths and shared paths delivered • No of bicycle parking facilities delivered [for stations only] • Projects open to traffic (completed) • Trips planned, including cycling on Trip Planner • Percentage decrease in the ratio of walking and bike riding incidents • Percentage increase in stakeholder and commuter sentiment score • Percentage reduction in journey times, as indicated by Transport Customer Survey
Transform walking and bike riding experiences for customers	• Positive to very positive experience as indicated by quantitative and qualitative feedback • Feedback indicates X% would take part in a walking or bike riding experience again
Generate insights into demand and performance of the walking and bike riding network	• Percentage increase in walkway usage • Percentage increase in cycleway usage • Percentage of walkways mapped in ArcGIS • Percentage of cycleways mapped in ArcGIS • Number of cameras deployed on walkways and cycleways across NSW • Percentage coverage of cycleways through sensors
Provide leadership and advocacy for walking and bike riding within NSW	• Number of projects including active transport infrastructure • Increase in funding for Active Transport projects

In addition, the teams used localised outcome metrics such as iteration goals and PI objectives to measure performance.

7.5.3 Outcomes Delivered

The following outcomes were delivered:

- portfolio, program and product governance,
- a consolidated view of active transport portfolio replacing multiple spreadsheets,
- SAFe processes (daily standups, iteration planning, backlog grooming, iteration reviews, PI planning, portfolio sync meetings),
- benefits tracking across the state using piezo and pneumatic tube sensors,
- The ability to manage Get NSW Active program grants,

- ways of working,
- flow metrics (end-to-end delivery of projects).
- additional SAFe processes (strategic portfolio review and participatory budgeting),
- project updates from councils using the Salesforce grants portal,
- improved work flow by reducing delays in development work streams,
- benefits tracking of investments delivered using cameras,
- model cycling volumes across the state, and
- data-driven investment decisioning, including both quantitative data and the qualitative voice of the customer.

Based on the outcomes delivered, I summarise our initial observations as follows. The DSIs benefited from a product mindset of not having to start and stop the work after ingesting each dataset as a project. The product delivered by DSIs continued to be enhanced, with features added at each iteration and each PI. Each dataset contributed to better benefits tracking and investment decisioning capability. This led us to our first guiding principle of *“focusing on product, not projects, and using agile methods and program management for delivery”*.

In alignment with SAFe Lean-Agile principles (Scaled Agile, 2023), we organised the teams around value for both engineering and DSIs. This reduced the handoffs and delays and we combined both engineering and technology teams to bring in efficiencies and increase focus on value. This led us to our second guiding principle of *“organising teams around value and not functions, which had traditionally been the practice in our organisation”*.

Traditionally, teams are formed and disbanded at the end of a project or a program. This brings in inefficiencies of ramping up the domain knowledge in the beginning and walking away with intellectual property at the end. In Active Transport, we created permanent teams and keep bringing in work to them. In next round of funding for the Get NSW Active program, we allocated a percentage of infrastructure budget to support persistent DSI team. This led us to our third guiding principle of *“funding value streams, which are multi-year and step away from the project-based funding required for individual business cases”*.

We stepped away from centralised decision-making and left it to those who knew the subject area well and were best positioned to make the decision. This was subject to financial delegation of authority as we still needed to follow corporate governance. This led us to our fourth guiding principle of “*empowering teams and decentralising decisions, which were frequent, time-critical and required local context*”.

I observed that visualising our portfolio and network data brought transparency to what we were doing and highlighted areas requiring attention. Visualisation helped us identify blockages in our flow. This included Kanban boards using Jira to show work in progress for all teams and portfolio and network analytics using PowerBI dashboards. This led us to our fifth guiding principle of “*visualising all aspects of the portfolio and providing transparency to all stakeholders*”.

Having the right level of data helped us making sensible portfolio decisions, such as where the resources should be deployed and where the next cycleway and walkway should be built. We were collecting both qualitative and quantitative data and to use for decision-making. This led to our sixth guiding principle of “*collecting sufficient data and using it for decision-making*”. The exploratory nature of data and DSIs is such that unless you try, you do not know what insights that data can deliver.

While DSI delivery timeframes are relatively short compared to those of construction projects, there is a need to optimise them. The only way is to collect data and identify activities on a critical path that causes delays. Faster delivery of DSIs results in faster delivery of value and benefits realisation and thus the flow needs to be measured. This led us to our seventh guiding principle of “*measuring flow and addressing areas causing delay*”.

Baselining capability allowed us to continuously improve. Once we commenced our DSIs, we continuously brought in changes to how we operated, in people, processes and tools. The data told us whether the changes were working or not. This led us to our eighth and final guiding principle of “*relentless improvement of people, processes and tools and having a growth mindset*”.

7.6 CONCLUSION AND RECOMMENDATIONS

In this section, I review our research question: “*What minimum viable governance is necessary for Data Science Initiatives to deliver envisaged value efficiently and effectively?*” and

summarise our findings from using Active Transport as a case study. I start with the findings from this chapter, discuss the limitations of our research, the implications for business managers and practitioners of the proposed governance and end with the recommendations.

7.6.1 Conclusion

In this Active Transport case study, I found that introducing SAFe (Scaled Agile, 2023) helped the newly-formed team to adopt new ways of working. The “command-and-control” mentality did not exist and traditional hierarchical governance structures were not created. I also found that the best-practice frameworks for portfolio management (Jenner & Kilford, 2011; Project Management Institute, 2017a) or IT governance (Agutter, 2020; ISACA, 2018) were more suited to stable environments and traditional command-and-control settings (Horlach et al., 2019). DSIs, on the other hand, which are innovative and exploratory, generally need agile delivery methods and have scope and goals elaborated in a continuous and iterative process. This makes traditional project governance, which is focused on successful delivery based on the iron triangle of time, cost and quality (Pollack et al., 2018), less suitable for DSIs and helped fill the gap I identified in the literature on the effective delivery of DSIs.

With only a two-day “Leading SAFe” training together, the team bonded and its members, who came from engineering and technology backgrounds and different ways of working, came to know each other better. While the technology teams were used to agile methods, the engineering teams followed waterfall methods with stage-gates. SAFe allowed us to work as one team. We started talking more about minimum viable product and business agility. By having daily stand-ups for each value stream, I was able to interact with all the team members to know what they were working on and if they faced any blocks on the day. Continuous communication and decision-making took place in these stand-ups. An iteration review at the end of two-week iteration to review progress brought in transparency about our iteration goals and what was achieved. Successes were acknowledged and celebrated. A PI planning ceremony allowed us to set up portfolio outcomes for each quarter and together the teams unpacked them into specific epics/features/stories for delivery into iterations. Fully empowered, the teams decided on “when” and “how” outcomes were achieved. Retrospectives at the end of the iteration and planning interval allowed us to continuously see what was done well and identify areas for improvement. All improvement areas were turned into actions and assigned to individuals to work on.

Our case studies show that the nature of governance changes along with the maturity of the organisation delivering DSIs. There is no one size of governance that fits all DSIs, even within the same organisation. Müller (2016) identifies four models of organisational project governance—process models, governance and governmentality-based models, nested models and layered models— which cover agency, institutional, stakeholder and transaction cost theories. He then describes the model of four-governance paradigms linking corporate governance with organisational project governance through an overlay of two dimensions – corporate governance orientation from shareholder to stakeholder and the control structures in a project’s parent organisation from a behaviour-to-outcome perspective. I extended this four-governance paradigm model to the DSIs and propose that, depending upon the maturity of the delivering organisation, DSIs can be governed using Versatile Artist or Agile Pragmatist paradigms.

Due to the exploratory and innovative nature of DSIs, I ruled out the conformist paradigm quadrant where the focus is on maximising efficiency by trusting process over the capabilities of individuals. Looking at the diverse skill set requirement of DSIs, as shown in Table 4-7, I also ruled out the Flexible Economist paradigm quadrant, where the focus is on flexible application of effective project management methods, tools, techniques and management approaches, often through a PMO. Our initial DSIs—Vanguard, CTABS and PTIPS Analytics— all required versatility from the delivery team to tailor or develop methodologies and work practices. Thus, I saw the versatile artist paradigm quadrant as suitable for organisations commencing their DSI journey with low process maturity and where stakeholder engagement still needed to be high. By our sixth and seventh DSI—MPR and Active Transport— the processes had matured and stakeholder orientation along with behaviour control through process compliance were appropriate. For organisations that were more mature in their DSI journey, the Agile Pragmatist paradigm was a suitable quadrant.

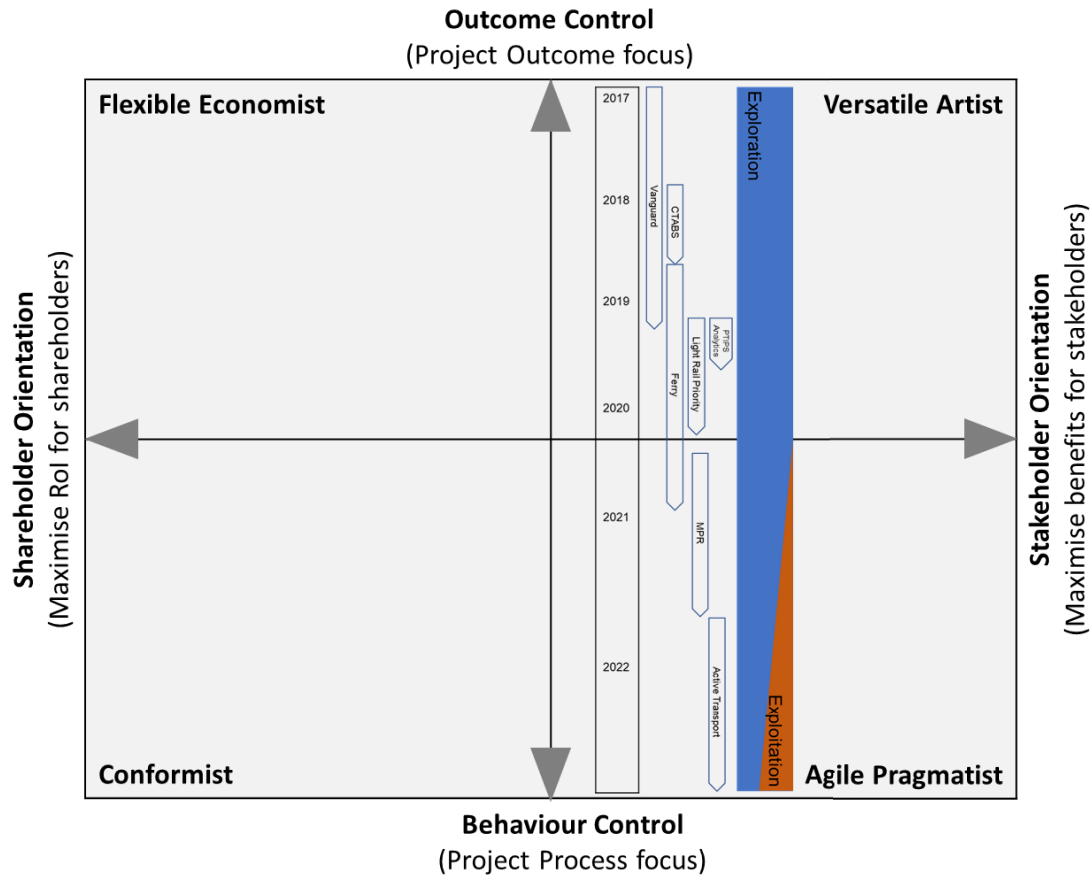


Figure 7-4. DSI Governance using Four-Governance Paradigms (Müller, 2016)

I concluded that using SAFe enabled the Active Transport portfolio to implement minimal viable governance of portfolio and product management and bring in a degree of business agility not previously experienced by team members and stakeholders. It allowed Active Transport to address the challenges it faced managing the portfolio efficiently and effectively, using data for benefits tracking and investment decisioning and establishing new ways of working with the new team.

7.6.2 Limitations and Implications of the Research

This chapter is based on the Active Transport portfolio of Transport for NSW and less than a year of implementing Scaled Agile and lean portfolio management. The \$1.2 billion portfolio consisted of 99% construction projects and 1% DSI. It is expected that as the portfolio matures and reliance on digitisation and data increases, the composition may change.

I believe that the guiding principles for MVG of DSIs identified here do not present an exhaustive and universal set and more may emerge as the field of data science advances. Future research could include validating the characteristics with other public and private sector organisations delivering DSIs in other countries. Another aspect is that DSIs are a more recent phenomenon and sit in a rapidly evolving technology and delivery space. This has an impact on the currency of the research work being done as some of the guiding principles may evolve as the maturity of DSI changes from being exploratory to exploitative.

The limited availability of methods and standards in DSI delivery has caused business managers and practitioners to chart their own path and thus introduce inconsistency in how DSIs are managed and governed in different organisations. With research such as this, it is expected that the standardisation of DSIs will increase and guide business managers and practitioners in the effective and efficient governance of the DSIs.

7.6.3 Recommendations

Based on this Active Transport portfolio case study, I recommend the eight guiding principles for minimum viable governance for DSIs, as shown in Table 7-5. While Active Transport used SAFe for delivering DSIs, I believe that the guiding principles are generic and can be applied by organisations using any other Agile delivery method. Table 7-5 shows how these guiding principles are aligned to the 12 principles mentioned in Table 7-2 from the Project Management Institute (2021b) and the Association of Project Management (2016).

Table 7-5
Guiding Principles for Minimum Viable Governance of DSIs

No	Description
(i)	Focus on product not projects – Use product management for DSIs and not project management for delivery [APM #1]
(ii)	Organise around value – Structure the teams around delivery of value and not functions [PMI #4, 7] [APM #2, 3, 4]
(iii)	Fund value streams – Instead of funding projects using business cases, fund the multi-year value streams [PMI #2, 4] [APM #2]
(iv)	Empower teams – Allow decision-making to be done by people who have the most knowledge and not based on hierarchy [PMI #1, 2, 5, 6] [APM #5, 6]
(v)	Visualise portfolio – Visualise all aspects of portfolio and make them available to all

stakeholders [APM #7, 11]

- (vi) **Data-driven decision-making** – Collect as much relevant data and use it for decision-making [PMI #3, 8, 10] [APM #8, 10, 12]
- (vii) **Measure flow** – Identify steps causing most delay in the flow and address them [PMI #3, 4, 5, 8, 11, 12]
- (viii) **Relentless improvement of people, process and tools** – Baseline capability and continuously improve [PMI #2, 3, 4, 7, 8, 9, 10, 11, 12] [APM #7, 9, 10, 12]

As a practitioner deeply involved in the governance transformation efforts within the organisation, I advocate for the adoption of the eight guiding principles outlined in this paper to establish minimum viable governance for data science initiatives (DSIs). Drawing upon firsthand experiences and observations, these principles are designed to strike a balance between enabling innovation and ensuring effective delivery, thereby maximizing the value derived from DSIs. Moving forward, I encourage organisations to embrace a product mindset and prioritise business agility in response to the evolving digital landscape.

8. Insights into Building Better Business Cases for DSIs

This chapter is based on my working paper “When Should Data Science Initiatives be Managed as Exploratory Projects? Insights for Building Better Business Cases” (Mathur, 2024).

8.1 ABSTRACT

Investment in data is increasing exponentially as organisations recognise the potential value that data can deliver. Technology has also matured in that use of big data is becoming a general-purpose technology and is no longer limited to a privileged few. While DSIs provide the mechanism to extract value from data, the success rate of their delivery is not keeping up with the need to realise value; Gartner estimates that 85% of projects delivering DSIs fail (Asay, 2017).

In this chapter, I draw on the concept of “exploratory projects” (Lenfle, 2008) to arrive at an explanation that traditional waterfall approaches to program management might actually heighten the risk of such failures. My investigation suggests that waterfall approaches to program management are likely to set up structural tensions between business case development, program design, delivery and benefits realisation, which decouple value creation from intent. I conclude by arguing that the business case for DSIs should be constructed differently; instead of the traditional focus on costs and benefits, decision-makers should understand the rationale behind such exploratory investment and that the benefits profile can be tracked across the investment life cycle. I also propose applying a product mindset to business cases such that value streams rather than individual projects are funded and empowered teams should be allowed to determine the scope of their DSIs. I believe this change will lead to a transformation in how many organisations view their DSI investments.

8.2 INTRODUCTION

Reflecting on the change programs I have delivered over 30 years, I can classify them as either exploratory or exploitative. Any organisation has to balance its portfolio with an appropriate mix of exploration and exploitation to succeed (March, 1991). Lenfle (2008) defines exploitative projects as those that focus on optimising the triple constraints of cost, quality and time to deliver new products and services whereas exploratory projects are those where from the outset neither the goals nor the means of attaining them are clearly defined. This framing allows us to reconsider how a business case for these two distinct categories of investment can be constructed.

In this chapter, I draw on in-depth case studies of seven DSIs at Transport for NSW (Transport) to theorise the boundary conditions to determine whether they would be better managed as exploratory rather than exploitative projects. The first DSI, with a waterfall (sequential) delivery approach was seen to be failing, triggered a switch to a fully agile approach, whereby the delivery teams reduced the planning horizon and scope of work delivered into two-week sprints. Table 4-1 summarises the seven DSIs and how the delivery uncertainty fell from “High” for the first four DSIs to “Medium” for the subsequent three. I argue that while the agile approach was used to manage uncertainty and risks in all the DSIs, the first four DSIs were particularly appropriate candidates to be managed as exploratory projects. As the program team grew more competent in managing the uncertainty and the risks, the subsequent DSIs could then potentially be managed as exploitative projects. Decision-makers should be aware, however, that while the balance of exploratory versus exploitative changed over time (as illustrated in *Figure 4-1*), the exploratory component is always present in DSIs.

I show here that DSIs have little similarity with initiatives defined in classical project management frameworks (Axelos, 2022; Project Management Institute, 2017b, 2021b). This is because the goals are unclear at the beginning, deadlines are hard to define and the schedule changes frequently. In addition, due to the rapidly changing technology landscape and a continuous re-setting of business expectations, experimentation is necessary throughout the pathway to a solution. Further, while the business will have some high-level objectives for the investment, the potential value to be delivered is largely unknown at the time that the business case is generated. The delivery costs thus carry a high degree of risk as the technology and path of delivery cannot be mapped *ex ante* to any degree of high fidelity. As decision-makers look for certainty around

costs and benefits, approval of the business case for such investments is difficult to obtain when seen through the lens of a “Change the Business” portfolio. I argue that business managers should not use the traditional costs and benefits metrics for decision-making and instead have a separate category for innovation and exploratory investments in which DSIs would belong. Failure to do this may also result in DSI investments not being approved by business managers or a proliferation of optimism bias in DSI business cases. This optimism bias has been documented in the planning of large infrastructure projects where a high level of misinformation about costs and benefits is used to justify projects to go ahead during the early stages of project development. This misinformation generates high risks for the project to be delivered (Flyvbjerg, 2007a). Business managers also need to acknowledge the role DSIs can play in capability development and experimentation and how data can be used for future decisioning in product and service development.

This chapter contributes to program management research by focusing on how to construct a business case for DSIs, which are innovative and exploratory in nature, that is well-understood by decision-makers. I argue that project portfolio management practices need to acknowledge that exploratory projects need a separate category in project prioritisation. I also highlight that for a product not yet constructed, the cost of associated DSIs should be embedded in the overall product development cost and not treated as an afterthought or separate technology project investment. As investment in DSIs is expected to increase significantly in the next decade, I expect this research will lead to improved decision-making processes for such investments.

8.3 LITERATURE REVIEW

In this section I review the literature about exploratory projects. I then introduce data science and related investment trends, DSIs and how some of their characteristics, such as uncertainty in terms of their goals, take them into the exploratory projects category and then review how DSI business cases can support project portfolio management.

8.3.1 Exploratory Projects

Exploratory projects can be characterised as projects for which neither the goals nor the means of attaining them are clearly defined from the outset (Lenfle, 2008). Delivery of such projects cannot be done effectively using standard project methodology, which largely focuses on

delivery of a defined scope, cost and schedule. The inclusion of an agile practice guide in the *PMBOK Guide Sixth Edition* (Project Management Institute, 2021b) is an acknowledgment by the Project Management Institute that waterfall methods do not allow for the effective risk management of ICT-enabled projects, thus causing high failure rates in delivery. As a result, organisations have moved to agile practices for delivery of ICT-enabled projects. CH Loch et al. (2011) demonstrate the limited relevance of standard risk management methods for projects with unforeseeable uncertainties, or unknown unknowns. Further, Christoph Loch and Sommer (2019) discuss the need for managers to adopt flexible execution methods for exploratory projects and the reluctance of supervising managers to provide this flexibility in the evolution of project goals. I think it is important in relation to the execution of DSIs that “being flexible about the means” does not imply “losing control” by top management.

Research projects focus on experimentation in areas which can be used in future projects. Development projects, on the other hand, deliver a unique product or service and thus benefits are relatively more easily understood by business managers, especially those applying a financial lens to the investment.

8.3.2 Data Science

Data science can be defined as the collection, preparation, analysis, visualisation, management and preservation of large collections of information to generate actionable insights (Dhar, 2013; Saltz & Stanton, 2017). Data analytics and business intelligence have been used by organisations to provide answers to specific questions based on existing data but data science goes one step further to parse large data sets, sometimes in unstructured ways, to expose insights and do predictive analysis.

8.3.3 Data Science Initiatives

I define DSIs as related projects or programs that involve the application of data science techniques and methods to address complex problems, generate insights or create new products or services. DSIs have unique characteristics which separate them from traditional ICT investments:

1. DSIs carry a high degree of uncertainty right from initiation through to their closing phases, except when the data is from a well-defined and structured source.

2. DSIs are often enablers for decision-making and may not make a direct benefit contribution.
3. Neither the goals of a DSI nor the means of attaining them are clearly defined from the outset, with the caveat that as the market matures, the emergence of pre-built solutions will reduce the uncertainty.
4. DSIs are not independent of each other and each DSI acts as an enabler for the next one.
5. The skills required to deliver a DSI are different from those required for a typical ICT program.
6. DSIs do not end and after initial delivery but convert into managing the product, the model and the data.

In the delivery of a DSI, an organisation and its teams go through stages of exploration, transition and exploitation. The process of delivering DSIs becomes more efficient as more are delivered and the organisation learns through the experience. With the exception of data from a well-defined and structured source, every new dataset carries uncertainty in its scope and quality. This uncertainty therefore introduces the underpinning exploration component to a DSI with a shift to the exploitative component as the organisation and its teams mature.

When I look at the journey of DSIs in an organisation, the initial DSIs belong to “exploratory projects” category due to uncertainties and unknowns. As the organisation delivers more DSIs, the number of the uncertainties and unknowns starts reducing. While the business goals may remain uncertain, repeatable patterns emerge for the underpinning technology and methods. I see a link between the maturity of a program team and a business to deliver DSIs and the need to manage them as “exploratory projects”.

In the previous chapter, I investigated DSI governance at project, program and portfolio level and found that the best-practice frameworks for portfolio management (Jenner & Kilford, 2011; Project Management Institute, 2017a) or IT governance (Agutter, 2020; ISACA, 2018) were more suited for stable environments and traditional command-and-control settings (Horlach et al., 2019). DSIs, which are innovative and exploratory, generally follow agile delivery methods and have their scope and goals elaborated in a continuous and iterative process. This makes traditional project governance focused on the successful delivery of the iron triangle, namely time, cost and

quality (Pollack et al., 2018), less suitable for DSIs. Instead, I proposed a minimum viable governance using SAFe (Scaled Agile, 2023) for portfolio and product management to bring in a degree of business agility, using the following guiding principles:

1. Focus on product not projects – use product management for DSIs, not project management, for delivery.
2. Organise around value – structure the teams around delivering value, not functions.
3. Fund value streams – instead of funding projects according to business cases, fund the multi-year value streams.
4. Empower teams – allow decision-making to be done by people who have the most knowledge, not based on hierarchy.
5. Visualise the portfolio – visualise all aspects of the portfolio and make the visualisations available to all stakeholders.
6. Data-driven decision-making – collect as much relevant data and use it for decision-making.
7. Measure flow – identify the steps causing the most delay in the flow and address them.
8. Relentless improvement of people, process and tools – establish a capability baseline and continuously improve.

I thus argue that due to the uncertainty around their benefits and the delivery itself, DSIs are typically exploratory projects. Other than the set-up of the foundational infrastructure, the delivery work packages of DSIs cannot be clearly defined at the outset and thus the schedule cannot be prepared in great detail. Nor can DSIs conform to the rational waterfall project approach in the delivery of a unique product, service or result within a specified period, defined budget and quality requirements. While iterative delivery has been proposed for software projects for some time (Mathur, 2005), DSIs lean more to being delivered in a progressive elaboration using agile methods.

In Chapter 6 - Framework to Manage DSIs, I proposed a DSI delivery framework with five core domains and integrating the PMI's *Standard for Program Management* (Project Management Institute, 2017b) for program management, the Prosci framework (Hiatt, 2006) for people change management, SAFe (Scaled Agile, 2023) for solution delivery, DAMA's DMBOK (Earley, 2017) for data management and CRISP-DM (Chapman et al., 2000) for data science processes, as shown

in *Figure 6-3*. This delivery framework is, I propose, a step towards the standardisation and efficient delivery of DSIs.

8.3.4 Business Case

Organisations use business cases for project portfolio decisions. A business case articulates the business drivers, costs, benefits and the factors affecting the achievability of a project (Axelos, 2022; ISACA, 2010). The benefits should be identified in sufficient detail to allow their contribution to the organisation's strategic measures be assessed against other potential investments. While financial metrics such as NPV and hurdle rates of return are commonly used in decision-making, Christensen et al. (2008) describe them as “innovation killers”. McGrath and McManus (2020) propose a DDP approach to integrate agility and innovation into an organisation while minimising disruption to existing business. Their DDP approach for digital transformations and product innovation has five key steps:

1. Define the operating experience: it is not just about digital,
2. Focus on specific problems: identify outcomes and progress metrics,
3. Identify your competition: cast a wide net,
4. Look for platforms: do not forget the ecosystem implications, and
5. Test your assumptions: failures are lessons too.

SAFe (Scaled Agile, 2023) recommends an iterative build-measure-learn cycle for product innovation and strategic investment. It uses Epic (see Chapter 2) instead of projects as a container for a significant solution development and insists on a definition of minimum viable product (MVP) to be created by product team. SAFe proposes creation of a lean business case as illustrated in *Figure 8-1* for review by the lean portfolio management function.

SCALED AGILE

SCALED AGILE

Lean Business Case

Epic Name: (Short description for the Epic)	Funnel Entry Date: (Date the Epic entered the funnel)	Epic Owner: (The name of the Epic Owner)
Epic Description: (Consider using the description from the Epic Hypothesis Statement as a starting point. Refer to the Epics article for more details)		
Business Outcome Hypothesis for MVP: (States the quantitative and qualitative benefits that the business can anticipate if the hypothesis is proven true)		Leading Indicators and KPIs (Establish Innovation Accounting metrics to provide leading indicators for the business outcome hypothesis)
Minimum Viable Product (MVP) Features (Feature or Capability) ...	Additional Potential Features (Feature or Capability) ...	Nonfunctional Requirements - List any new or changed NFRs ...
Sponsors: (List key business sponsors who will be supporting the initiative)		
Users and Markets Affected: (Describe the user community and any markets affected)		
Impact on Products, Programs and Services: (Identify products, programs, services, teams, departments, etc. that will be impacted by this Epic)		
Impact on Sales, Distribution, Deployment: (Describe any impact on how the product is sold, distributed, or deployed)		
Analysis Summary: (Brief summary of the analysis that has been formed to create the business case.		Go / No-Go: (Go, or No-Go recommendation)
Forecasted Story Points for the Epic's MVP (Estimated story points for the MVP of the epic)		Forecasted Monetary Cost for the Epic's MVP: (Example: Estimated story points * cost per story point for MVP features)

costs, spinning success scenarios while ignoring the possibility of mistakes.” J. P. Kotter (2012) estimates that up to 70% of change initiatives fail to deliver on the benefits they set out to achieve. This leads me to argue that DSIs should not be subject to financial metrics, not least because they are prone to rigging to get business cases approved. Rather, alternative systems such as DDP and a lean startup cycle should be used to support intelligent decisions for future growth.

8.3.5 Summary of Literature Review

In the previous section, I investigated exploratory, research and development projects. I then explored data science and related investment trends. I described DSIs, including their unique characteristics, the minimum viable governance and their delivery framework, and explored traditional and lean-agile business cases to support project portfolio management.

This view of the literature motivated me to ask the following research questions:

“How should the business cases for DSIs be structured? Is there a difference in business cases when DSIs are managed as exploratory projects versus exploitative projects? Is there a difference in business cases when DSIs cover products already constructed versus not constructed?”

8.4 RESEARCH SETTING AND METHODS

8.4.1 Research Methods

In this chapter I share my work at Transport for NSW (Transport) and my lived experiences of and reflections on constructing and delivering business cases for DSIs as case study research (Eisenhardt, 1989; Yin, 2018). This multi-methods approach (Hunter & Brewer, 2015; Straits & Singleton, 2018) allowed me to generate insights for building better DSI business cases.

The seven case studies at Transport covered the six years between 2017 and 2022 inclusive. This period incidentally maps to my joining Transport in late 2016 as a program manager and commencing work on delivering my first DSI in 2017. I was the program manager for the first six DSIs used as case studies and then business sponsor and portfolio director for the seventh DSI, which was completed in December 2022.

8.4.2 Research Setting

This research was situated in the Operational Systems division (now merged with Customer Strategy and Technology division) and Active Transport branch in Transport. The researcher's journey mirrors that of Transport as an organisation. When he joined Transport as a program manager, he had significant experience in delivering ICT transformation programs using both waterfall and agile methods but none in delivering data-related programs. This chapter tracks the delivery of DSIs, the challenges he faced in understanding some of the unique characteristics of DSIs, his ability to construct business cases and get them approved, his ability to influence near real-time data-driven decision-making in Transport and growth in maturity of Transport to harness the value of data underpinned by his own ability to deliver DSIs.

8.5 DATA COLLECTION AND ANALYSIS

Seven DSIs were used for data collection which the I managed at Transport. The data sources included documentation, archival records, direct observations, participant-observation and physical artefacts. While the scale of the DSIs was different, together they paint a good picture of unique characteristics and business cases, as shown in Table 4-1 *Summary of Transport DSIs*.

The business case benefits, the benefits delivered and the benefits complexity are summarised in Table 4-2.

Starting with the first DSI, called Vanguard, its vision was to consolidate and disseminate data and information to contribute to a public transport network where customers and staff felt safe and always travelled with a valid ticket. It was focused on increasing revenue through improved fare compliance, customer satisfaction and security outcomes. Vanguard consolidated data from multiple sources to paint a picture of fare evasion and security. It constructed a fines lifecycle by linking tickets scanned, fines and cautions issued and tracking collections from Revenue NSW. This provided Transport with the information about (i) where revenue leakage was taking place, specifically which mode and what time of the day so that transport officers could be deployed more effectively and (ii) the profile of users not tapping on when boarding public transport so that they could be better educated with targeted campaigns. Vanguard also constructed a picture of crime and incidents across NSW transport. It combined crime data from the Bureau of Crime Statistics and Research (BOCSAR) and various other modes on low level incidents to paint a

wholistic picture. This enabled Transport to deploy the Police Transport Command at the right time and in the right areas.

In the Vanguard Success Story video, Tony, who became the sponsor in 2018, explained, “[The] Vanguard program provided an innovative opportunity to consolidate transport crime statistics, security incidents and fare compliance data into one system. The challenge was that data sets were not centralised and held by several government agencies and key transport operators. File structures differed and data collation validation systems were not automated. There were some inefficiencies within the Security and Revenue Protection teams where time was lost collecting, cleaning and collating data rather than conducting analysis and developing strategies and operational outcomes in the Transport Cluster. So, we looked to find a better solution.”

Tony closed with, “Vanguard has consolidated key datasets. It has greatly enhanced our capability to analyse trends, identify hot spots and share these insights with our partners.” While Vanguard did not deliver the business case scope in full and ingested only six of the possible 21 data sources, it was perceived as a successful delivery. It laid the foundation of data management and DSI delivery and became a showcase for people, process and technology outcomes. As a result, Vanguard was nominated for the 2018 Transport (Cluster) Awards in the Safety Category and was a finalist in the Project Management Institute’s 2019 Awards in the “Innovation in Project Management” category.

When I look back, I inherited a business case that took 12 months to be developed and approved. It followed a traditional ICT investment structure with hard costs and benefits defined to justify the investment. The benefits outlined were (i) improved fare compliance and reduced revenue loss caused by fare evasion, (ii) predictive, targeted and multi-modal deployment of Police Transport Command and transport officers, (iii) an evidence base that could be used by transport operators when requesting police assistance, (iv) a single-source-register of all security incidents that would provide better intelligence to agencies and inform the development of crime prevention initiatives and (v) better outcomes for Transport customers through improved evidence collection and communication between business managers, including a safer and more compliant customer environment.

Table 8-1
Vanguard Business Case Quantified Benefits

Benefit type	Current	Target	Benefit measure	Benefit owner
Improved fare compliance	6.40%	5%	Non-compliance rate	Security and revenue protection
Improved fine payment	32%	45% (in Year 7)	% paid on time	Sydney Trains, Transport

Vanguard showed “Improved fare compliance” as a business case benefit (see Table 8-1) but the delivery team could not establish the direct benefit contribution. Though the team could provide analytics on fare evasion to its internal Security & Revenue Protection team, to transport officers and to the NSW Police, they were not accountable for actions taken to reduce fare evasion, including changing the behaviour of travelling public. The team subsequently identified some deep socio-economic factors that led to fare evasion in certain parts of Sydney and New South Wales, including finding repeat offenders. Key observations for us were (i) Vanguard was at best an enabler of the benefits outlined in the business case; the sponsor of this DSI could not be held responsible for improved fare compliance and improved fine payment benefit metrics; (ii) The scope of work in the business case contained non-functional data management requirements only and no relationship to how the benefits were to be delivered; (iii) the schedule in the business case had a 13-month delivery deadline, which was found inadequate considering the uncertainty around data readiness. Ultimately, it took slightly over 27 months to deliver the program; (iv) the Vanguard business case had a waterfall delivery approach which did not reflect what was required to manage the uncertainty of a DSI; (v) The original risk register did not identify the key risks encountered in the delivery phase; (vi) Vanguard did not deliver the full scope of 21 data sources and ingested only six of them, and (vii) it was difficult to track the benefits profile and trace it directly to the program.

The second DSI was the Ferry program, the vision for which was to implement an evidence-based management of the ferry contract and to improve customer experience. The program used data coming from the ferries every 30 seconds to show their real-time location on Sydney Harbour, their on-time or delayed status and to provide service performance reporting. The Ferry program delivered foundational Operational Data Lake (ODL) storage capability for sharing data and business insights to other programs and stakeholders, including service delivery, intelligent congestion management program, integrated planning and customer services. An environment using Microsoft Azure Cloud was set up and real-time vessel tracking and ferry on-time running

dashboards were delivered. This delivery coincided with new Sydney Ferries operator incorporated into Transport's systems.

One observation on the business case was that it took six months to be developed. Due to internal restructuring at Transport, the approval took another 12 months. The proof of concept was approved as part of the business case development and was completed. Key observations were (i) the Ferry program required new cloud-based Microsoft technology infrastructure in place for solution enablement, which increased the business case costs; (ii) Transport's finance department had little appetite for any investment with a benefit-to-cost ratio (BCR) of less than 1; (iii) the technology landscape was evolving and it was difficult for the Ferry program to deliver an architecture that was future-proof, and (iv) user acceptance testing of the solution showed the quality of data received from the ferries was not good enough for dashboards to be operationalised. A different solution was required but this was dependent on the ferry operator and was not delivered due to personnel changes at Transport.

The third DSI, called Community Trip Allocation and Booking System (CTABS), was terminated due to its non-delivery but it provided interesting insights nonetheless. The business objectives of this DSI were to (i) obtain visibility of community transport services in NSW; (ii) understand the customers by answering the questions of "Who?", "How?", "Why?" and "Where?"; (iii) understand the trips and travel patterns; (iv) assess service quality; (v) investigate opportunities to improve service delivery; (vi) determine if CTABS had resulted in operational efficiencies and (vii) assist in managing contracts. The scope included transferring consolidated CTABS data daily from its disaster recovery site to Transport to enable data analytics and verification of provider self-reporting.

The project used Microsoft Azure platform which management at Transport were keen to try out on a small, self-contained initiative. As this technology was new to Transport, there were limited in-house skills and the project was severely impacted. The development environment using Transport's Azure offering could not be set up in three months and to deliver some business outcomes it was agreed to set up a stand-alone Azure environment for development and production purposes. Using data from the CTABS Disaster Recovery Environment, trip dashboards were produced and published internally within Transport.

Our observations on the business case were that the investment required was small and the funding was available, so a light business case was acceptable to commence the project. Key observations were (i) Transport's Regional and Outer Metropolitan division's risk appetite was high considering the small investment requirement; (ii) only a light business case was required to commence the project; (iii) both the sponsor and the product owner of this DSI were involved in the technology decision-making; (iv) the project underestimated the time needed to implement appropriate data management as there was no underpinning foundation available; (v) a diminishing return on investment caused the project to be shut down though the scope was not fully delivered and (vi) the project delivered sufficient insight to Transport to commence other investments using Microsoft Azure platform.

In the fourth DSI, the PTIPS, the analytics proof of concept was recognised as successful by the business sponsor and regarded as a showcase for agile delivery and the setup of a solid foundation for future DSIs. PTIPS supports the operational requirements of all public transport buses in New South Wales, receiving location data from every bus every 10 seconds. The business objectives of this project were to (i) validate an analytics solution using Azure ODL; (ii) provide self-service capability to bus contract managers and operators with a minimum six months of PTIPS data and (iii) determine the operational expenditure (OPEX) requirements from January 2020 onwards. The original scope included the development of three dashboards: (i) a "cancelled trips" and "incomplete trips" (shift) summary report; (ii) a heat map of complex trips and looped services and (iii) depot departure report. This DSI delivered 10 complex PowerBI dashboards to the Transport's Greater Sydney division. The ability to move PTIPS operational data from Amazon Web Services (AWS) to Azure environments was implemented and the entire data orchestration operationalised.

The investment required for this DSI was small and the funding was available, so a briefing note was acceptable to commence the project. Our observations on the business case were (i) that the appetite to get self-service analytics capability was high considering the small investment required; (ii) only a briefing note was required to commence the project; (iii) instead of creating a detailed business requirements document, an interactive and iterative approach was adopted for the dashboards development; (iv) Azure architecture from the Ferry program proof of concept was extended to deliver this project, highlighting the dependency and benefit from previous DSIs; (v)

sponsor support was strong and (vi) to bridge the skills gap, two vendors were engaged to deliver underpinning infrastructure and visualisations.

The business objectives of the fifth DSI, the Light Rail Priority Project, were to (i) support optimising Sydney Light Rail journey times; (ii) provide light rail with enhanced priority at intersections; (iii) increase the visibility of light rail vehicles to the Traffic Management Centre (TMC) and the Sydney Coordinated Adaptive Traffic System (SCATS); (iv) support a decrease in Sydney traffic congestion and (v) implement a hardware-free solution for all SCATS intersections. The scope included (i) updating the existing PTIPS priority engine to predict light rail services and scoring of light rail vehicles based on a policy-based algorithm; (ii) Storing data into an accessible repository; (iii) provide dashboards to monitor the operational effectiveness of the priority engine and (iv) provide real-time operational support to all services. The project delivered the functionality in production environment. The performance tuning for key intersections was completed to deliver priority to the light rail.

Our observations of the business case were that this investment was mandatory for the CBD and Sydney East Light Rail (CSELR) project as it delivered priority to light rail vehicles at 56 intersections and associated analytics to measure the effectiveness of the solution. The compliance nature of the investment meant a briefing note was sufficient to secure funding. Key observations were (i) linking analytics to key business outcome made the investment a compliance requirement; (ii) a briefing note was required to commence the project; (iii) the project extended the PTIPS used for buses only to a multi-modal system and allowed a light rail mode; (iv) the Azure architecture from the PTIPS analytics proof of concept was extended to deliver this project, highlighting the benefits outcome relationship to previous DSIs and (v) PTIPS resources were leveraged to deliver this project and a vendor engaged to deliver visualisations.

The sixth DSI, the MPR program, commenced in January 2020 to “ensure data management and architectural consistency of ODL across multiple performance reporting business cases”. Its initial scope included bus (metro), ferry, light rail and bus (regional) performance reporting and the scope was extended in June 2020 to include Sydney Metro performance reporting as well as data ingestion and self-service projects. The MPR program was one of the most successful DSIs delivered at Transport. Terry, the light rail performance reporting business sponsor, nominated the project team for operational systems rewards and recognition. The program was a finalist in the

Intelligent Transport Systems Australia Awards in the “Excellence in Transport Data Award” category. The *MPR Success Story* was published by our delivery partners [Agile Analytics](#) and [data-driven](#), showcasing the strength of partnership and use of technology to deliver data-driven business outcomes. Key outcomes, publications and achievements are listed in Figure 8-2.



Figure 8-2. MPR Program Outcomes, Publications and Achievements

The MPR program closed in June 2021 having delivered a scalable big-data platform to store, process and service the analytics needs of Transport contract managers, operators and data analysts.

Our observations of this DSI were that the MPR program did not have one business case. The initial team was set up using the Ferry program business case and the subsequent business cases of bus (metro), light rail, bus (regional) and Sydney Metro performance reporting, as well as the data ingestion and self-service projects, leveraged the delivery capability. Key observations were (i) the Transport’s Operational Systems’ appetite for consistency in data management and architecture for ODL was high; (ii) the business sponsors focused on business outcomes and left the technical governance to a separate forum; (iii) the sponsors accepted that uncertainty existed around delivery of the data right quality, including with appropriate business rules and visualisations and that the team would deliver the best possible outcome in the shortest amount of time (which was a marked shift from a waterfall-oriented fixed time-cost-scope mindset) and (iv) the business cases funded delivery capability (or a SAFe Agile release train).

The last and seventh DSI described in this chapter was the Active Transport Data and Analytics program. The Active Transport portfolio consisted of \$1.2 billion worth of cycling and walking infrastructure, with all its economic, health and sustainability benefits, to be delivered between 2014 and 2027 across New South Wales. The delivery of investments was either through 128 councils for smaller projects or through the Infrastructure and Place Division of Transport for major projects. Through its Get NSW Active program, Active Transport issues grants of \$50 million to \$75 million annually to councils to deliver walking and cycling infrastructure on local roads. The role of Active Transport is program management of this major and politically visible stream of more than 200 investments. The PRUAIP, for example, was a significant \$198 million investment to improve Parramatta Road, which is a major state road through multiple local councils in Sydney. The role of the Active Transport branch in relation to the PRUAIP, was management of 32 projects, including coordinating all approvals in Transport. Major projects such as Parramatta Light Rail, Sydney Metro West and the 16 Regional Cities Services Improvement program deliver walking and cycling as part of their scope and the role of Active Transport is to track the delivery of scope and conformance to walking and cycling policies and procedures.

In the case of the bus, ferry, light rail, Sydney trains and metro modes, usage can be derived by tracking Opal Card tap-on and tap-off. However, quantitative assessment of for walking and cycling trips is challenging as the start and end of each trip triggers no measurable event. Piezo-electric and pneumatic tube counters were deployed on some cycleways to count cycle usage but out of 100-plus cycleways and walkways delivered since 2017, usage of only 14 permanent and nine pop-up cycleways was being tracked daily in July 2021. Some of the counters were broken and could not be fixed as their maintenance contracts had expired. While the Covid-19 pandemic gave Active Transport a once-in-a-generation impetus, the ability to measure the walking and cycling volumes in the Greater Sydney and Regional and Outer Metropolitan areas did not exist. To assess demand for the infrastructure across the state was similarly challenging in the absence of any sensors while any surveys used were annual or at best biannual, thus did not provide any real-time pulse of the walking and cycling demographic. This meant projects were being approved despite the absence of underlying data for both usage by and voice of customers. Active Transport's data and analytics roadmap was finalised in July 2021 and the Data and Analytics program launched to deliver it.

The objectives of the Data and Analytics program were (i) to use data for benefits tracking and (ii) provide data-driven decisioning capability. The program delivered portfolio, program and product governance; a consolidated view of Active Transport's portfolio replacing multiple spreadsheets; SAFe processes (daily standups, iteration planning, backlog grooming, iteration reviews, PI planning, portfolio sync meetings); benefits tracking across the state using piezo and pneumatic tube sensors; the ability to manage walking and cycling grants; ways of working and flow metrics (end-to-end delivery of projects).

The Active Transport portfolio used the SAFe (Scaled Agile, 2023) framework for portfolio and product management and incorporated lessons from previous DSIs. The key observations were (i) funding from an unused business case was used initially to fund proof of concepts; (ii) the business case was product-centric, not project-centric; (iii) the DSI required large investment in sensors and CRM software to capture data was embedded in the wider infrastructure business case and (iv) the exploratory nature of DSIs was accepted by management.

Based on the outcomes delivered, I summarise our initial observations as follows. Utilising product management for DSIs shifts the focus from temporary project-based approaches to a more continuous and iterative product-centric mindset. Unlike projects that have defined start and end dates, DSIs benefit from ongoing enhancements and adaptations, making product management a more suitable approach for driving long-term value and sustained innovation. This led us to our first guiding principle of *“using product management for DSIs and not project management for delivery”*.

By funding multi-year value streams instead of individual projects, organisations can better predict costs and allocate resources strategically. This approach acknowledges the inherent complexity and uncertainty of DSIs, allowing for flexibility and continuity in funding to support ongoing development and evolution of data capabilities over time. This led us to our second guiding principle of *“instead of funding projects using business cases, fund the multi-year value streams to better predict costs”*.

Recognising the exploratory nature of DSIs, it is essential to assign benefits that are SMART (Specific, Measurable, Attainable, Relevant, and Time-based). This ensures that business cases accurately reflect the potential outcomes and impact of DSIs, enabling stakeholders to make informed decisions based on realistic expectations and achievable goals. This led us to our third

guiding principle of “*considering the exploratory nature of DSIs, assign only those benefits which are SMART (Specific, Measurable, Attainable, Relevant and Time-based)*”.

Decentralising decision-making empowers teams with the most knowledge to make informed decisions on detailed scope, dataset selection, business rules, and visualisation techniques. This approach fosters agility, innovation, and ownership of outcomes, enabling teams to respond quickly to evolving requirements and deliver value more effectively. This led us to our fourth guiding principle of “*allow decision-making to be done by people who have the most knowledge and not based on hierarchy such as decision on detailed scope, datasets to ingest, business rules and visualisation*”.

Establishing a new category in portfolio management specifically for exploratory projects within discretionary investments helps differentiate them from exploitative projects. This distinction allows organisations to allocate resources and prioritise initiatives based on their strategic alignment and risk profile, ensuring that exploratory DSIs receive the appropriate level of support and investment. This led us to our fifth guiding principle of “*establish a new category in portfolio management for exploratory projects within Discretionary investments to differentiate them from exploitative projects*”.

Collecting and leveraging relevant data for decision-making is critical for building better business cases for DSIs. By incorporating data-driven insights into the business case development process, organisations can identify opportunities, assess risks, and forecast outcomes more accurately, leading to more informed investment decisions and improved project outcomes. This led us to our sixth guiding principle of “*collect as much relevant data and use it for decision-making*”.

Agile methods are inherently more suitable for the delivery of exploratory DSIs due to their iterative and adaptive nature. By embracing Agile principles and practices, organisations can respond quickly to changing requirements, incorporate feedback from stakeholders, and deliver value incrementally, ultimately increasing the success rate of DSIs and maximising their impact on business outcomes. This led us to our last and seventh guiding principle of “*agile methods are more suitable for delivery of exploratory DSIs*”.

8.6 CONCLUSION AND RECOMMENDATIONS

In this section, I review the research questions: “*How should the business cases for DSIs be structured? Is there a difference in business cases when DSIs are managed as exploratory projects versus exploitative projects? Is there a difference in business cases when DSIs cover products already constructed versus not constructed?*” and summarise the findings of using seven Transport DSIs as case studies. I first start with findings from this chapter, discuss the limitations of the research, the implications on business managers and practitioners of the business cases and end with the recommendations.

8.6.1 Conclusion

My experiences at Transport from 2017 to 2022 have allowed me to reflect how we constructed and delivered business cases for traditional ICT investments versus those for DSIs. In various roles I have filled as program and portfolio manager in consulting, banking and finance and tourism sectors, the focus was on hard costs and benefits. Governance of benefits realisation pose challenges for ICT investments (Jenner, 2012; Standards Australia, 2016; Thorp, 2003) even when they are exploitative, with clear line of sight on the costs and benefits. The problem is compounded for exploratory investments such as DSIs when both costs and benefits cannot be quantified upfront (See Chapter 5 - Unique Characteristics of DSIs).

My first DSI Vanguard business case reflected an exploitative investment. It had benefits metrics of improved fare compliance and fine payment (Table 5-3), which neither the sponsor nor DSI delivery team could deliver. As we created subsequent business cases for CTABS, Ferry, Light Rail Priority, PTIPS Analytics and MPR, we included capability delivery in the business cases while keeping the exploratory nature of DSIs front and centre of what we were doing. I could explain the uncertainty in requirements, data and schedule to the sponsors and take them through the journey of Agile delivery. In the next five business cases mentioned above, the infrastructure and data already existed and we provided the insights. The business case of Active Transport, which was my seventh DSI, was unique in that I was both the DSI business sponsor and responsible for delivering \$1.2 billion Active Transport (walking and cycling) infrastructure across the state. The data was not being captured and limited insights were available. I included the cost of technology (sensors and software licences to capture the walking and cycling data and voice of

customer) and DSI as part of the infrastructure investment cost. I had shifted from a project to a product mindset and funded value streams delivering Active Transport infrastructure and DSI products.

Agreement by the Transport Leadership Team on 31st January 2022 for Active Transport to fund nine value streams, as shown in Figure 8-3, was a major shift in our thinking of how DSIs should be funded and delivered. The data and analytics value stream was shown to be an enabler in the delivery of all Active Transport infrastructure. This value stream included all technology and DSI investments and did not have a separate business case as its costs were included in the value streams delivering the infrastructure. I took this approach as DSIs should not be an afterthought and should be considered part of main product investment, which in the case of Active Transport, is delivering walking and cycling infrastructure.



Figure 8-3. Active Transport Value Streams

I conclude that moving from a project to a product mindset, funding value streams and acknowledging the exploratory nature of DSIs has allowed Transport to build better business cases. While the organisation and teams mature and DSIs move from being exploratory to exploitative, the exploratory aspect of DSIs is always there – only the percentage reduces. The Active Transport Business case took the value stream and product lifecycle into account such that delivery capability through skilled resources was maintained to support the DSI, not just the initial delivery.

8.6.2 Limitation and Implications of Research

The case studies included in this research used seven DSIs from one public sector organisation, Transport for NSW. Future research could include DSIs from other public and private sector organisations to validate the findings of this research or if their business cases are treated differently. While exploratory projects have been around for a long time, DSIs are a more recent phenomenon and sit in a rapidly evolving technology and delivery space. This has an impact on the currency of the research work being done and some of the guiding principles may evolve as the dial shifts for DSIs from being exploratory to more exploitative, as evidenced already in Transport and reflected in *Figure 4-1*.

The limited availability of methods and standards in the delivery of DSIs has caused business managers and program managers to chart their own path and thus introduce inconsistency in how DSIs are treated and delivered in different organisations. With the emergence of research such as this, it is expected that the standardisation on DSIs will increase and provide guidance to business managers in the development and treatment of business cases and to practitioners in the efficient delivery of DSIs in relation to when they should be managed as exploratory projects.

8.6.3 Recommendations

There is a rich literature about program management for ICT-enabled programs and proven delivery frameworks which have matured over the past three decades (Axelos, 2022; Project Management Institute, 2016, 2017b). This research makes a significant contribution to the theory and practice of the emerging data-science domain.

This research used data and observations from seven programs at Transport to provide insights into DSI business cases. It is evident that DSIs cannot have the same criteria as those used for other ICT programs in terms of business case development and approval. The benefits cannot be easily attributed to the investment and most of the DSIs are enablers. This class of programs carries a high degree of uncertainty right from the through to the closing phases. DSIs can be categorised as “exploratory projects” (Unique Characteristics of DSIs), like innovation or mining initiatives, rather than the “exploitative projects” category that the majority of ICT-enabled initiatives fall into. The nuggets or the insights to be discovered are known only when you get to them. I propose that a new category of “exploratory projects” be established within discretionary

investments. This category is not applicable to organisations where delivery of DSIs is the core-business.

The current program management literature does not adequately support delivery of DSIs and instead focuses on risk elimination and rapid delivery of business outcomes. I propose reviewing the competence of the program team and business on managing risks and uncertainties, as DSIs are best managed as exploratory projects using Agile methods instead of exploitative projects using waterfall methods. Business managers need to be able to distinguish between exploratory versus exploitative projects at the time of investment approvals.

Based on these seven Transport case studies, I recommend the seven guiding principles for building better business cases for DSIs as shown in Table 10-2. While Transport used SAFe for delivering DSIs, I believe that the guiding principles are generic and can be applied by organisations using any other Agile delivery method.

Table 8-2
Guiding Principles for Building Better Business Cases for DSIs

No	Description
(i)	Focus on product not projects – Use product management for DSIs and not project management for delivery.
(ii)	Fund value streams – Instead of funding projects using business cases, fund the multi-year value streams to better predict costs.
(iii)	Acknowledge exploratory nature – Considering the exploratory nature of DSIs, assign only those benefits which are SMART (Specific, Measurable, Attainable, Relevant and Time-based).
(iv)	Empower teams – Allow decision-making to be done by people who have the most knowledge and not based on hierarchy such as decision on detailed scope, datasets to ingest, business rules and visualisation.
(v)	Create new category for exploratory projects – Establish a new category in portfolio management for exploratory projects within Discretionary investments to differentiate them from exploitative projects.
(vi)	Data-driven decision-making – Collect as much relevant data and use it for decision-making.
(vii)	Use Agile methods for delivery – Agile methods are more suitable for delivery of exploratory DSIs.

Additional research is still required to develop approaches to delivering DSIs. This research will deliver a significant contribution to the body of knowledge for program management relevant to both literature and practitioners. Without this work, there will be more failed programs,

dissatisfied sponsors and much-needed investment in this emerging domain will be delayed, as will the benefits that will flow from harnessing the data and the nuggets in it.

9. Using DSIs for Benefits Tracking and Investment Decisioning – A Transport for NSW Case Study

This chapter is based on my published paper “Using Data Science Initiatives to deliver Smart Infrastructure and improve Customer Experience – A Transport for NSW Case Study” (Mathur, 2021b). While this chapter does not address the research questions directly, it demonstrates the value that DSIs can potentially deliver to an organisation. I describe how I used the initiatives that I implemented at Active Transport, using the DSI delivery framework, minimum viable governance and business case discussed in this thesis, as a case study.

9.1 ABSTRACT

Significant effort goes into identifying benefits in business cases for infrastructure projects. Tracking these benefits is complex and is often ignored once the infrastructure is delivered. In addition, few organisations are harnessing DSIs effectively for data-driven decision-making. This chapter describes how Transport for NSW (Transport) used DSIs to track the benefits from its investment in Active Transport infrastructure and to identify future investments, using a case-study approach.

Transport uses cameras and AI/ML to track walking and cycling patterns throughout New South Wales. It uses Strava data to model cycling trips, including kilometres, and DSpark data to model walking trips, including kilometres. The modelled counts are calibrated using camera counts. By combining these counts with per-kilometre social/health, economic and environmental benefits, the value of this Active Transport infrastructure is demonstrated. Listening to the voice of the customer and overlaying existing infrastructure onto current and potential walking and cycling usage discussed in this study identified previously missing links that could then be used for investment decisioning.

The chapter informs program and change management practitioners in how to use DSIs for benefits tracking and investment decisioning.

9.2 INTRODUCTION

Significant effort goes into identifying benefits for transport infrastructure projects; these often include social/health, economic and environmental categories, over a 20-to-30-year time horizon. Tracking these benefits is complex and thus most projects face challenges in demonstrating them once the infrastructure has been delivered, not least because organisations need both people and processes in place to carry out the task. Torres, Khemici, and Paré (2017) compare benefits realisation to a hot potato game whereby each player takes responsibility by taking a hot potato in their hand and passing it on to the next player as the project moves into a different phase of its life cycle. This paper focuses on process, specifically the data side of benefits tracking. Data limitations can hinder effective assessment, as demonstrated by a study of four cycling infrastructure investments in Glasgow City Council (Hong, McArthur, & Livingston, 2020). In addition, any decision-making relies on multiple qualitative and quantitative factors. A survey by EY (EY, 2015) found that 81% of businesses agree that data should be at the heart of all decision-making but only 31% restructure their operations to harness the power of such data.

This chapter highlights how Transport for NSW (Transport) used DSIs to track benefits in its investment in Active Transport infrastructure and to identify future investments. Transport deploys cameras on each Active Transport project to track the pre- and post-walking and cycling counts using AI/ML. It uses Strava Metro data (<https://metro.strava.com/>) to model cycling trips and kilometres (K. Lee & Sener, 2021) and DSpark telecommunications data (<https://www.dspark.com.au/>) to model walking trips and kilometres. The modelled counts are calibrated using camera counts to provide data per statistical area, council and state. By combining the counts with per-kilometre social/health, economic and environmental benefits, the value of Active Transport infrastructure is demonstrated.

Overlaying existing infrastructure with current and potential walking and cycling usage and listening to the qualitative customer sentiment through the Voice of Customer solution identifies infrastructure gaps for use in investment decisioning.

9.3 LITERATURE REVIEW

9.3.1 Data Science Initiatives (DSIs)

DSIs are defined as projects or programs that involve the application of data science techniques and methods to address complex problems, generate insights or create new products or services. The term ‘big data’ has become ubiquitous and several authors have attempted to standardise its definition (De Mauro et al., 2016; Ward & Barker, 2013). Business intelligence is an umbrella term that includes applications, tools, infrastructure and practices to enable access to and analysis of information to optimise performance and decision-making (Gartner, 2023; Halper, 2015; Larson & Chang, 2016). Data science can be defined as the collection, preparation, analysis, visualisation, management and preservation of large collections of information to generate actionable insights (Donoho, 2017; Saltz & Stanton, 2017). DSIs often include implementation of AI and ML. I see different organisations at different levels of maturity in how they exploit data for benefits tracking and investment decisioning and change management should account for where an organisation is in their maturity journey. I argue that DSIs are typically exploratory projects due to the uncertainty around benefits and the delivery itself.

9.3.2 Active Transport Metrics

Compared to other modes of transport, such as bus, ferry, light rail, heavy rail and metro, the benefits provided by Active Transport are unique as they are driven solely by the number of kilometres cycled or walked. To realise benefits from Active Transport investments, a framework was developed by Transport (B. Lee et al., 2023) to measure for per-kilometre benefits, including social/health, economic and environmental aspects, as shown in Table 9-1.

Table 9-1
Framework for per-km Active Transport benefits

Social/Health	Economic	Environmental	Future Benefits
Physical activity benefits	Vehicle operating cost savings	Air pollution	Injury benefit/disbenefit (beyond health)
Ambient air quality health benefits	Parking cost savings	Climate change	Public transport savings
Air quality exposure	Roadway provision cost savings	Well-to-tank emissions	15-min neighbourhood benefits

Road traffic injury (health)		Noise	Accessibility benefits
Congestion cost savings		Soil and water	Amenity benefits
		Nature and landscape	Property price increases
		Urban effects	Retail benefits
		Biodiversity	Tourism benefits

The framework was used to derive the financial impact of Active Transport benefits for urban and rural areas across walking, cycling and micro-mobility devices (e-Bike and e-Scooter), as shown in Table 9-2.

Table 9-2
Financial impacts for per-km Active Transport benefits for urban and rural areas

[illegible]

Provider/ maintenance savings	Roadway provision cost savings				0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Environment benefits	Greenhouse gas emissions	99-233g CO ₂ per km		✓	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	Well-to-tank emissions	-			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Noise	-		✓	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00
	Soil and water				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Nature and landscape				0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02
	Urban effects				0.01	0.01	0.01	0.01	-	-	-	-
	Biodiversity				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total	Total	-	-	-	6.52	2.25	2.11	0.67	6.21	1.94	1.80	0.36

9.4 RESEARCH SETTING AND METHODS

9.4.1 Research Methods

The research focused on showing how Transport for NSW (Transport) uses DSIs to (1) track benefits in its investment in Active Transport infrastructure and (2) to identify future investments. Such a focus requires deep engagement in the field, observing and interacting with decision-makers, business stakeholders, program managers and delivery team members. As a result, I chose a multi-methods research methodology appropriate for exploratory research when the aim was to gain familiarity with a problem or to generate new insights for future research (Eisenhardt, 1989). I chose a single case study method of a large portfolio as this provided the opportunity to enhance understanding and simultaneously enable other researchers to apply these findings in other contexts as well as the potential to generalise the findings (Straits & Singleton, 2018; Yin, 2018). The DSIs chosen for this case study were delivered between July 2021 and March 2022, bringing in-depth insights into the program life cycle. As the Active Transport portfolio is delivered, additional insights will continue to emerge.

My position in Transport gave me good access to data to conduct the case study, to understand how DSIs can assist organisations in benefits tracking – a step that is essential for good governance but not yet widely practiced. Similarly, and despite the hype around using big data, organisations fail to harness its power for decisioning. In this case study, I therefore explored investment decisioning in the context of Transport. While I led the delivery of several DSIs outside the Active Transport portfolio, for the purposes of this chapter I used only Active Transport DSIs to inform and validate the findings.

Using an interpretive research tradition associated with case studies, ontological and epistemological assumptions of DSIs emerged. I used data generated from sensors and combined them with datasets from multiple sources to calibrate for tracking benefits. The observations took place over a period of 18 months between July 2021 and December 2022.

9.4.2 Research Setting

Our research was situated within Active Transport, a sub-portfolio of a larger portfolio of Transport for NSW (Transport). It consists of engineering projects delivering walking and cycling infrastructure across New South Wales as well as DSIs for benefits tracking and supporting investment decisions. The portfolio aligns with the cycling strategy and action plan of the New South Wales Government to move towards a more sustainable Sydney (City of Sydney, 2018).

As stated in earlier chapters, the Active Transport portfolio consists of \$1.2 billion worth of infrastructure to be delivered between 2014 and 2027 across New South Wales. In the Get NSW Active (Walking and Cycling) program, Active Transport issues \$50 million to \$75 million in grants annually to councils to deliver walking and cycling infrastructure on local roads.

9.5 DATA COLLECTION AND ANALYSIS

9.5.1 Active Transport Benefits Tracking

Capturing walking and cycling counts

One of the key metrics for tracking the benefits of Active Transport infrastructure is the tally of people using it. Up until October 2021, Transport used piezo-electric counters for permanent cycleways and pneumatic tube counters for pop-up cycleways. A proof of concept was done using smart cameras with AI/ML to determine count of cyclists and pedestrians in late 2021. The solution used a Meraki camera with Cisco Edge Computing devices that provided a unique count at multiple sites in the camera frame every 30 seconds, as shown in *Figure 9-1*. Testing revealed that there was insufficient computing power in the device and the count was lower than the actual usage. The processing unit was upgraded and it improved the accuracy level to 95% for sites closer to the camera. This solution is now being implemented for each Active Transport project to capture walking and cycling counts at key locations.

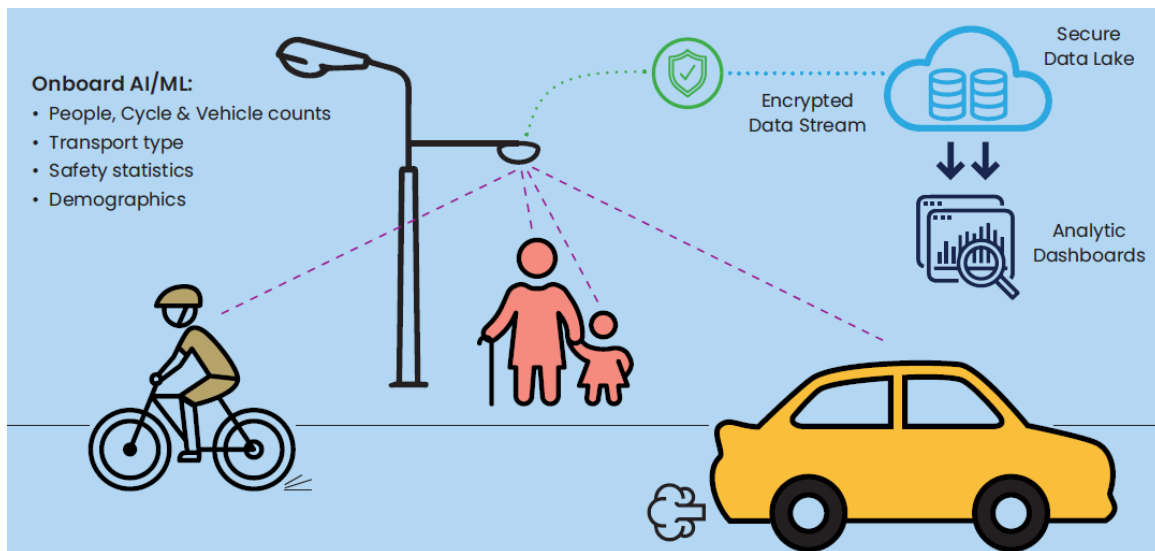


Figure 9-1. Camera with AI/ML to capture cycling and walking counts

Key features of the solution were:

- Configuring multiple sites with one camera at a complex intersection,
- Capturing counts of more than 20 defined vehicle types. The solution could be trained to identify new vehicle types such as e-Scooters and cargo bikes,
- Near real-time counts and availability of data on an Azure platform every 30 seconds, and
- Sharing analytics with internal Transport and external council and public users simultaneously.

Privacy concerns were addressed by not storing the camera footage and importing only the data counts to the Azure platform for visualisation and benefits tracking.

Modelling Benefits across NSW

In Transport, daily/weekly patronage is an important metric for all modes - bus, ferry, light rail, metro and trains. Patronage is available via Opal, a contactless fare collection system for public transport services, or via credit card. However, in relation to Active Transport, there is no clear start and end of a trip and accessing this metric is a unique challenge. We used Strava data to model cycling trips and kilometres (Sabeti, 2021) and DSpark data provided by Optus Telecommunications to model walking trips and kilometres. Figure 9-2 shows a schematic flow chart of the data integration process and the ML model inputs and outputs used to calculate cycling trips

and kilometres and *Figure 9-3* shows a visualisation of modelled cycling volumes across the Sydney metropolitan region.

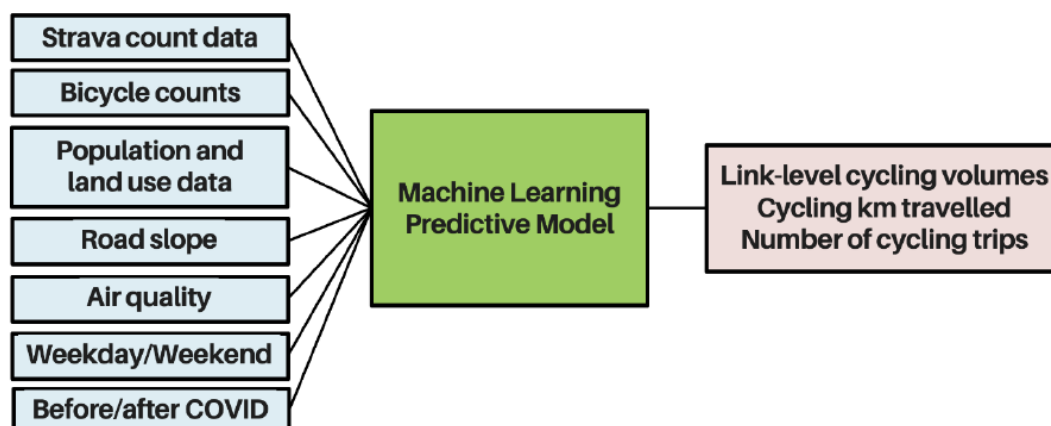


Figure 9-2. ML model inputs and outputs for cycling trips and kilometres (Saber, 2021)

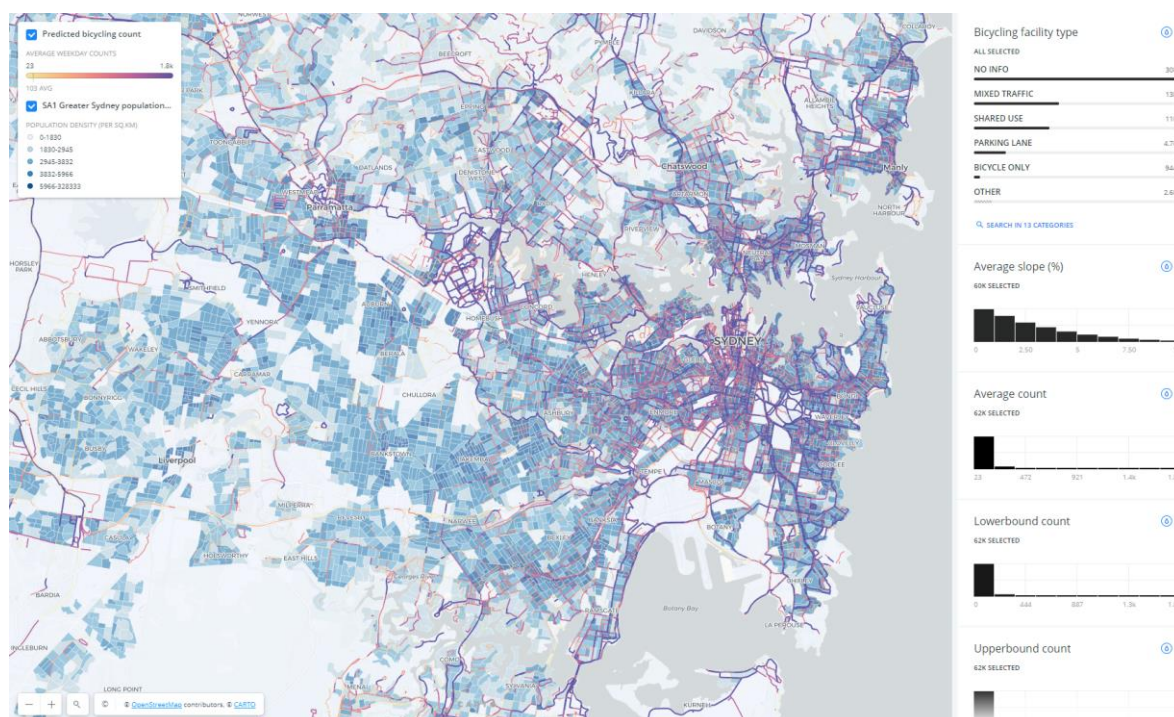


Figure 9-3. Average weekday cycling counts in Sydney metropolitan region

Like the cycling model, *Figure 9-4* shows a schematic flow chart of the data integration process and the ML model inputs and outputs used to calculate walking trips and kilometres.

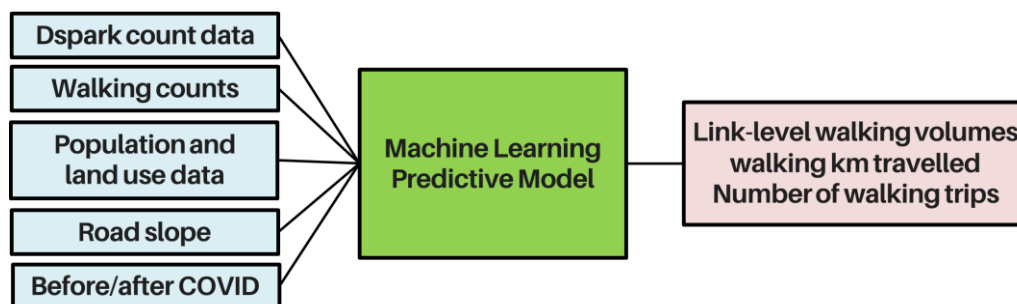


Figure 9-4. ML model inputs and outputs for walking trips and kilometres

By using benefit multipliers for walking and cycling from Table 9-2, we could model the Active Transport benefits across statistical areas, local government areas and across the state.

9.5.2 Active Transport Investment Decisioning

Multiple qualitative and quantitative factors were used to decide where new walking and cycling infrastructure should be built. The decision-making takes place at both local councils and centrally at Transport. Active Transport provides two datasets – Voice of Customer to measure customer sentiment and Cycleway Finder to assist cyclists in route planning.

What the customers are saying

A Voice of Customer solution was implemented to undertake a social and digital media scan of all walking and cycling mentions in New South Wales. The solution provides an “Always On” real-time Active Transport pulse and calculates a walking, cycling and overall Active Transport sentiment score. The aggregate scoring, as shown in Figure 9-5, includes real-time surveys and tracking of walking and cycling mentions in social media as well as the integration of historical offline questionnaires.

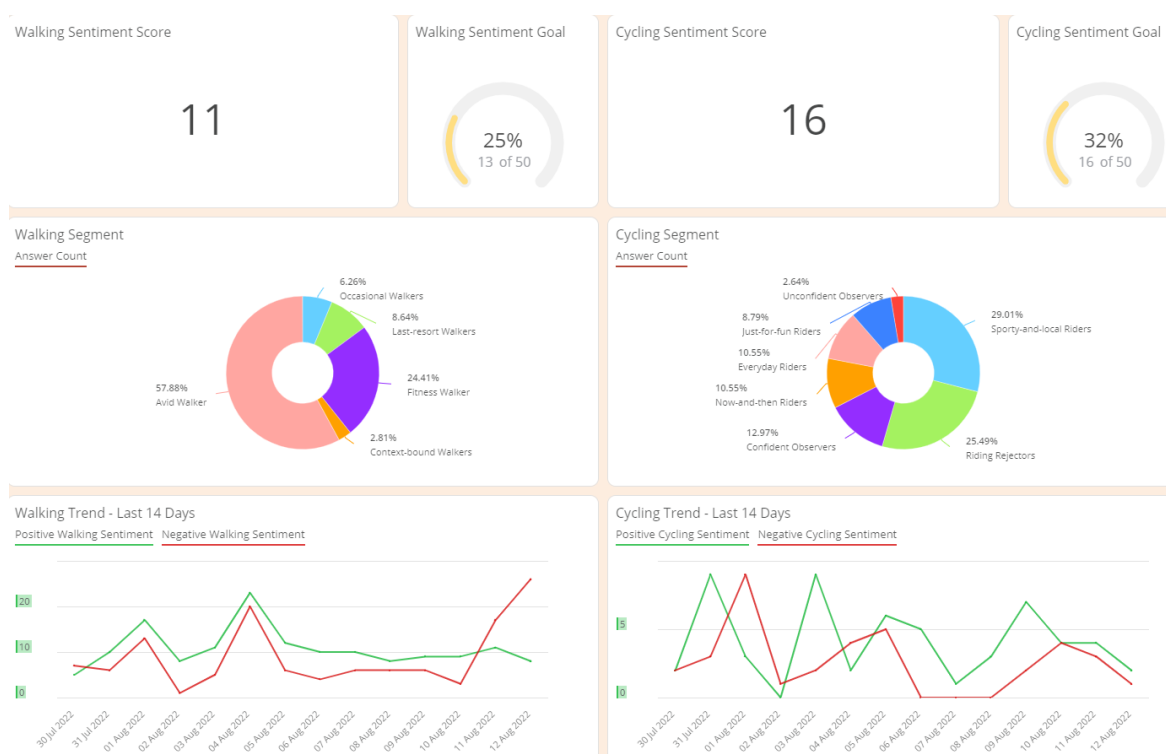


Figure 9-5. Active Transport Voice of Customer

To calculate the walking sentiment score, we took the total percentage of people who had a positive sentiment and subtracted the percentage of people with a negative sentiment of all people who answered walking questions or mentions that were included in the walking social listening topic in the Voice of the Customer solution. The scores ranged from 100 (everyone positive, no negative) to -100 (all negative, no positive). This methodology ignored the neutrals and we set the sentiment target score at 50 as a high benchmark. The cycling sentiment score was calculated similarly, with people answering that they cycle or mentions included in the cycling social listening topic in the Voice of the Customer solution. The commuter sentiment score combined both walking and cycling sentiment and was the total percentage of positive walking and cycling minus the total percentage of negative walking and cycling.

This Voice of the Customer provided us not only with the trending sentiment scores but also customer comments about Active Transport infrastructure, which allowed new projects to be added to the portfolio pipeline or identified infrastructure requiring maintenance.

Where are the gaps in the network for walking and cycling?

Our Cycleway Finder (*Figure 9-6*) allowed us to see where gaps existed in the bike riding corridor network. When this information was overlaid with the heat maps of modelled cycling usage, it became a powerful tool for investment decisioning.

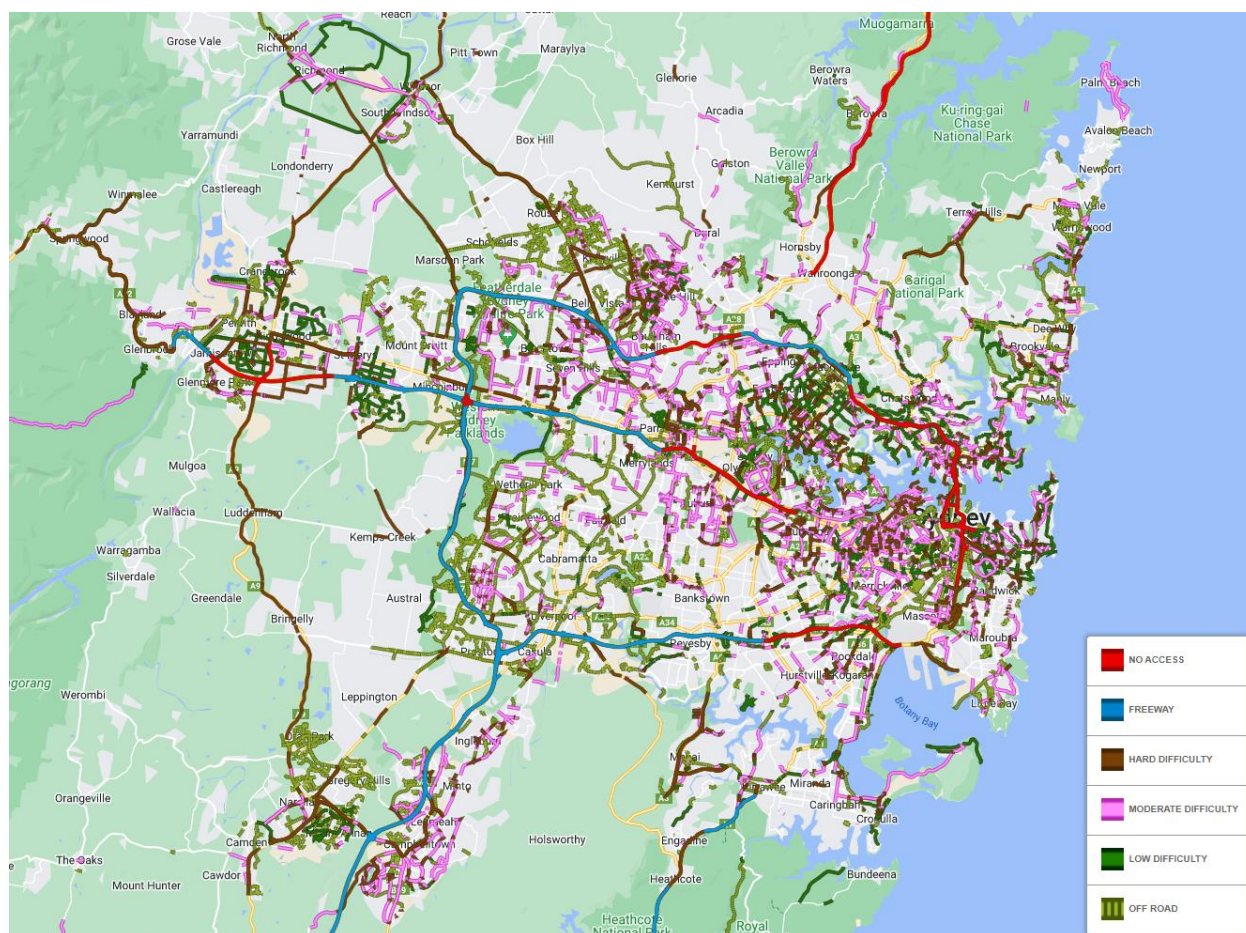


Figure 9-6. Cycleway Finder

Similarly, heatmaps of movement data, as shown in *Figure 9-7*, enabled us to assess gaps in the walking and mobility ecosystem. While we did not have a record of every footpath that exists in New South Wales, big data analysis of movement allowed us to identify investment opportunities to promote walking and to create 15-minute neighbourhoods, which deliver a more connected, liveable and walkable community.

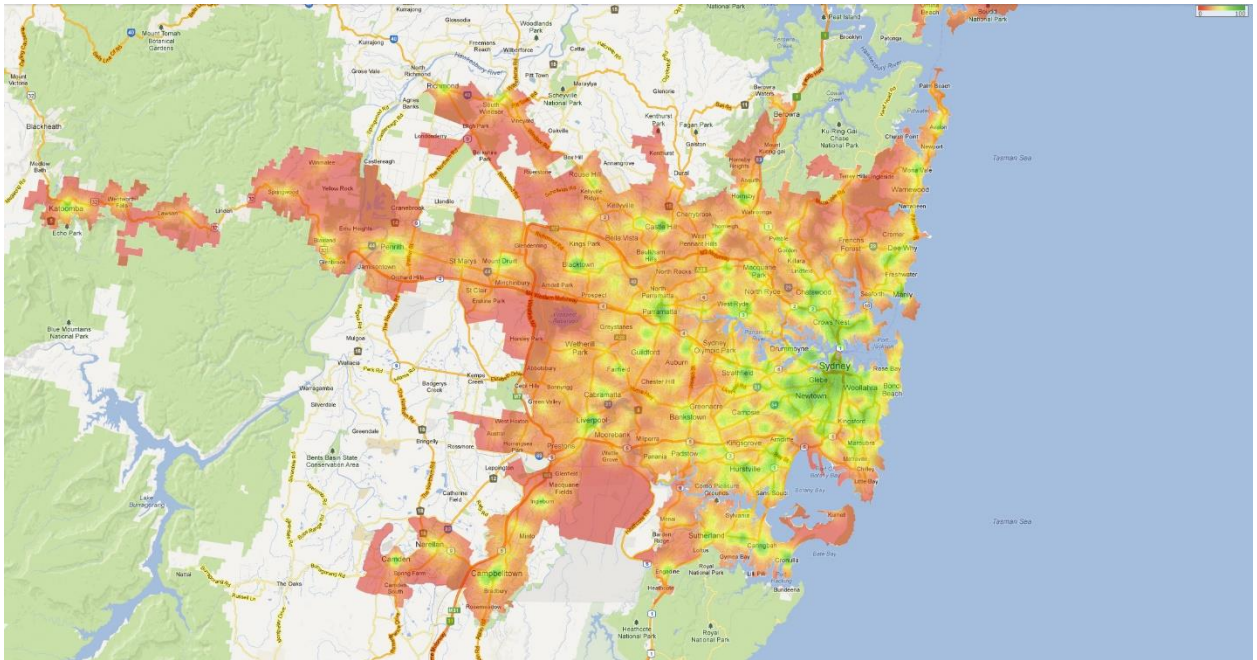


Figure 9-7. Mobility Heatmap of Sydney

9.6 CONCLUSION AND RECOMMENDATION

9.6.1 Conclusion

Despite the hype around data science, organisations struggle to use it effectively in delivering outcomes. This chapter highlights how Transport used DSIs to track benefits from investments in Active Transport infrastructure and to identify gaps in the network and hence future investments. Most of the infrastructure projects had previously faced challenges in demonstrating any benefits in their business cases. Transport used DSIs to overcome this challenge and to demonstrate the benefits on a per-day basis across statistical areas, councils and the state.

Notwithstanding their limitations, the typical 18-to-24-month lead times for infrastructure project delivery and the benefits modelled make the case for Active Transport investments strong.

9.6.2 Limitations and Implications of this Research

This research used DSIs from a public sector organisation in Australia as a case study to identify challenges in using data and DSIs for benefits tracking and investment decisioning. Future research can include validating the challenges with other public and private sector organisations delivering DSIs in other countries.

The modelling approach using Strava and DSpark datasets had limitations; the Strava app is used by certain demographics of cyclists and thus does not represent the entire population. We used about 100 existing counters and cameras to calibrate the Strava data but the accuracy of the model will improve when additional counters and cameras are deployed on the network. Similarly, DSpark data excludes all trips made by mobile phone users under the age of 16 and excludes recreational trips where the start and end points are the same. Additional cameras will help us calibrate the DSpark data for higher accuracy.

9.6.3 Recommendation

The findings presented in this chapter underscore the transformative potential of DSIs in addressing the longstanding challenges of benefits tracking and investment decisioning within infrastructure projects, particularly in the realm of Active Transport. Leveraging DSIs, Transport successfully tracked the benefits accrued from investments in Active Transport infrastructure while simultaneously identifying critical gaps in the network for future investments. However, to fully capitalise on these insights and drive meaningful change, several recommendations emerge:

- (i) **Refinement of Benefits Tracking Mechanisms:** Transport should continue refining its benefits tracking mechanisms to ensure accuracy, granularity, and comprehensiveness in capturing the diverse range of benefits associated with Active Transport infrastructure. This may involve further calibration of models, integration of additional data sources, and validation exercises to enhance the reliability of benefit estimations.
- (ii) **Integration of Real-time Data Streams:** Embracing real-time data streams can significantly enhance the timeliness and relevance of benefits tracking efforts. Transport should explore opportunities to integrate real-time data sources, such as IoT sensors, mobile applications, and social media sentiment analysis, to capture dynamic shifts in usage patterns, user sentiments, and emerging trends that could influence investment decisions.

(iii) **Stakeholder Engagement and Collaboration:** Effective benefits tracking and investment decisioning require active engagement and collaboration with stakeholders across government agencies, local councils, community groups, and the general public. Transport should foster a culture of collaboration, transparency, and information sharing to leverage collective expertise, insights, and resources in optimising Active Transport infrastructure investments.

(iv) **Continuous Monitoring and Evaluation:** The journey towards optimised benefits tracking and investment decisioning is iterative and ongoing. Transport should establish robust mechanisms for continuous monitoring and evaluation to assess the effectiveness of DSIs, identify areas for improvement, and adapt strategies in response to evolving challenges and opportunities.

(v) **Capacity Building and Knowledge Sharing:** Building internal capacity and expertise in data science, analytics, and decision support methodologies is essential for sustaining the momentum gained through DSIs. Transport should invest in training programs, knowledge sharing initiatives, and partnerships with academic institutions and industry experts to empower staff with the requisite skills and knowledge to harness the full potential of data-driven decision-making.

(vi) **Alignment with Strategic Objectives:** Finally, it is imperative to ensure alignment between benefits tracking outcomes, investment decisions, and broader strategic objectives related to sustainability, public health, economic development, and social equity. Transport should integrate benefits tracking processes into strategic planning frameworks to ensure that investment decisions are guided by overarching organisational goals and societal priorities.

By adopting these recommendations, Transport can further leverage DSIs to drive evidence-based decision-making, optimise resource allocation, and maximise the societal impact of Active Transport infrastructure investments. This holistic approach will not only enhance the efficiency and effectiveness of benefits tracking and investment decisioning but also pave the way for a more sustainable, equitable, and resilient transportation system for the future.

10. Discussion and Conclusion

10.1 INTRODUCTION

This thesis has described the challenges I faced while delivering DSIs at Transport for NSW. As a result of my investigation, it contributes a body of knowledge that can be used by practitioners, business managers and academics to overcome uncertainties and complexities in delivering DSIs. Using the key and supplementary research questions, I have addressed the knowledge gap and extended the literature on program management and data science. The contributions made by this research are summarised in Table 10-1 and are explained in more detail in the section that follows.

Table 10-1
Summary of Research Contributions

Research Issue	Status of Research Issue in the Extant Literature	Contribution of this Research
RQ - Why do DSIs face challenges delivering envisaged value when using traditional processes for managing ICT-enabled programs? How do these challenges manifest within the delivery process?	The uncertainty in DSIs is not evident and acknowledged in the current literature and the traditional methods for <i>exploitative projects</i> used to deliver DSIs.	An addition: Identified unique characteristics of DSIs and categorised them as <i>exploratory projects</i> . (Refer Chapter 5. Unique Characteristics of DSIs)
SQ1 - What design principles should be incorporated in a DSIs delivery framework so that program managers can adopt a predictive path to realise value from such investments?	Current literature and practitioner guides support delivery of ICT-enabled projects but do not adequately support DSIs, causing a high failure rate of up to 85%.	An addition: Proposed a new integrated DSI delivery framework to cater for inherent risk and uncertainty in DSIs consisting of five domains - program management, change management, data science, data management and scaled agile domains.

*(Refer Chapter 6.
Framework to Manage
DSIs)*

SQ2 - What is the minimal viable governance required for DSIs to ensure they deliver envisaged value efficiently and effectively?

Sufficient literature exists to support corporate, portfolio, program and agile governance though limited in managing innovation and exploration and addressing dynamic nature of governance.

An extension: Extended the Goldilocks Theory of Governance and proposed guiding principles for minimum viable governance of DSIs.

(Refer Chapter 7. Minimum Viable Governance for DSIs)

An extension: Extended the four-paradigm governance model to DSIs to reflect maturity of delivery organisation and dynamic nature of governance.

(Refer Chapter 7. Minimum Viable Governance for DSIs)

SQ3 - How should the business cases for DSIs be structured?

The current literature supports business cases for programs and products with finite scope, cost and schedule adequately but is limited in supporting product life cycle.

An extension: Extended product mindset and product life cycle to DSIs and proposed guiding principles for building better business cases for DSIs.

(Refer Chapter 8. Insights into Building Better Business Cases for DSIs)

Other Contributions

Significant investment is taking place in DSIs but good uses showing value are limited.

Researchers face challenges in dissemination and impact of their

An advance: Demonstrated how DSIs can be used for benefits tracking.

(Refer Chapter 9. Using DSIs for Benefits Tracking and Investment Decisioning)

work.

– *A Transport for NSW
Case Study*)

An advance: Created a microcredential course “Management and Governance of DSIs” at UTS to share this research with researchers and practitioners.
(Refer Section 10.4.1. Establishment of Management and Governance of DSI Course)

The response to the key research question (RQ) highlights the main contribution this thesis makes to the literature on program management, namely, the management and governance of DSIs as exploratory projects (Lenfle, 2008). In the emerging field of data science, identification of the unique characteristics of DSIs has important implications for research and practice throughout their life cycle. This research used data and observations from seven programs at Transport as case studies to outline the unique characteristics of DSIs. Section 10.2.1 shows that, depending on the maturity and competence of the delivery organisation, DSIs can be categorised as “exploratory projects” (Lenfle, 2008; Maniak, Midler, Lenfle, & Le Pellec-Dairon, 2014) similar to innovation initiatives, rather than “exploitative projects”, into which the where majority of traditional ICT-enabled initiatives fall.

In response to the supplementary question SQ1, I propose a new integrated DSI delivery framework with delivery methods that caters for their inherent risk and uncertainty. Section 10.2.2 shows how DSIs can be managed using this delivery framework. Supplementary question SQ2 allowed me to extend the Four-Paradigm Governance Model (Müller, 2009, 2016) to DSIs. Governance of exploratory projects often requires processes and a mindset different from that of exploitative projects. Section 10.2.3 shows how the Versatile Artist and Agile Pragmatist paradigms of the framework can be used for the governance of DSIs, depending upon the maturity of the delivering organisation. In response to supplementary question SQ3, I provide

eight guiding principles on how DSI business cases could be constructed and highlight that a shift from project-to-product life cycle suits DSIs better. Section 10.2.4 shows how adopting a product mindset helps in the structuring of DSI business cases.

In this final chapter, I also discuss how investigation into the management and governance of DSIs as exploratory projects contributes to theory and practice, has important implications for practitioners, business managers and academics and provides useful avenues for future research. The thesis then concludes by reflecting back on the motivation for the research questions.

10.2 CONTRIBUTIONS TO THEORY

10.2.1 Classification of DSIs As Exploratory

The case studies show that due to the uncertainties they carry throughout their life cycle, DSIs generally cannot be managed as exploitative initiatives. Chapter 5 demonstrated how case studies of seven DSIs spanning six years combined with semi-structured interviews with practitioners helped me identify six unique characteristics of DSIs. One key characteristic is that DSIs carry a high degree of uncertainty from initiation through to their closing phases, except for when the data is from a well-defined and structured source. While the high-level target outcomes of a DSI are clear, the how-to is largely unknown because the data attributes, data quality and business rules of a DSI are often unknown at the start and only when a data discovery exercise is undertaken using sample data can certainty be enhanced.

There is growing concern that applying an incorrect approach to project management in the face of uncertainty is both common and detrimental to their performance. Atkinson et al. (2006) raised the concern that common project management practice does not address fundamental sources of uncertainty, especially in “soft” projects where flexibility and tolerance of vagueness are necessary. They attributed uncertainty to a lack of information, ambiguity, the characteristics of project parties, trade-offs between trust and control mechanisms and varying agendas during the different stages of a project’s life cycle. As I delivered the seven DSIs, I saw this uncertainty and ambiguity throughout; they prevented me from creating detailed schedules and left me to operate in an environment of vagueness. The scope of future

sprints could be decided only after the team had reviewed the characteristics of the new dataset.

Innovative and exploratory projects face challenges when organisations attempt a standardised and routinised approach to manage uncertainty instead of applying an adaptive model and treating innovative projects as “voyages of discovery” (Davies et al., 2018, p. 969). As my attempt to deliver the first DSI, called Vanguard, using traditional ICT delivery methods in which I was so experienced failed, I had to adopt a path that required continuous discovery and regular re-planning. Lenfle (2008) characterised exploratory projects as projects for which neither the goals nor the means of attaining them are clearly defined from the outset and then outlined five management principles to deliver them. When I carried out the semi-structured interviews in my research to identify the unique characteristics of DSIs, the lack of clarity around goals and the path to attain them was acknowledged by all survey participants. CH Loch et al. (2011) have raised concerns about using standard risk management methods for novel projects with unforeseeable uncertainties, complexities and unknown unknowns. They argue that traditional and heavyweight project management models requiring detailed upfront planning and schedules are not suitable for delivering novel innovation projects, which are often characterised by divergence and unforeseeable uncertainties. After attempting a detailed schedule for my first DSI, I gave up rigorous front-end planning and took on a more adaptive approach to planning and delivery.

In more recent debates on project management approaches, Mahmoud-Jouini, Midler, and Silberzahn (2016) propose design thinking as a possible solution to address the innovative context of projects and they suggest three imperatives for project management: managing the explorative phase, managing the involvement of stakeholders in the projects and managing the project in relation to the strategising process of the firm. Van de Ven (2017) explains that the process of innovation is messy and complex and cannot be reduced to a simple sequence of stages or phases, as is implied by the stage gate model used in many organisations for managing innovation. Similarly, my research into a classification of innovation projects (Shenhar, 2001; Shenhar et al., 2004) also supported a view that one project management approach does not fit all projects. Gartner’s study indicating that 85% of big-data projects fail (Asay, 2017) highlighted the problem in DSI management of addressing their

uncertain and exploratory nature and the outcomes expected by the business managers. This was confirmed by the seven case studies discussed in this thesis which point to the empirical knowledge gap about how DSIs should be managed and governed in this rapid growth area of investments by organisations that want to gain strategic advantage using data.

In response to the key research question, I identified in Chapter 5 six unique characteristics of DSIs, which are summarised in Table 10-2.

Table 10-2
Unique Characteristics of DSIs

No	Description
(i)	DSIs carry a high degree of uncertainty from initiation through to their closing phases, except for when the data is from a well-defined and structured source.
(ii)	DSIs are often enablers for decision-making and may not have a direct benefit contribution.
(iii)	Neither the goals nor the means of attaining them are clearly defined from the outset of a DSI, with the caveat that as the market matures, the emergence of pre-built solutions will reduce the uncertainty.
(iv)	DSIs are not independent of each other and each one acts as an enabler to the next one.
(v)	The skills required to deliver a DSI are different from those required for a typical ICT program.
(vi)	DSIs do not end and after initial delivery but convert into managing the product, model and data.

Taking into account these unique DSI characteristics and the uncertainties in their life cycle, I propose categorising DSIs as “exploratory projects” (Lenfle, 2008; Maniak et al., 2014), similar to innovation initiatives, rather than as “exploitative projects”, into which a majority of ICT-enabled initiatives generally fall. I also acknowledge that exploration-to-exploitation is a continuous spectrum, as shown in *Figure 4-1 Transport DSIs Chronology*, in which DSIs operate depending on the maturity and competence of the delivery organisation. As elaborated in sections 10.2.2, 10.2.3 and 10.2.4, this categorisation has a flow-on effect on how DSI business cases are developed, how DSIs are managed and governed, how the benefits realisation takes place and, ultimately, how the success rate of DSIs is measured.

10.2.2 Management of DSIs

As 85% of big data projects are predicted to fail (Asay, 2017), DSI management requires methods and frameworks not currently available from the literature and best practices that support their exploratory and innovative nature. Therefore, developing a

suitable framework to develop and deliver DSIs could improve the probability of their successful delivery.

In response to supplementary question SQ1, I reviewed frameworks used for program management (Artto et al., 2009; Axelos, 2022; Project Management Institute, 2017b; Thiry, 2016), change management (Hiatt, 2006; J. Kotter, 2007; Kübler-Ross, 2009; Lorenzi & Waterman, 1985), scaled agile (Ambler & Lines, 2016; Kniberg & Ivarsson, 2012; Larman & Vodde, 2016; Scaled Agile, 2023; Sutherland, 2019), data management (Earley, 2017) and data science (Azevedo & Santos, 2008; Chapman et al., 2000; Fayyad et al., 1996; Foroughi & Luksch, 2018; Mason & Wiggins, 2010; Rollins, 2015; SAS Institute, 2009; Severtson et al., 2017). I found each of these could play a role in improving DSI delivery.

The case studies from Transport, as shown in *Figure 4-1*. Transport DSIs Chronology, demonstrate that process integration of the five domains— program management, change management, data science, data management and scaled agile domains, could contribute to the successful delivery of DSIs. Such integration also allows for the gradual transition from exploration to exploitation as organisational maturity in managing DSIs improves. The research also demonstrates that exploration always remains even though the percentage reduces with maturity. This implies that there is always going to be some uncertainty about the scoping, costing, scheduling and data quality of data in DSI management. This has implications for practitioners delivering DSIs and business managers who want to realise benefits from investment in DSIs. The research identified that traditional project and program management frameworks did not provide adequate support to practitioners and business managers to manage this uncertainty and therefore a progressive elaboration of scope is required throughout a DSI lifecycle.

Having a framework that addresses the uncertainty and innovative nature of DSIs could help improve the success rate of delivery. In Chapter 6 - Framework to Manage DSIs, I proposed an integrated framework, as shown in *Figure 10-1* DSI Delivery Framework.

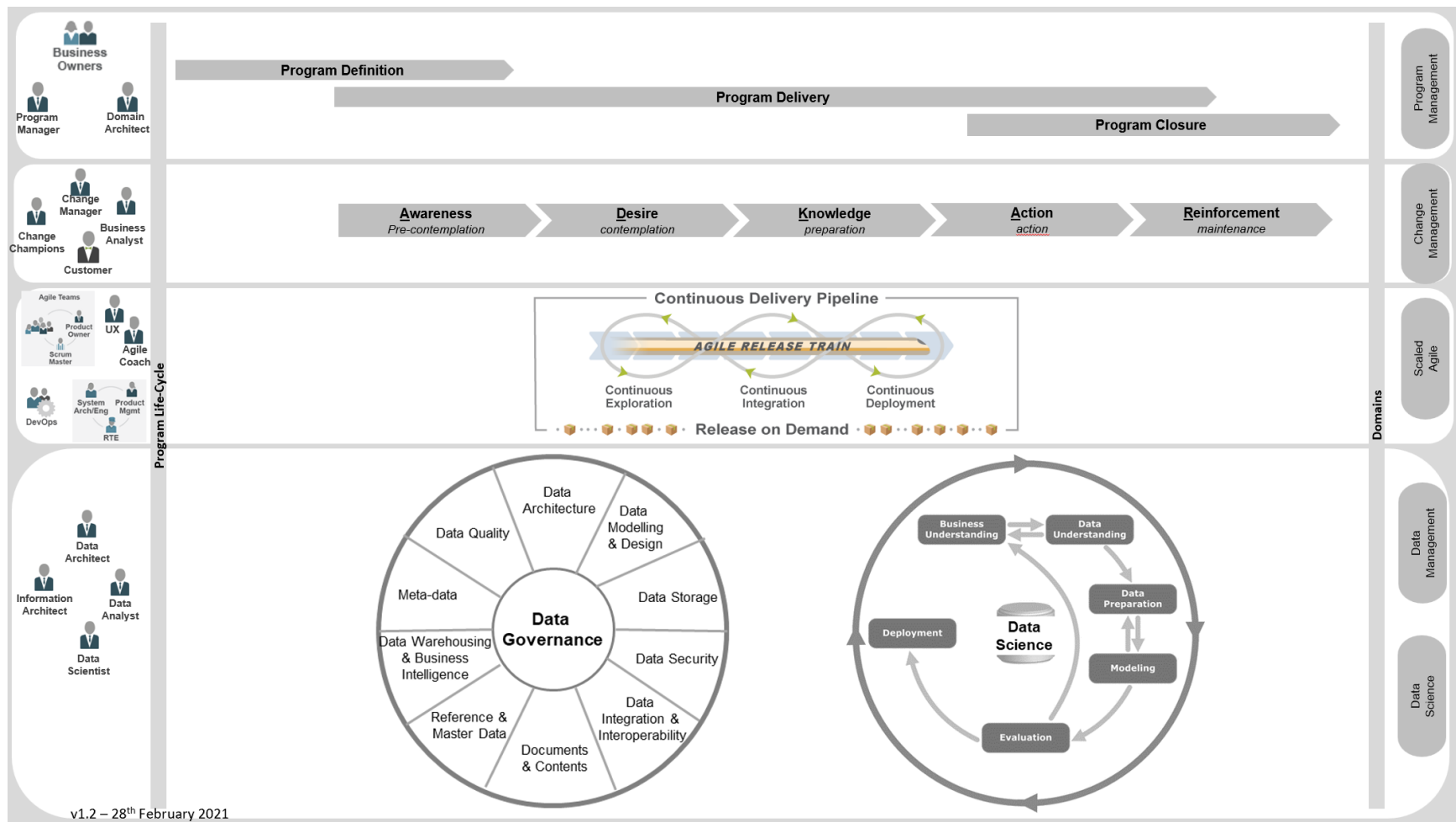


Figure 10-1. DSI Delivery Framework

I also provide a summary of DSI delivery processes in the framework developed across the program phases, as shown in Table 10-3 *DSI Delivery Framework – Summary of Key Processes across five domains and program phases*.

Table 10-3
DSI Delivery Framework – Summary of Key Processes across five domains and program phases

Domain	Program Phases		
	Definition	Delivery	Closure
1. Program Management	Program Planning		Program Closure
2. Change Management	Change Management Planning	System Demo (Showcase)	
3. Scaled Agile		- PI Planning - Continuous Exploration - Continuous Integration - Continuous Deployment - Release on Demand	
4. Data Management	Data Management Planning	- Implement Data Governance - Establish Data Architecture - Data Modelling & Design - Manage Data Storage & Operations - Implement Data Security - Implement Data Integration & Interoperability - Implement Document & Content Management - Publish Reference & Master Data - Implement Data Warehousing & Business Intelligence - Implement Metadata Management	

- Implement Data Quality
5. Data Science
- Business Understanding
 - Data Understanding
 - Data Preparation
 - Modelling
 - Evaluation
 - Deployment

In addition, I identified the specific skills required in each of the program management, change management, scaled agile, data management and data science domains, as shown in Table 10-4 *Roles and Responsibilities across DSI Domains*. Depending upon the type of DSI, the skill level and amount of time required for each role varies but all are essential for successful delivery. For example, when the Azure platform for DSIs was not set up, the PTIPS Analytics DSI team needed a dedicated full-time domain architect. Once the infrastructure was in place, the cloud and security engineers provided the ongoing platform support and the MPR and Active Transport DSIs did not need a dedicated domain architect on the team. This example illustrates that the skills required for DSIs vary based on the maturity of the organisation and requires both practitioners and business managers to be aware of this requirement.

Table 10-4
Roles and Responsibilities across DSI Domains

Domain	Role	Responsibility	DSI Essential	ICT Essential
Program Management	Business Owner	Has the primary business and technical responsibility for governance, compliance, and return on investment (ROI) for a solution.	Yes	Yes
Program Management	Product Manager	Owns vision and roadmap and defines features and releases.	Yes	Yes
Program Management	Program Manager	Responsible for overseeing the achievement of larger organisational goals by coordinating efforts between different projects without managing any one of them.	Yes	Yes
Change Management	Customer	Buyers of a solution. Internal customers are part of the enterprise whereas external customers are outside the enterprise and can be Business-to-Business (B2B), Business-to-Professional (B2P) or Business-to-Consumer (B2C).	Yes	Yes
Change Management	Change Manager	Leads change management and adoption of product and processes.	Yes	Yes
Change Management	Change Analyst	Supports Change Manager by deep-diving into stakeholder identification, impact assessment and taking them through the change journey.	Yes	Yes
Scaled Agile	Domain Architect	A specialist with deep knowledge within a particular domain of their expertise. A domain could be 'Data Services,' 'Process Design,' 'Integration Services', 'Domain Expert for SAP', etc.	Yes	Yes
Scaled Agile	Product Owner	Responsible for defining and prioritising stories to streamline the execution of program priorities while maintaining the conceptual and technical integrity of the features or components.	Yes	Yes
Scaled Agile	Scrum Master	Servant leaders and coaches for an agile team who help remove impediments and foster an environment for high-performing team dynamics, continuous flow, and relentless improvement.	Yes	Yes
Scaled Agile	Business Analyst	Guides businesses in improving processes, products, services and software through data analysis. Acts as an interface between IT and the business to help bridge the gap and improve efficiency.	Yes	Yes
Scaled Agile	Tester	Responsible for the quality of software development and deployment and performing automated and manual tests to ensure the software is fit for purpose.	Yes	Yes
Scaled Agile	Cloud Engineer	Responsible for duties associated with cloud computing, including design, planning, management, maintenance and support. Can encompass a few different roles such as cloud architect, cloud software engineer, etc.	Yes	No
Scaled Agile	Security Engineer	Identifies threats and vulnerabilities in systems and software, develops and implements solutions to defend against hacking, malware and ransomware, insider threats and all types of cybercrime.	Yes	No
Data Management	Data Owner	Has the authority and accountability for the information assets. Decides who has the right to access and edit data and how it's used and be responsible for overseeing and protecting a data domain.	Yes	No
Data Management	Data Custodian	Responsible for the safe custody, transport, storage of the data and implementation of business rules and is generally a technical role. Can also be called Database Administrator (DBA), Data Modeller, ETL Developer. Acts as the proxy for the Data Owner where Data Owner is external.	Yes	No
Data Management	Data Steward	Responsible for data content, context, and associated business rules and is generally a business role.	Yes	No
Data Science	Information Architect	Implements information structure, features, functionality, UI and focuses on structural design and implementation of an infrastructure for processing information assets.	Yes	No
Data Science	Data Architect	Responsible for data architecture and data integration and may work at the enterprise level or functional level. Work on the structural design of an infrastructure specific to collecting data, pulling it through a lifecycle and pushing it into other meaningful systems.	Yes	No
Data Science	Data Scientist	Analytical data expert who has the technical skills to solve complex problems and curiosity to explore what problems need to be solved. Part mathematician, part computer scientist and part trend-spotter straddling both the business and IT worlds.	Yes	No
Data Science	Machine Learning (ML) Engineer	Focuses on researching, building and designing self-running artificial intelligence (AI) systems to automate predictive models and deploy into production environment.	Yes	No
Data Science	Data Engineer	Works with multiple databases to capture and process live, streaming and distributed data. Designs and develops data collection, management, and search-and-retrieval systems in order to support the collection, processing, exploitation, analysis and dissemination of complex datasets.	Yes	No
Data Science	Data Analyst	Responsible for providing descriptive statistics, probability models, and other quantitative assessments of raw, processed, and generated data. Employs a combination of traditional statistical and machine learning/artificial intelligence techniques to analyse complex data sets in support of analytics, collection and managerial activities.	Yes	No
Data Science	Data Communications Specialist	Responsible for communicating and presenting summaries of structured and unstructured data in visual, text-based and interactive formats. Requires strong technical knowledge for implementing data visualizations using technologies such as PowerBI, Tableau, etc.	Yes	No

The framework, processes and skills contribute to a comprehensive identification of how DSIs can be delivered to deliver their intended benefits. In Chapter 9, I demonstrated how applying aspects of the framework I developed did deliver the intended benefits in Active Transport case study.

While the DSI delivery framework was presented and shared with academics and practitioners at the 6th PMI Research and Academic Virtual Conference 2021 and the Project Management Research Conference (PMDOK2022), I intend to offer it to professional organisations such as the Project Management Institute, the Australian Institute of Project Management and the Scaled Agile Academy for inclusion in their body of knowledge so that practitioners and academics can leverage the outcomes of this research.

The development of an integrated framework is a significant contribution to the literature in addressing the inherent risks of innovation which this class of programs carries. It also adds to the current literature of program management and data science, addressing the gap in relation to the management of DSIs. While practitioners have been provided with frameworks to support management of ICT programs (Artto et al., 2009; Axelos, 2022; Project Management Institute, 2017b; Thiry, 2016) and data scientists had proposed methods to develop data science procedures (Chapman et al., 2000; Fayyad et al., 1996; Mason & Wiggins, 2010; Rollins, 2015; SAS Institute, 2009; Severtson et al., 2017), an integrated framework for program management practitioners to support end-to-end DSI delivery was still missing. This lack of an integrated framework impacted the delivery of some of the initial Transport case studies described in this thesis where failure to plan and govern DSIs, to engage stakeholders continuously in the product being developed, to continuously explore, integrate and deploy the solution, to manage and govern data and lack of data preparation, modelling and evaluation all had a negative impact on their success. By integrating processes from program management (Project Management Institute, 2017b), people change management (Hiatt, 2006), Scaled Agile (SAFe) for solution delivery (Scaled Agile, 2023), DAMA's DMBOK for data management (Earley, 2017) and CRISP-DM for data science (Chapman et al., 2000), a significant gap has been addressed through this research.

10.2.3 Governance of DSIs

Too much governance and behavioural control can stifle innovation in organisations. Too little governance using only outcome control can waste precious organisational resources. Business agility demands empowerment of people to take decisions on initiatives designed to deliver innovative products and services. In response to supplementary question SQ2, I investigated the appropriate level of governance for DSIs - just right to allow exploration while not stifling innovation. I propose two contributions to the literature on DSI governance.

First, no governance can be implemented in vacuum and I reviewed the hierarchy of corporate (Claessens, 2006; John & Senbet, 1998; Müller et al., 2016; Tricker & Tricker, 2015), portfolio (Cooper et al., 1997; Jenner & Kilford, 2011; Martinsuo, 2013; Project Management Institute, 2016), program (Artto et al., 2009; Axelos, 2022; Project Management Institute, 2017b; Thiry, 2016) and agile (Association of Project Management, 2016; Collyer & Warren, 2009; Horlach et al., 2019; Kersten, 2018; Luna et al., 2014) governance. Governance of DSIs must take existing governance hierarchies into account. In Chapter 7, I extended the Goldilocks Theory of Governance (Mullay, 2010). Here, I propose minimum viable governance for DSIs to ensure that they have a balanced governance strategy from ideation through to delivery and benefits realisation that supports their exploratory nature. The eight guiding principles for minimum viable governance for DSIs are summarised in Table 10-5.

Table 10-5
Guiding Principles for Minimum Viable Governance for DSIs

No	Description
(i)	Focus on product not projects – Use product management for DSIs and not project management for delivery.
(ii)	Organise around value – Structure the teams around delivery of value and not functions.
(iii)	Fund value streams – Instead of funding projects using business cases, fund the multi-year value streams.
(iv)	Empower teams – Allow decision-making to be done by people who have the most knowledge and not based on hierarchy.
(v)	Visualise portfolio – Visualise all aspects of portfolio and make them available to all stakeholders.
(vi)	Data-driven decision-making – Collect as much relevant data and use it for decision-making.

(vii) **Measure flow** – Identify steps causing most delay in the flow and address them.

(viii) **Relentless improvement of people, process and tools** – Baseline capability and continuously improve.

While my focus in this thesis has been management and governance of DSIs, I believe that some of the concepts and guiding principles developed through this study can be extended to other ICT initiatives as well. My experience in delivering such initiatives has shown that ongoing maintenance of the project output faces challenges due to lack of resources and funding. The *2018 Chaos Report* by the Standish Group (Johnson, 2018) found that 61% of software projects are challenged, meaning they are completed but over budget, over time and/or with less than the originally specified features. This can lead to ongoing challenges during operation and maintenance due to limited resources and funding. Agile methods for software delivery, which emphasise iterative and incremental development, can help address these challenges by enabling persistent teams and funding value streams. Similarly, using empirical data for decision-making could become a norm for the ICT industry. A survey by NewVantage Partners (NewVantage Partners, 2022) found that 91.6% of executives in the industry believe that big data and AI will be a key competitive differentiator for their organisations in the future. This indicates that there is a growing recognition of the importance of data-driven decision-making in the industry.

Second, the case studies show that the nature of governance changes as organisations become more mature in delivering DSIs. They also show that there is no one size of governance that fits all DSIs, even within the same organisation. This led me to explore how the proposed minimum viable governance and the need for dynamic governance of DSIs can be linked to existing governance models. Müller (2009, 2016) identified four models of organisational project governance: process models, governance and governmentality-based models, nested models and layered models, which cover agency, institutional, stakeholder and transaction cost theories. He then described the model of four-governance paradigms linking corporate governance with organisational project governance through an overlay of two dimensions – corporate governance orientation from shareholder to stakeholder and the control structures in a project's parent organisation from a behaviour-to-outcome perspective. The conformist paradigm assumes efficiency is maximised by trusting process over the capabilities of individuals; the Flexible Economist paradigm aims for a flexible

application of the most effective project management methods, tools, techniques and management approaches, often through a PMO; the Versatile Artist paradigm is characterised by organisations that employ senior and experienced project managers, who develop their methodologies and work practices in accordance with project needs and the Agile Pragmatist paradigm emphasises process compliance with a time-phased delivery of functionalities. I extend this Four-Paradigm Governance Model (Müller, 2009, 2016) to DSIs and propose that, depending upon the maturity of the delivering organisation, DSIs can be governed using the Versatile Artist or Agile Pragmatist paradigms, as shown in *Figure 10-2*. DSI Governance using the Four-Paradigm Governance Model.

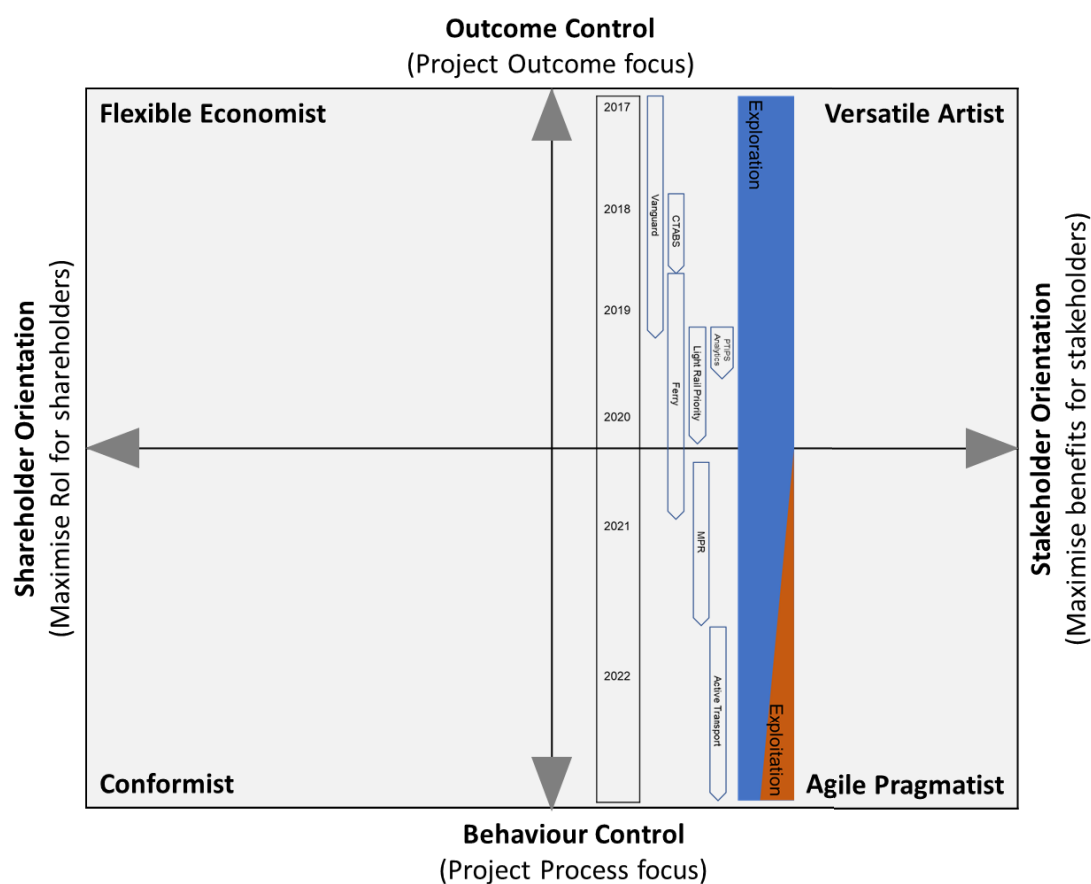


Figure 10-2. DSI Governance using the Four-Paradigm Governance Model

Previous applications of the Four-Paradigm Governance Model (Aubry, Müller, & Glückler, 2011; Joslin & Müller, 2015; Müller & Lecoivre, 2014) show these paradigms to be mutually exclusive and state that organisations governing projects tend to fit into one of the four paradigms. This can lead to ignoring the context in which projects are managed. For example, in exploring PMOs through community of practice

theory, Aubry et al. (2011) place them and associated governance in the conformist, versatile artist or agile paradigms but do not show if or how requirements evolve over time. In the context of DSIs, my case studies challenge this position and argue that for organisations commencing their DSI journey with low process maturity and requiring high stakeholder engagement, governance using Versatile Artist paradigm is suitable. The implication here is that as the processes are still evolving, the DSI needs experienced program managers, scrum masters, information architects and other specialists to ensure delivery governance. As the organisational maturity in delivering DSIs improves, and behaviour control is through process compliance, the Agile Pragmatist paradigm becomes more appropriate. The need for experienced program managers, scrum masters and information architects can be balanced by more mature processes, which in turn deliver a process-led governance. *Figure 10-2*, where the timeline of the case studies is overlayed across the Versatile Artist and Agile Pragmatist paradigms, illustrates this transition. While my governance hypothesis is based on the seven Transport DSIs, future research can validate this hypothesis against a larger cross-industry sample of DSIs. This will help the proposed DSI governance model to be generalisable and applicable globally in the emerging field of data science. Future research could also validate a contingency provision as DSIs transition from one paradigm to another and the hypothesis that contingency requirements are reduced when the organisation is operating in the Agile Pragmatist paradigm.

10.2.4 Business Cases for DSIs

The case studies show that DSIs deliver a product that needs to be managed and maintained throughout their life cycle. Instead of using a project life cycle, which has a finite scope, cost and schedule (Project Management Institute, 2021b), I propose associating a product life cycle to DSIs. The key implication here is that organisations should fund multi-year value streams instead of discrete projects.

The exploratory and innovative nature of DSIs does not allow business cases for DSIs to work well compared to exploitative projects, where costs and benefits are comparatively easier to obtain. In response to supplementary question SQ3, in Chapter 8 - Insights into Building Better Business Cases for DSIs, I proposed seven guiding principles for building better business cases for DSIs. These are summarised in Table 10-6 *Guiding Principles for Building Better Business Cases for DSIs*.

Table 10-6
Guiding Principles for Building Better Business Cases for DSIs

No	Description
(i)	Focus on product not projects – Use product management for DSIs and not project management for delivery.
(ii)	Fund value streams – Instead of funding projects using business cases, fund the multi-year value streams to better predict costs.
(iii)	Acknowledge exploratory nature – Considering the exploratory nature of DSIs, assign only those benefits which are SMART.
(iv)	Empower teams – Allow decision-making, such as on detailed scope, datasets to ingest, business rules and visualisation, to be done by people who have the most knowledge and not based on hierarchy.
(v)	Create new category for exploratory projects – Establish a new category in portfolio management for exploratory projects within discretionary investments to differentiate them from exploitative projects.
(vi)	Data-driven decision-making – Collect as much relevant data and use it for decision-making.
(vii)	Use Agile methods for delivery – Agile methods are more suitable for delivery of exploratory DSIs.

Implementation of minimum viable governance and a product life cycle can help improve the success rate of DSIs and address the gap in the literature on how such programs can be governed. Kersten (2018) challenged the traditional project management methods that have served us well for over a century and introduced flow framework, flow metrics and value stream management to measure the performance of projects and structure the delivery teams. I extend this project-to-product concept to how DSI business cases need to be framed and ultimately delivered. Guiding principles (i), (ii), (iii) and (vii) in Table 10-6 show how specifically this concept can be applied to DSIs. Thus, I contribute to the current literature in how to develop business cases for DSIs and how to measure their success from delivery of output to delivery of value.

10.3 CONTRIBUTIONS TO PRACTICE

This research had a strong focus on the management and governance of DSIs for business managers and practitioners who need to be informed about the differences between DSIs and ICT-enabled programs so that they adapt methods to improve the chance of successful business outcomes. Limited availability of methods and standards in delivery of DSIs has caused business managers and portfolio, program and change

management practitioners to chart their own path and thus introduce inconsistency in how DSIs are treated and delivered in different organisations. With this research providing guiding principles and framework, I expect the different categorisation of DSIs will provide guidance to business managers and practitioners in the effective management and governance of DSIs.

As evidenced in *Figure 4-1. Transport DSIs Chronology*, organisations and teams go through the stages of exploitation, transition and exploration in the delivery of DSIs. The process of delivering DSIs becomes more efficient as more are delivered, and the organisation learns through experience. With the exception of data from well-defined and structured sources, every new dataset carries uncertainty in its scope and quality. This uncertainty brings in the need for combining the exploration component of a DSI with exploitation as the organisation and its teams mature.

When I reflect on the journey of DSIs in an organisation, the initial DSIs fell significantly into the “exploratory projects” category due to their uncertainties and unknowns. As the organisation delivered more DSIs, the number of uncertainties and unknowns started reducing. While the business goals may still have been uncertain, repeatable patterns emerged to underpin the technology and methods. I saw a link between the maturity of a program team and business to deliver DSIs and the need to manage them as “exploratory projects”.

I thus argue that DSIs are typically exploratory projects due to uncertainty around their benefits and the delivery itself. Other than setting up a foundation infrastructure, the delivery work packages of DSIs cannot be clearly defined at the outset and thus the schedule cannot be prepared in full detail. DSIs cannot conform to the traditional waterfall approach to projects so that they deliver a unique product, service or result within a specified period, defined budget and defined quality requirements. While iterative delivery has been proposed for software projects for some time (Kruchten, 2004), DSIs lend themselves rather to being delivered in a progressive elaboration using agile methods. This progressive elaboration allows a data scientist to get continuous feedback from a product owner on the outcomes and not wait for an extended period only to find out that what has been delivered is not what was desired.

With their focus on risk elimination and rapid delivery of business outcomes, current program management practice guides (Artto et al., 2009; Axelos, 2022; Project Management Institute, 2017b; Thiry, 2016) do not adequately support DSI delivery. Using my experience of the seven DSIs discussed in this research, I saw that they could not be effectively delivered as exploitative projects using waterfall methods and hence believe that business managers need to be aware of exploratory vs exploitative projects at the time of investment approval. Practitioners need a framework that adequately supports DSI delivery.

I reviewed portfolio management practices in relation to DSIs. Reflecting on my experience at Transport, I recommend that business managers be made aware that significant effort is required to build DSI delivery capability and that effort should be considered in business cases. The initial investments may not deliver the full scope but are likely to lay the foundation for future success. This includes not only maturity in the skills of the team but also better data management, data governance and reuse of datasets, thus contributing to richer decision-making capability in relation to future investments.

10.3.1 Unique Characteristics of DSIs

I expect business managers and practitioners will use the seven characteristics defined in Table 1-2 *Unique Characteristics of DSIs* in DSI management and governance to address the uncertainty, innovation and inherent risks associated with their delivery. I also expect practitioners to define the success factors suitable to this class of programs for metrics rather than follow traditional ICT success factors.

10.3.2 Management of DSIs

As mentioned in Section 10.2.2, practitioners have had frameworks to support management of ICT programs and data scientists have had methods to develop data science procedures. Yet, big-data projects have had a low success rate. The seven case studies at Transport showed that failure to plan and govern DSIs, to continuously engage stakeholders in the product being developed, to continuously explore, integrate and deploy the solution, failure to manage and govern data and the lack of data preparation, modelling and evaluation all impacted the success rate of my DSIs. I expect practitioners will configure the integrated DSI delivery framework as shown in *Figure 10-1 DSI Delivery Framework*, Table 10-3 *DSI Delivery Framework* –

Summary of Key Processes across five domains and program phases and in Table 10-4 *Roles and Responsibilities across DSI Domains* to support innovation and address the inherent risks in DSIs.

10.3.3 Governance for DSIs

Business managers and practitioners have struggled to find a balance between implementing too much governance, which stifles innovation and too little governance, whereby precious organisation resources are wasted. As evidenced in the Active Transport portfolio described in Chapter 7, implementing minimum viable governance of portfolio and product management and bringing in a degree of business agility allowed Active Transport to address the challenges it faced, namely, managing the portfolio efficiently and effectively, using data for benefits tracking and investment decisioning and establishing new ways of working with the new team. This approach can be used by business managers and practitioners in other organisations as well.

I expect business managers and practitioners to leverage the seven principles defined in Table 10-5 *Guiding Principles for Minimum Viable Governance for DSIs* and *Figure 10-2 DSI Governance* using the Four-Paradigm Governance Model when implementing governance in their organisation.

10.3.4 Business Cases for DSIs

The exploratory and innovative nature of DSIs does now allow business cases for DSIs to stack well against exploitative projects where costs and benefits are comparatively easier to obtain. As evidenced in the Transport DSIs, moving from project-to-product mindset, funding value streams and acknowledgment of exploratory nature of DSIs allowed Transport to build better business cases. This approach can be used by business managers and practitioners in other organisations as well.

I expect business managers and practitioners to leverage the learnings from Transport DSIs and the seven guiding principles, as defined in Table 10-6 *Guiding Principles for Building Better Business Cases for DSIs*.

10.3.5 Using DSIs for Benefits Tracking

Significant effort goes into identifying benefits for infrastructure projects in business cases. Tracking these benefits is complex and often ignored once the infrastructure is delivered. Also, few organisations are harnessing benefits tracking

effectively for data-driven decision-making. In Chapter 9 - Using DSIs for Benefits Tracking and Investment Decisioning – A Transport for NSW Case Study, I highlighted how Transport was using DSIs to track the benefits from its investment in Active Transport infrastructure and to identify future investments.

I expect business managers and practitioners to leverage the learnings from Transport DSIs and to use DSIs to track the benefits of their infrastructure projects.

10.4 OTHER CONTRIBUTIONS

10.4.1 Establishment of Management and Governance of DSI Course

Considerable effort goes into conducting research and it is important that the value of research is realised through effective dissemination. Ross-Hellauer et al. (2020) outline 10 rules that researchers can follow to disseminate their work in novel and engaging ways, and hence increase the impact of their research on science and society. While I used the traditional outputs like research articles, book chapters and conference presentations, I also explored alternatives, as suggested in Rule 6 (Ross-Hellauer et al., 2020). A Google search for “data science courses” showed universities in Australia offering one or more technical courses on data science. However, bridging technology and program management is still an uncharted area with very few universities offering such courses. The Data Science Process Alliance (2022) is one private institute offering courses that bridge data science and project management but they covered only data science and agile domains and excluded program, change and data management.

Using content from my research, I created a microcredential (Mathur & Sankaran, 2021) at the University of Technology Sydney. The course helps organisations and individuals currently not well equipped to manage programs and projects that use big data to drive strategic decision-making. Providing governance structures and processes to manage DSIs, this course is aimed at senior management and program managers to plan and deliver data science-based programs and projects effectively and it uses case studies from organisations that have delivered such initiatives. The course is offered over four days and has the following structure:

Day 1 – Data Science Initiatives (DSIs) and Data Management

Activity	Duration	Time
Welcome and Concept Mapping	15 min	6:00 – 6:15
Lecture – Unique Characteristics of DSIs	15 min	6:15 – 6:30
Lecture – Data Management	30 min	6:30 – 7:00
Breakout Session	15 min	7:00 – 7:15
Case Study – Major Infrastructure Projects	45 min	7:15 – 8:00
Case Study – Transport for NSW (Vanguard and MPR)	30 min	8:00 – 8:30
Assessment 1A – Data Management Quiz	15 min	8:30 – 8:45
Concept Mapping	15 min	8:45 – 9:00

Day 2 – Agile Frameworks and Change Management

Activity	Duration	Time
Reflection	15 min	6:00 – 6:15
Lecture – Agile Frameworks	30 min	6:15 – 6:45
Lecture – Change Management	15 min	6:45 – 7:00
Breakout Session	15 min	7:00 – 7:15
Case Study – Transport for NSW (Active Transport)	45 min	7:15 – 8:00
Assessment 1B – Agile Frameworks & Change Management Quiz	15 min	8:00 – 8:15
Concept Mapping	15 min	8:15 – 8:30

Day 3 – Data Science Methodologies and Delivery Framework

Activity	Duration	Time
Reflection	15 min	6:00 – 6:15
Lecture – Data Science Methodologies	30 min	6:15 – 6:45
Lecture – DSI Delivery Framework	15 min	6:45 – 7:00
Breakout Session	15 min	7:00 – 7:15
Case Study – Salesforce	45 min	7:15 – 8:00
Lecture – DSI Roles and Responsibilities and Technology Landscape	30 min	8:00 – 8:30

Assessment 1C – Data Science Methodologies and DSI

Delivery Framework Quiz (5%)

Assessment 1D – DSI Roles and Responsibilities and Technology Landscape (5%)	15 min	8:30 – 8:45
--	--------	-------------

Assessment 2 – Peer-reviewed proposal (20%)

Concept Mapping	15 min	8:45 – 9:00
-----------------	--------	-------------

Day 4 – Bringing it together

Activity	Duration	Time
Reflection	10 min	6:00 – 6:10
Case Study – Agile Analytics	40 min	6:10 – 6:50
Case Study – Microsoft	40 min	6:50 – 7:30
Preparation for DSI Case Study Presentation	30 min	7:30 – 8:00
Assessment 3 – DSI Case Study Presentations (60%)	30 min	8:00 – 8:30
Feedback on Presentations	15 min	8:30 – 8:45
Close	15 min	8:45 – 9:00

The course has been offered twice so far, in 2021 and 2022, providing an alternative way of disseminating the outcomes of my research. A total of 18 practitioners and researchers have attended so far and, based on the student survey results, it was well received.

10.5 RESEARCH LIMITATIONS

This research used a multi-methods approach, combining learnings from delivering DSIs through seven case studies at Transport and semi-structured interviews with practitioners from five organisations. These insights were then used to develop a DSI delivery framework, as implemented in the Active Transport branch of Transport. However, the research also has several limitations, which should be acknowledged.

First, the sample size was limited to seven DSIs from one public sector organisation in Australia. This may affect the generalisability of research findings, as noted by Sandberg (2005) that truth is always unfinished in the interpretive tradition, and researchers cannot generate absolute truth claims. The case studies spanning six years combined with semi-structured interviews with practitioners from other

organisations allowed me to validate some of the research findings. However, the DSI characteristics identified in Chapter 5 - Unique Characteristics of DSIs may not be an exhaustive and universal set, and more may emerge or change as additional DSIs are studied.

Second, the study developed the DSI delivery framework and guiding principles for minimum viable governance and developing better business cases using DSIs from only one public sector organisation. As no two organisations operate in the same manner, research in another setting may yield slightly different results. Future research could focus on identifying key factors that may influence the delivery of DSIs in different settings, such as differences in organisational culture, governance structures or resource availability.

Third, DSIs are a recent and rapidly evolving phenomenon in the technology of data science and delivery space. The currency of the research work may be impacted as some of the findings could change over time and as the maturity of DSIs change from being exploratory to exploitative. The DSI delivery framework and guiding principles may need to be updated or adapted in response to emerging trends and technologies.

In the next section, areas of future research are proposed to address these limitations.

10.6 FUTURE RESEARCH

In conclusion, this chapter and the thesis have highlighted several areas for future research that warrant attention in both the literature and practice of managing and governing DSIs. To enhance current approaches, it is recommended that the DSI characteristics, the DSI delivery framework, and the guiding principles be validated with other public and private sector organisations delivering DSIs in Australia and other countries as well as with further theoretical explorations. This can be achieved by using a larger sample of case studies or mixed-methods research using quantitative and qualitative methods.

Cross-industry case studies can be conducted to validate the applicability and effectiveness of the developed DSI delivery framework, minimum viable governance principles, and business case guidelines. By examining DSIs in industries such as

finance, healthcare, manufacturing, or retail, the versatility and generalisability of the research outcomes can be demonstrated.

A comparative analysis between DSIs implemented in different organisational contexts can be done to identify common challenges, best practices, and adaptations of the framework and principles. This comparative approach can provide insights into how variations in organisational culture, governance structures, or resource availability impact the delivery and management of DSIs.

There is potential to collaborate with researchers and organisations from different countries and regions to implement and evaluate the proposed framework and principles in diverse cultural and regulatory environments. International collaboration can enrich the research findings and ensure that the recommendations are relevant and applicable on a global scale.

Further research is also needed to identify success factors specific to this class of programs. This includes advancing the literature on exploration, management, governance and portfolio management of DSIs, as highlighted in Section 10.2 – Contributions to Theory.

In addition, future research should investigate how organisations delivering DSIs can transition from the explorative to the exploitative stage more quickly, how changes in governance can affect DSI delivery and the need for contingency provision. Answering these research questions will provide valuable insights for practitioners and help advance the field of DSI management and governance.

10.7 CONCLUSION

I conclude that the current literature does not adequately cover management and governance of DSIs and that business managers and practitioners need to be informed about the differences between DSIs and ICT-enabled programs so that they can adapt methods to improve the chance of successful business outcomes.

Program management for ICT-enabled programs has rich literature and proven delivery frameworks which have matured since the 1990s (Axelos, 2022; Project Management Institute, 2016, 2017b). However, the current program management literature does not adequately support the delivery of innovative and exploratory DSIs, focusing instead on risk elimination and rapid delivery of business outcomes of

exploitative initiatives. In 2022, the global big-data analytics market size was valued at US\$271.83 billion. The market is projected to grow from US\$307.52 billion in 2023 to US\$745.15 billion by 2030 (Fortune Business Insights, 2023). The availability and adoption of a suitable delivery framework can potentially reduce wastage of this \$60 billion annual spend on DSIs. In Chapter 5 - Unique Characteristics of DSIs, I identify six unique characteristics of DSIs to be used in their delivery.

I drew on my own struggles in delivering DSIs to create an integrated DSI delivery framework. In Chapter 6 - Framework to Manage DSIs, I provide details of the design principles and the framework, including the processes.

In Chapter 7 - Minimum Viable Governance for DSIs, I used an Active Transport case study to introduce minimum viable governance (MVG) of portfolio and product management. This allowed Active Transport to address the challenges it faced, namely, managing the portfolio efficiently and effectively, using data for benefits tracking and investment decisioning and establishing new ways of working with the new team. I therefore recommend the eight guiding principles for minimum viable governance for DSIs. I also argue that “being flexible on the means” does not imply “losing control” of DSIs from a top management perspective.

My experiences at Transport from 2017 to 2022 have allowed me to reflect on how we constructed and delivered business cases for traditional ICT investments compared to those for DSIs. As discussed in Chapter 8 - Insights into Building Better Business Cases for DSIs, business cases for traditional ICT investments typically focus on hard costs and benefits and, while exploitative, pose challenges in the governance of benefits realisation. The problem is compounded for exploratory investments such as DSIs where both costs and benefits cannot be quantified upfront. I conclude that moving from a project to a product mindset, funding value streams and acknowledgment of exploratory nature of DSIs allows better business cases to be built. While the organisation and teams mature and DSIs move from being exploratory to exploitative, the exploratory aspect of DSIs remains, only the percentage of uncertainty reduces. Based on the Transport case study, I recommend the eight guiding principles for building better business cases for DSIs. While Transport used SAFe for delivering DSIs, I believe that the guiding principles are generic and can be applied by organisations using any other agile delivery method.

While big-data and data science are attracting significant investment, successful implementation and use cases that show the value derived from them need to be shared so that organisations can learn from them. In Chapter 9 - Using DSIs for Benefits Tracking and Investment Decisioning – A Transport for NSW Case Study, I share how Transport used DSIs to track the benefits from its investment in Active Transport infrastructure and to identify future investments. I see special relevance to organisations in harnessing the power of data science and DSIs for benefits tracking, an area which gets ignored as programs are closed and teams move on to the next initiative.

I also conclude that organisations and teams go through stages of exploration, transition and exploitation in the delivery of DSIs and the process of delivering DSIs becomes more efficient the more are delivered. With the exception of data from a well-defined and structured source, every new dataset carries uncertainty in scope and quality. This brings in the need to have an underpinning exploration component in a DSI combined with a shift to exploitation as the organisation and teams mature.

Figure 10-3. Management and Governance of DSIs summarises the findings and conclusions of this thesis, showing the DSI delivery framework in the middle, expanding on unique characteristics, providing guiding principles on governance and business cases, followed by a list of the skills required to deliver them.

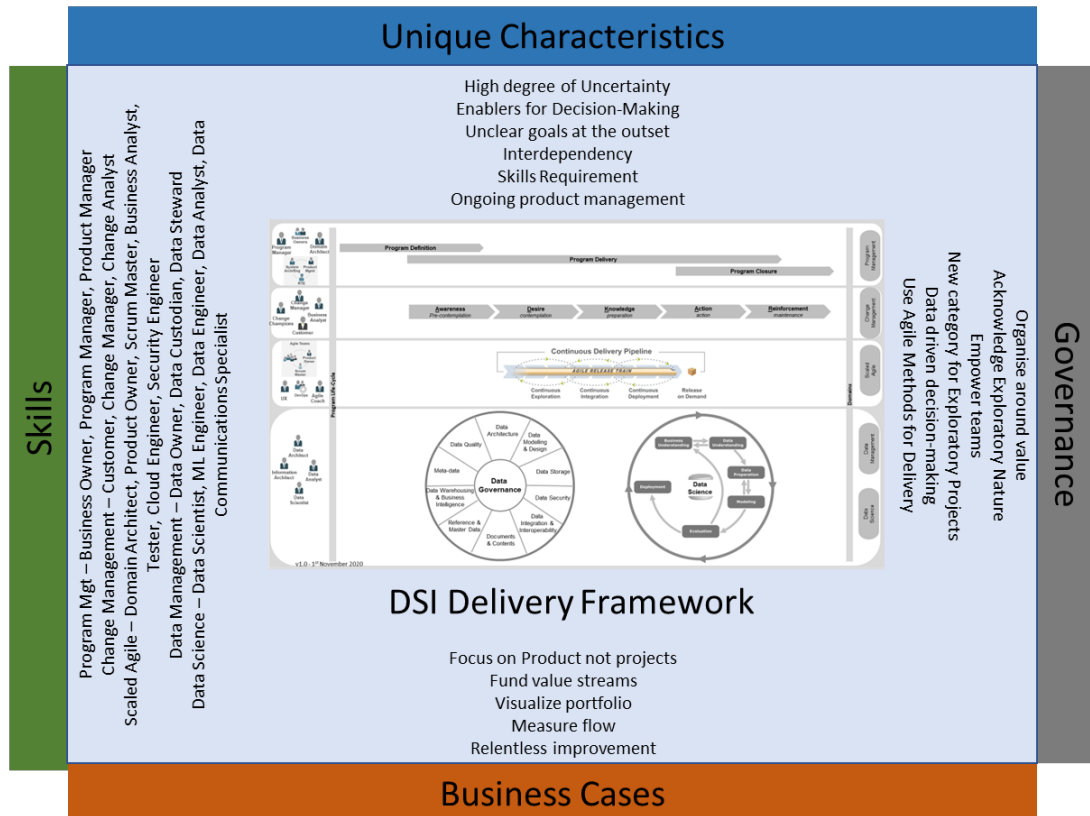


Figure 10-3. Management and Governance of DSIs

This research is but a small addition to the body of knowledge for DSI program management relevant to both literature and practitioners. Additional case studies in other sectors are required to enrich this work. What is quite clear is that without further research, there will be more failed DSIs and dissatisfied sponsors and delays in the much-needed investment in this emerging domain as well as in the benefits that will flow from harnessing the data and the nuggets in it.

11. Appendices

11.1 APPENDIX A: ANALYSIS OF ARTICLES ON BIG DATA

Analysis of selected articles on Big Data				
Reference	Management Field	Type of Research	Main Findings	Relevance to research
George et al. (2014)	Big Data Management (Editorial from a Special Issue of Academy of Management Journal (2014)	Stage setters for future work on big data based on article published in the special issue	• Big data will change the landscape of economic policy and research • Continuous ubiquitous nature of data can help address how business processes and opportunities evolve in real time • Has the potential to transform management theory, and practice through granular data	• Dynamic nature of big data and impact on how long-term programs are delivered and manage changes • Dynamics of program implementation will be affected by continuous and ubiquitous nature of data

Jamil and Carvalho (2019)	Project Management	Review in a book chapter using case studies	• Importance of data for decision making and knowledge sharing in projects • Growth of databases will need new types of software to store, manage and analyze • How relationship between project and knowledge management relationship can be improved using data analytics	• Will impact on knowledge management strategies within programs
Rathore, Ahmad, Paul, Wan, and Zhang (2016)	Urban Planning	Developed a Model to implement Big Data Analytics using IoT for Smart City and Urban Planning using multiple datasets	• The systems developed by the authors will benefit both citizens and authorities in making fast decisions	• How to develop a model to enhance programs to support urban planning and infrastructure
Saltz and Shamshurin (2016)	Big Data Process Methodologies	Literature Review	• No agreed standard for these projects but a process methodology might be useful • Success factor to execute big data projects	• Success factors could be useful?

Viaene and Van den Bunder (2011)	Managing Business Analytics Projects	Data gathered through in-depth interviews of experienced project managers	• Cope with uncertain changing requirements • Engaging stakeholders and managing expectations • Intelligent experimentation	• Challenges of big data analytics project that need to be taken into account
Waller and Fawcett (2013)	Supply Chain Design and Management (SCM)	Review	• Has significant implications to SCM • Difficult to see direct connections to research and teaching	• Skills required by data scientists • Examples of causes • Relevant theories
Westergren, Saarikko, and Blomquist (2017)	The IoT Guide	Case Studies	• In 2020 there will be at least 20 billion connected devices in the world (Gartner) • IoT will have significant impact on business strategies • Vertical and horizontal strategies will face conflicts due to different actors planning while at the same time new ways of collaboration can become easier	• The impact of the phenomenal growth of IoT device and the challenges to be faced in handling all. IoT will change business strategies which has an effect on how programs will support realization of strategy.

Zhou, Fu, and Yang (2016)	Smart Energy Management from big data	Review of state of research	➤ Large amounts of data being collected in the energy sector due to IoT and cloud technologies ➤ Volume of data growing rapidly I complex structures and forms ➤ Big data analytics will play an important role in smart grid management ➤ Big data analytics is a multi -disciplinary field ➤ Many barriers to data sharing ➤ Governance of data is lacking	➤ Programs that include energy management will be impacted by the advances in big data ➤ Data governance needs careful consideration
------------------------------	---	--------------------------------	---	---

11.2 APPENDIX B: SEMI-STRUCTURED INTERVIEW GUIDE

General information

Participant Name:

Organisation:

Role:

Interview Date & Time:

Consent form(s) completed: Yes/No Audio Consent? Yes/No

Researcher Background

As you know, I have been delivering data analytics programs at Transport for last four years. In my new role as Director – Active Transport at Transport NSW, I am managing GIS, Insights and Walking & Cycling Program.

I am undertaking a PhD at UTS in the efficient delivery of data science programs. This interview is one of the significant steps of my research methodology.

Research Overview

Information and Communications Technology (ICT) programs are increasingly integrating a data-science component. Initial observations suggest that there are subtle, but crucial differences in the effective delivery of such Data Science Initiatives (DSIs) when compared to other ICT-enabled programs. A review of the literature shows a lack of accepted frameworks for Program Managers to adopt when managing DSIs and navigate through the exploratory and innovative nature of such investments.

This research will investigate the characteristics of DSIs across all phases of a Program Life-Cycle and contribute to an evidence-based delivery framework to increase the likelihood of DSIs creating the envisaged value.

Data Science Initiatives (DSIs) in context of this research includes investments in Data Science, Data Analytics, Business Intelligence, Big-Data, and Artificial Intelligence.

To meet the challenge described above, the thesis will focus on investigating the following research question:

Why do Data Science Initiatives (DSIs) face challenges delivering envisaged value when using traditional processes for managing ICT-enabled programs? How do these challenges manifest within the delivery process? How is the gap in literature across relevant domains contributing to these challenges?

Following outcomes are expected from this research:

- Identification of unique characteristics due to which DSIs face challenges delivering envisaged value when using traditional processes for managing ICT-enabled programs; and
- Creation of a body of knowledge including a framework for delivery of DSIs acknowledging their exploratory and innovative nature and thus improving the likelihood of delivering their envisaged value

This interview will focus on your experiences in delivering DSIs.

Interview Overview

This research interview is designed to be fairly open-ended and I want to encourage the information to flow in a relatively informal way. However, to keep within the time available I may need to steer the discussion to ensure we cover the main elements.

The overall structure of the interview includes 4 sections as follows:

1. Knowing more about you, your organisation and typical projects you have delivered
2. Your views of DSIs and how they compare with other ICT Programs. What frameworks/methodologies did you use.
3. I have made some assumptions on DSIs and would like to get your views on them
4. Before we close this interview, we will have a chance cover any other comments/thoughts.

Section 1 – Interviewee Background

1. Could you please tell me about your background? How long have you been working in ICT sector? What type of industries have you worked in?
2. What is a typical week in your role look like? What are the challenges you face?

Section 2 – Comments on DSIs

1. What is DSI to you? Please share your experiences of a DSI Program you have worked on. What was your role and what was the size of Team/Budget/Duration?
2. How long have you been managing DSIs? What has changed in terms of people, process and technology?
3. How do you compare an ICT Program with DSI? What similarities do you see? What differences do you see? Has this changed over time? If so, why?
4. What Delivery Frameworks/Methodologies have you used for DSIs. Why did you choose them? Is there a criterion you use when choosing the Delivery Framework / Methodology for a DSI?
5. Were the DSIs you delivered considered successful by Sponsor/Client. What made them successful? When were they considered not to be successful?
6. What key risks and issues have you seen in DSIs?
7. Anything you could have done differently to deliver DSIs?

Section 3 – Validation of DSI Assumptions

While delivering DSIs, I found some common characteristics which I would like to validate with you. Also, I would like to see if you have observed something else.

What are your thoughts on the following assumptions?

1. DSIs carry high degree of uncertainty right from initiation through to closing phases.
2. DSIs are enablers for decision making & may not have a direct benefit contribution

3. Neither the goals nor the means of attaining them are clearly defined from the outset for a DSI
4. DSIs are not independent of each other and act as an enabler to next one
5. Skills required to deliver a DSI are different to typical ICT program
6. DSIs do not end and after initial delivery convert into continuous business improvement initiative

Have you observed any other unique characteristic in DSIs?

Section 4 - Closing

1. Is there any other part of DSI which you would like to discuss?
2. Can you think of other people I should talk to about DSIs?

11.3 APPENDIX C: NAVIGATING HYPERLINKS

This document uses hyperlinks throughout to support cross-referencing. Instructions on how to navigate links are described below depending on whether the document is opened in a browser-based or a desktop PDF reader.

In browser-based PDF Reader

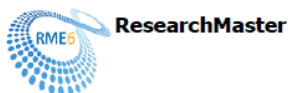
- Open the PDF document in your web browser (e.g., Chrome, Edge, Firefox).
- Move your cursor over the hyperlink text or area until it changes to a hand symbol, indicating a clickable link.
- Click the hyperlink once. Your browser will open a new tab or window, taking you to the linked webpage.
- To return to the PDF document, use the browser's "Back" button or close the new tab/window.

In desktop PDF Reader (e.g., Adobe Acrobat Reader)

- Open the PDF document in your desktop PDF reader.
- Move your cursor over the hyperlink text or area until it changes to a hand symbol.
- Click the hyperlink once. If it's a web link, your default web browser will open, directing you to the linked webpage. If it's an internal link within the same PDF, the viewer will jump to the relevant section or page.
- To return to the PDF document, use your PDF reader's navigation controls or Alt + Tab (Windows) or Command + Tab (Mac) to switch between applications.

11.4 APPENDIX D: ETHICS APPROVAL

Ethics approval was granted by UTS prior to commencing the research (Application ID ETH20-5468, dated 28th January 2021).



Human Ethics Application

Application ID :	ETH20-5468
Application Title :	Managing Data Science Initiatives as Exploratory Projects: A New Approach to Program Management
Date of Submission :	
Primary Investigator :	Prof Shankar Sankaran (Chief Investigator)
Other Personnel :	Mr Sandeep Mathur (SResearch Student)
	Prof Ivor Tsang (Co-Supervisor)
	Dr Samuel Christopher Macaulay (Co-Supervisor)

Dear Applicant,

Re: ETH20-5468 - "Managing Data Science Initiatives as Exploratory Projects: A New Approach to Program Management"

Your local research office has reviewed your application and agreed that it now meets the requirements of the National Statement on Ethical Conduct in Human Research (2007) and has been approved on that basis. You are therefore authorised to commence activities as outlined in your application, subject to any conditions detailed in this document.

You are reminded that this letter constitutes ethics approval only. This research project must also be undertaken in accordance with all [UTS policies and guidelines](#) including the Research Management Policy.

Your approval number is UTS HREC REF NO. ETH20-5468

Approval will be for a period of five (5) years from the date of this correspondence subject to the submission of annual progress reports.

The following standard conditions apply to your approval:

- Your approval number must be included in all participant material and advertisements.
- Any advertisements on Staff Connect without an approval number will be removed.
- The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project to the Ethics Secretariat (Research.Ethics@uts.edu.au).
- The Principal Investigator will notify the UTS HREC of any event that requires a modification to the protocol or other project documents, and submit any required amendments prior to implementation. Instructions on how to submit an amendment application can be found [here](#).
- The Principal Investigator will promptly report adverse events to the Ethics Secretariat. An adverse event is any event (anticipated or otherwise) that has a negative impact on participants, researchers or the reputation of the University. Adverse events can also include privacy breaches, loss of data and damage to property.
- The Principal Investigator will report to the UTS HREC annually and notify the HREC when the project is completed at all sites.
- The Principal Investigator will notify the UTS HREC of any plan to extend the duration of the project past the approval period listed above through the progress report.
- The Principal Investigator will obtain any additional approvals or authorisations as required (e.g. from other ethics committees, collaborating institutions, supporting organisations).
- The Principal Investigator will notify the UTS HREC of his or her inability to continue as Principal Investigator including the name of and contact information for a replacement.

This research must be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research.

You should consider this your official letter of approval.

If you have any queries about this approval, or require any amendments to your approval in future, please do not hesitate to contact your local research office or the Ethics Secretariat.

Ref: 12a

12. Glossary

Agile. A term used to describe a mindset of values and principles as set forth in the Agile Manifesto.

Agile Release Train (ART). A long-lived Agile team that incrementally develops, delivers, and often operates one or more solutions in a value stream. Each ART is a virtual organisation (typically 50 – 125 people) organised around the enterprise's significant Development Value Streams and exist solely to realise the promise of that value by building and delivering solutions that benefit the customer.

Artificial Intelligence. Applies advanced analysis and logic-based techniques, including machine learning, to interpret events, support and automate decisions, and take actions.

Big Data. It is the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.

Business Case. A value proposition for a proposed project that may include financial and nonfinancial benefits.

Business Intelligence. It is an umbrella term which includes applications, tools, infrastructure, and practices to enable access and analysis of information to optimize performance and decision-making.

Benefits Realization. Ensuring that program delivers intended outcomes and value.

Change Management. Change management is the application of a structured process and set of tools for leading the people side of change to achieve a desired outcome.

Data Management. Consists of the practices, architectural techniques, and tools for achieving consistent access to and delivery of data across the spectrum of data subject areas and data structure types in the enterprise, to meet the data consumption requirements of all applications and business processes.

Data Science. Data Science is the collection, preparation, analysis, visualization, management, and preservation of large collections of information to generate actionable insights.

Data Science Initiatives. Data Science Initiatives (DSIs) are defined as related projects or programs that involve the application of data science techniques and methods to address complex problems, generate insights, or create new products or services.

Deep Learning. It is a subset of machine learning, which is essentially a neural network with three or more layers.

Exploratory Projects. Projects for which neither the goals nor the means of attaining them are clearly defined from the outset.

Exploitative Projects. Projects which focus on optimizing cost-quality-time triple constraints to deliver new products and services.

Machine Learning. It is an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. ML focuses on the development of computer programs that can access data and use it learn for themselves.

Portfolio Management. The centralized management of one or more portfolios to achieve strategic objectives.

Program Management. The application of knowledge, skills, tools, and techniques to a program to meet the program requirements and to obtain benefits and controls not available by managing projects individually.

Project Management. The application of knowledge, skills, tools, and techniques to project activities to meet the project requirements.

SAFe. A framework for business agility which integrates the power of Lean, Agile, and DevOps into a comprehensive operating system that helps enterprises by delivering innovative products and services faster, more predictably, and with higher quality.

Scaled Agile. A set of organizational and workflow patterns for implementing agile practices at an enterprise scale.

Value Stream. Represent the series of steps that an organization uses to implement solutions that provide a continuous flow of value to a customer.

13. References

- Agar, M. H. (1986). *Speaking of ethnography* (Vol. 2): Sage.
- Agutter, C. (2020). *ITIL Foundation Essentials ITIL 4 Edition-The Ultimate Revision Guide*: IT Governance Publishing Ltd.
- Al-Sai, Z. A., Abdullah, R., & Husin, M. H. (2020). Critical success factors for big data: a systematic literature review. *IEEE Access*, 8, 118940-118956.
- Aljumah, A. I., Nuseir, M. T., & Alam, M. M. (2021). Organizational performance and capabilities to analyze big data: do the ambidexterity and business value of big data analytics matter? *Business Process Management Journal*, 27(4), 1088-1107.
- Alvarez, S., Afuah, A., & Gibson, C. (2018). Editors' Comments: Should management theories take uncertainty seriously? *Academy of Management*, 43(2), 169-172.
- Ambler, S. W., & Lines, M. (2016). *The disciplined agile process decision framework*. Paper presented at the Software Quality. The Future of Systems- and Software Development: 8th International Conference, SWQD 2016, Vienna, Austria, January 18-21, 2016, Proceedings 8.
- American Psychological Association. (2019). Publication manual of the american psychological association,(2020). *American Psychological Association*, 428.
- Artto, K., Martinsuo, M., Gemünden, H. G., & Murtoaro, J. (2009). Foundations of program management: A bibliometric view. *International Journal of Project Management*, 27(1), 1-18.
- Asay, M. (2017). 85% of big data projects fail, but your developers can help yours succeed. *Big Data Tech. Rebutlic*, 11, 1-5.
- Association of Project Management. (2016). *Directing Agile Change*: APM.
- Atkinson, R., Crawford, L., & Ward, S. (2006). Fundamental uncertainties in projects and the scope of project management. *International Journal of Project Management*, 24(8), 687-698.
- Atlassian. (2020). Beyond the basics of scaling agile. Retrieved from <https://www.atlassian.com/agile/agile-at-scale>
- Aubry, M., Müller, R., & Glückler, J. (2011). Exploring PMOs through community of practice theory. *Project Management Journal*, 42(5), 42-56.
- Axelos. (2022). *Managing successful programmes* (5th ed.): PeopleCert.
- Azevedo, A. I. R. L., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. *IADS-DM*.
- Becker, D. K. (2017, 11-14 December 2017). *Predicting outcomes for big data projects: Big Data Project Dynamics (BDPD): Research in progress* Paper presented at the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA.
- Belling, S. (2020). Approaches to Scaling Agile: Determining when and how to scale. *Succeeding with Agile Hybrids: Project Delivery Using Hybrid Methodologies*, 139-150.

- Brady, T., & Davies, A. (2004). Building project capabilities: from exploratory to exploitative learning. *Organization studies*, 25(9), 1601-1621.
- Cao, H., & Folan, P. (2012). Product life cycle: the evolution of a paradigm and literature review from 1950–2009. *Production Planning & Control*, 23(8), 641-662.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS Incorporated*, 9, 13.
- Chirumalla, K. (2017). *Challenges in managing new product introduction projects: An explorative case study*. Paper presented at the DS 87-2 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 2: Design Processes, Design Organisation and Management, Vancouver, Canada, 21-25.08. 2017.
- Christensen, C. M., Kaufman, S. P., & Shih, W. C. (2008). Innovation killers: how financial tools destroy your capacity to do new things. *Harvard Business Review* 86(1), 98-105, 137.
- Cilliers, E. J., Sankaran, S., Armstrong, G., Mathur, S., & Nugapitiya, M. (2021). From Urban-Scape to Human-Scape: COVID-19 Trends That will Shape Future City Centres. *Land*, 10(10), 1038.
- City of Sydney. (2018). Cycling Strategy and Action Plan: For a more sustainable Sydney. Retrieved from <https://www.cityofsydney.nsw.gov.au/strategies-action-plans/cycling-strategy-and-action-plan>
- Claessens, S. (2006). Corporate Governance and Development. *The World Bank Research Observer* 21(1), 91-122.
- Clegg, S., Killen, C. P., Biesenthal, C., & Sankaran, S. (2018). Practices, projects and portfolios: Current research trends and new directions. *International Journal of Project Management*, 36(5), 762-772.
- Cois, C. A., Yankel, J., & Connell, A. (2014, 13-15 October 2014). *Modern DevOps: Optimizing software development through effective system interactions*. Paper presented at the 2014 IEEE International Professional Communication Conference (IPCC), Pittsburgh, PA, USA.
- Collyer, S., & Warren, C. M. (2009). Project management approaches for dynamic environments. *International Journal of Project Management*, 27(4), 355-364.
- Cooper, R. G., Edgett, S. J., & Kleinschmidt, E. J. (1997). Portfolio management in new product development: Lessons from the leaders—I. *Research-Technology Management*, 40(5), 16-28.
- Data Science Process Alliance. (2022, 27/12/2022). Data Science Team Lead Plus. Retrieved from <https://www.datascience-pm.com/courses/>
- Davies, A., Manning, S., & Söderlund, J. (2018). When neighboring disciplines fail to learn from each other: The case of innovation and project management research. *Research Policy*, 47(5), 965-979.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*.
- Devkar, G., Sankaran, S., Ke, Y., Ninan, J., Mathur, S., Tsang, I., & Vuorinen, L. (2021). *Data Analytics to Evaluate Public Value from Megaprojects*. Paper presented at the The 6th PMI Research and Academic Virtual Conference 2021, Mumbai, India. <https://www.pmi.org.in/RAC2021/>
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73.

- digital.ai. (2021). 15th State of Agile Report. Retrieved from <https://digital.ai/resource-center/analyst-reports/state-of-agile-report>
- Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766.
- Drury-Grogan, M. L. (2014). Performance on agile teams: Relating iteration objectives and critical decisions to project management success factors. *Information and Software Technology*, 56(5), 506-515.
- Duncan, R. B. (1976). The ambidextrous organization: Designing dual structures for innovation. *The management of organization*, 1(1), 167-188.
- Earley, S. E. (2017). DAMA-DMBOK: Data management body of knowledge. *Technics Publications*.
- Ebert, C., & Paasivaara, M. (2017). Scaling Agile. *IEEE Software*, 34(6), 98-103.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review* 14(4), 532-550.
- EY. (2015). Becoming an analytics-driven organization to create value. Retrieved from https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/digital/ey-global-becoming-an-analytics-driven-organization.pdf
- 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.
- Flyvbjerg, B. (2007a). How optimism bias and strategic misrepresentation in early project development undermine implementation. *Concept* 41.
- Flyvbjerg, B. (2007b). How optimism bias and strategic misrepresentation in early project development undermine implementation. *concept*, 41.
- Foroughi, F., & Luksch, P. (2018). Data Science Methodology for Cybersecurity Projects. *arXiv preprint arXiv:1803.04219*.
- Fortune Business Insights. (2023). Big Data Analytics - Market Research Report. Retrieved from <https://www.fortunebusinessinsights.com/big-data-analytics-market-106179>
- Franková, P., Drahosová, M., & Balco, P. (2016). Agile Project Management Approach and its Use in Big Data Management. *Procedia Computer Science*, 83, 576-583.
- Gartner. (2019, 5 November 2019). Gartner Top 10 Data and Analytics Trends for 2019. Retrieved from <https://www.gartner.com/smarterwithgartner/gartner-top-10-data-analytics-trends>
- Gartner. (2020). Gartner Information Technology Glossary. Retrieved from <https://www.gartner.com/en/information-technology/glossary>
- Gartner. (2023). Gartner Glossary: Business Intelligence (BI) Services. Retrieved from <https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi-services>
- George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. In: Academy of Management Briarcliff Manor, NY.
- Gibbert, M., & Ruigrok, W. (2010). The “what” and “how” of case study rigor: Three strategies based on published work. *Organizational Research Methods* 13(4), 710-737.
- Halper, F. (2015). Next-Generation Analytics and Platforms for Business Success. Retrieved from <https://tdwi.org/webcasts/2015/01/next-generation-analytics-and-platforms-for-business-success.aspx>

- Hatchuel, A. (2001). Toward Design Theory and Expandable Rationality: The Unfinished Program of Herbert Simon. *Journal of Management and Governance*, 5, 260-273.
- Hiatt, J. (2006). *ADKAR: a model for change in business, government, and our community*: Prosci.
- Hong, J., McArthur, D. P., & Livingston, M. (2020). The evaluation of large cycling infrastructure investments in Glasgow using crowdsourced cycle data. *Transportation*, 47(6), 2859-2872.
- Horlach, B., Schirmer, I., & Drews, P. (2019, June 8-14, 2019). *Agile portfolio management: design goals and principles*. Paper presented at the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden.
- Hunter, A., & Brewer, J. D. (2015). Designing multimethod research. In *The Oxford handbook of multimethod and mixed methods research inquiry*.
- ISACA. (2010). The Business Case Guide: Using VAL IT 2.0.
- ISACA. (2018). *COBIT® 2019 Framework: Introduction and Methodology*: ISACA.
- Jamil, G. L., & Carvalho, L. F. M. (2019). Improving Project Management Decisions with Big Data Analytics In *Handbook of Research on Expanding Business Opportunities With Information Systems and Analytics* (pp. 45-65): IGI Global.
- Jenner, S. (2012). *Managing Benefits*: The Stationery Office.
- Jenner, S., & Kilford, C. (2011). *Management of portfolios*: The Stationery Office.
- John, K., & Senbet, L. W. (1998). Corporate governance and board effectiveness. *Journal of Banking & Finance* 22(4), 371-403.
- Johnson, J. (2018). *CHAOS report: decision latency theory: it is all about the interval*: Lulu.com
- Joslin, R., & Müller, R. (2015). Relationships between a project management methodology and project success in different project governance contexts. *International Journal of Project Management*, 33(6), 1377-1392.
- Kersten, M. (2018). *Project to product: How to survive and thrive in the age of digital disruption with the flow framework*: IT Revolution.
- Khoza, L., & Marnewick, C. (2023). A conceptual framework for scaled agile success. *International Journal of Project Organisation and Management*, 15(1), 99-128.
- Knapp, J., Zeratsky, J., & Kowitz, B. (2016). *Sprint: How to solve big problems and test new ideas in just five days*: Simon and Schuster.
- Kniberg, H., & Ivarsson, A. (2012). Scaling Agile@Spotify. Retrieved from <https://treze.es/wp-content/uploads/2018/12/scaling-agile-spotify.pdf>
- Kotter, J. (2007). Leading change: Why transformation efforts fail. *Harvard Business Review* 86, 97-103.
- Kotter, J. P. (2012). *Leading change*: Harvard Business Press
- Kruchten, P. (2004). *The Rational Unified Process: An Introduction* Addison-Wesley Professional.
- Kübler-Ross, E. (2009). *Death: The Final Stage*: Simon and Schuster.
- Lappi, T., Karvonen, T., Lwakatare, L. E., Aaltonen, K., & Kuvaja, P. (2018). Toward an improved understanding of agile project governance: A systematic literature review. *Project Management Journal*, 49(6), 39-63.
- Larman, C., & Vodde, B. (2016). *Large-scale scrum: More with LeSS*: Addison-Wesley Professional.

- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700-710.
- Laukat, T., & Wedel, F. (2020). *Software vendor selection using proof of concepts* Paper presented at the IT-Management in the Digital Age, Wedel, Germany.
- Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *Academy of Management Annals* 4(1), 109-155.
- Lee, B., Rosario, L. D., Wu, H., Roberts, L., Arnold, T., Mathur, S., & Pettit, C. (2023). Assessing the monetary value of active transport and e-micromobility per unit of distance travelled: a systematic review. *Manuscript submitted for publication*.
- Lee, K., & Sener, I. N. (2021). Strava Metro data for bicycle monitoring: a literature review. *Transport reviews*, 41(1), 27-47.
- Leffingwell, D. (2007). *Scaling Software Agility : Best Practices for Large Enterprises* Upper Saddle River, NJ: Addison Wesley.
- Lenfle, S. (2008). Exploration and project management. *International Journal of Project Management*, 26(5), 469-478.
- Lenfle, S. (2016). Floating in space? On the strangeness of exploratory projects. *Project Management Journal*, 47(2), 47-61.
- Loch, C., DeMeyer, A., & Pich, M. (2011). *Managing the unknown: A new approach to managing high uncertainty and risk in projects*: John Wiley & Sons.
- Loch, C., & Sommer, S. (2019). The Tension Between Flexible Goals and Managerial Control in Exploratory Projects. *Project Management Journal*, 50(5), 524-537.
- Lorenzi, P., & Waterman, R. H. (1985). In Search of Excellence. *Journal of Organizational Behavior*, 6(3), 246-247.
- Lovullo, D., & Kahneman, D. (2003). Delusions of success. *Harvard Business Review* 81(7), 56-63.
- Luna, A. J. d. O., Kruchten, P., Pedrosa, M. L. d. E., Neto, H. R., & de Moura, H. P. (2014). State of the Art of Agile Governance: A Systematic Review. *International Journal of Computer Science & Information Technology (IJCSIT)*, 6.
- Mahmoud-Jouini, S. B., Midler, C., & Silberzahn, P. (2016). Contributions of design thinking to project management in an innovation context. *Project Management Journal*, 47(2), 144-156.
- Maniak, R., Midler, C., Lenfle, S., & Le Pellec-Dairon, M. (2014). Value management for exploration projects. *Project Management Journal*, 45(4), 55-66.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science* 2(1), 71-87.
- Martinsuo, M. (2013). Project portfolio management in practice and in context. *International Journal of Project Management*, 31(6), 794-803.
- Mason, H., & Wiggins, C. (2010). A Taxonomy of Data Science. *Data Science @ISiM*(9). Retrieved from https://introdasci.dlilab.com/pdf/A_Taxonomy_of_Data_Science.pdf
- Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*, 37(3), 101284.

- Mathur, S. (2005). *How PMBOK-RUP-ITIL integration contributes to successful software development projects*. Paper presented at the PMI® Global Congress 2005—Asia Pacific, Singapore. <https://www.pmi.org/learning/library/pmbok-rup-til-software-development-7573>
- Mathur, S. (2007). *Project portfolio management techniques*. Paper presented at the PMI® Global Congress 2007—Asia Pacific, Hong Kong, People's Republic of China.
- Mathur, S. (2019). *Unique Characteristics of Data-Analytics Programs*. Paper presented at the PMI Australia Conference, Gold Coast. <https://www.pmiaustraliaconference.org.au/wp-content/uploads/2019/06/Sandeep-Mathur.pdf>
- Mathur, S. (2021a). *A Framework to Manage Data Science Initiatives*. Paper presented at the 21st International Conference on Knowledge, Culture and Change in Organizations, Auckland, New Zealand. https://cgscholar.com/cg_event/events/M21/proposal/54339
- Mathur, S. (2021b). *Using Data Science Initiatives to deliver Smart Infrastructure and improve Customer Experience – A Transport for NSW Case Study*. Paper presented at the PGCS - Project and Program Management Symposium, Canberra, ACT, Australia.
- Mathur, S. (2024). *When Should Data Science Initiatives be Managed as Exploratory Projects? Insights for Building Better Business Cases*. Working Paper No. 1.
- Mathur, S., Ninan, J., Vuorinen, L., Ke, Y., & Sankaran, S. (2021). An exploratory study of the use of social media to assess benefits realization in transport infrastructure projects. *Project Leadership and Society*, 2, 100010.
- Mathur, S., & Sankaran, S. (2021, 5th October 2021). Management and Governance of Data Science Initiatives (UTS Course). Retrieved from <https://open.uts.edu.au/uts-open/study-area/built-environment/management-and-governance-of-data-science-initiatives2/>
- Mathur, S., Sankaran, S., & MacAulay, S. (2023). *Project Delivery for General-Purpose Technologies: New Insights from Data Science Initiatives*. Working Paper No. 3.
- Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2021). *A Framework to Manage Data Science Initiatives*. Paper presented at the The 6th PMI Research and Academic Virtual Conference 2021, Mumbai, India.
- Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2022). *Minimum Viable Governance for Data Science Initiatives - A Transport for NSW Case Study*. Paper presented at the IPMA 10th Research Conference, Belgrade, Serbia.
- Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2023). **Minimum Viable Governance for Data Science Initiatives - A Transport for NSW Case Study**. In *Research Handbook on the Governance of Projects* (pp. 488): Edward Elgar Publishing Limited.
- Mathur, S., Sankaran, S., MacAulay, S., & Tsang, I. (2024). *Unique characteristics of Data Science Initiatives*. Paper presented at the IPMA 28th International Congress on Project Management, Arandjelovac, Serbia.
- McGrath, R., & McManus, R. (2020). Discovery-Driven. *Harvard Business Review* 98(3), 124-133.
- Mullaly, M. (2009). The '3 Bears' Theory of Project Management. *ProjectManagement.com*. Retrieved from <https://www.projectmanagement.com/articles/246650/the--3-bears--theory-of-project-management>

- Mullay, M. (2010). The Goldilocks Theory of Governance: How To Get To 'Just Right'. *ProjectManagement.com*. Retrieved from <https://www.projectmanagement.com/articles/258853/the-goldilocks-theory-of-governance--how-to-get-to--just-right-#> =
- Müller, R. (2009). *Project Governance* Gower Publishing, Aldershot, UK.
- Müller, R. (2016). Governance Models and Paradigms. In *Governance and Governmentality for Projects* (pp. 50-64): Routledge.
- Müller, R., & Lecoivre, L. (2014). Operationalizing governance categories of projects. *International Journal of Project Management*, 32(8), 1346-1357.
- Müller, R., Turner, J. R., Andersen, E. S., Shao, J., & Kvalnes, Ø. (2016). Governance and ethics in temporary organizations: The mediating role of corporate governance. *Project Management Journal*, 47(6), 7-23.
- Myrbakken, H., & Colomo-Palacios, R. (2017, 4-5 October 2017). *DevSecOps: A Multivocal Literature Review*. Paper presented at the Software Process Improvement and Capability Determination: 17th International Conference, SPICE 2017, Palma de Mallorca, Spain.
- NewVantage Partners. (2022). Data and AI Leadership Executive Survey 2022. Retrieved from <https://www.businesswire.com/news/home/20220103005036/en/NewVantage-Partners-Releases-2022-Data-And-AI-Executive-Survey>
- Ninan, J., Ke, Y., Sankaran, S., Mathur, S., Vuorinen, L., & Devkar, G. (2022). Social media for improving metro rail project operations. In *Social Media for Project Management*: CRC Press, Taylor and Francis.
- O'Reilly III, C. A., & Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *Academy of management Perspectives*, 27(4), 324-338.
- Obie, H. O., Chua, C., Avazpour, I., Abdelrazek, M., Grundy, J., & Bednarz, T. (2019). A study of the effects of narration on comprehension and memorability of visualisations. *Journal of Computer Languages*, 52, 113-124.
- Pollack, J., Helm, J., & Adler, D. (2018). What is the Iron Triangle, and how has it changed? *International Journal of Managing Projects in Business*.
- Pope, C., Ziebland, S., & Mays, N. (2000). Analysing qualitative data. *Bmj*, 320(7227), 114-116.
- Project Management Institute. (2016). *Governance of Portfolios, Programs, and Projects* Global standards.
- Project Management Institute. (2017a). *The standard for portfolio management* (4th ed.).
- Project Management Institute. (2017b). *The standard for program management* (4th ed.).
- Project Management Institute. (2019). *Benefits Realization Management : A Practice Guide*.
- Project Management Institute. (2021a). Beyond Agility: Flex to the Future. *Pulse of the Profession*. Retrieved from <https://www.pmi.org/learning/library/beyond-agility-gymnastic-enterprises-12973>
- Project Management Institute. (2021b). *A Guide to the Project Management Body of Knowledge (PMBOK guide) and The Standard for Project Management* (7th ed.): Project Management Institute, Inc.
- Rathore, M. M., Ahmad, A., Paul, A., Wan, J., & Zhang, D. (2016). Real-time medical emergency response system: exploiting IoT and big data for public health. *Journal of Medical Systems* 40(12), 283.

- Reddy, R. C., Bhattacharjee, B., Mishra, D., & Mandal, A. (2022). A systematic literature review towards a conceptual framework for enablers and barriers of an enterprise data science strategy. *Information Systems and e-Business Management*, 20(1), 223-255.
- Reinertsen, D. (2009). *The Principles of Product Development FLOW: Second Generation Lean Product Development*: Celeritas Publishing.
- Rollins, J. (2015). *Foundational Methodology for Data Science*: IBM.
- Ross-Hellauer, T., Tennant, J. P., Banelytė, V., Gorogh, E., Luzi, D., Kraker, P., . . . Vignoli, M. (2020). Ten simple rules for innovative dissemination of research. *PLoS computational biology*, 16(4), e1007704.
- Saberi, M. (2021). Mapping bicycling ridership across Sydney metropolitan area with Strava data. Retrieved from <https://medium.com/strava-metro/mapping-bicycling-ridership-across-sydney-metropolitan-area-with-strava-data-aada5b6954d7>
- Saltz, J. S., & Shamshurin, I. (2016, 5-8 December 2015). *Big data team process methodologies: A literature review and the identification of key factors for a project's success*. Paper presented at the 2016 IEEE International Conference on Big Data (Big Data), Washington D.C., USA.
- Saltz, J. S., & Stanton, J. M. (2017). *An introduction to data science*: Sage Publications.
- Sandberg, J. (2005). How do we justify knowledge produced within interpretive approaches? *Organizational research methods*, 8(1), 41-68.
- Sankaran, S., Lawler, C., Ang, K., & Mathur, S. (2021). Governance, Portfolio and Program Management (Master's Course). Retrieved from <https://handbook.uts.edu.au/subjects/15346.html>
- SAS Institute. (2009). *Getting Started with SAS Enterprise Miner 6.1*: SAS Institute.
- Scaled Agile. (2023). Scaled Agile Framework (SAFe) 6.0. Retrieved from <https://www.scaledagileframework.com/>
- Scholz, R. W., & Tietje, O. (2002). *Embedded case study methods: Integrating quantitative and qualitative knowledge*: Sage.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534.
- Severtson, W. R. B., Franks, L., & Ericson, G. (2017). What is the team data science process? *Microsoft Azure*.
- Sheffield, J., & Lemétayer, J. (2013). Factors associated with the software development agility of successful projects. *International Journal of Project Management*, 31(3), 459-472.
- Shenhar, A. J. (2001). One size does not fit all projects: Exploring classical contingency domains. *Management Science* 47(3), 394-414.
- Shenhar, A. J., Dvir, D., Morris, P., & Pinto, J. (2004). How projects differ and what to do about it. *The Wiley Guide to Project, Program and Portfolio Management* 1265-1286.
- Small, M. L. (2009). How many cases do I need?' On science and the logic of case selection in field-based research. *Ethnography*, 10(1), 5-38.
- Standards Australia. (2016). SA/SNZ TS 8019:2016 Governance of Benefits Realization for IT-enabled investments.
- Stobierski, T. (2021). What's the difference between Data Analytics & Data Science. *Harvard Business School - HBS Online*. Retrieved from <https://online.hbs.edu/blog/post/data-analytics-vs-data-science>

- Straits, B. C., & Singleton, R. (2018). *Social research: Approaches and fundamentals*: Oxford University Press.
- Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, 57(1), 103146.
- Sutherland, J. (2019). SCRUM@ Scale Guide. *The definitive guide to Scrum@ Scale: scaling that works*. Scrum Inc.
- Taborda, L. J., & Mohammed, M. (2021). *Exploring Agile Project Governance*. Paper presented at the ANZAM Conference, Perth, WA, Australia.
- The Agile Alliance. (2001). Manifesto for Agile Software Development. Retrieved from <https://agilemanifesto.org/>
- Thiry, M. (2016). *Program Management*: Routledge.
- Thorp, J. (2003). *The Information Paradox: Realizing the Business Benefits of Information Technology*: McGraw-Hill Ryerson.
- Torres, A. R., Khemici, N., & Paré, M. (2017). The hot potato game: roles and responsibilities for realizing IT project benefits. *The Journal of Modern Project Management*, 5(2).
- Transport for NSW. (2018). Future Transport Strategy 2056. Retrieved from <https://future.transport.nsw.gov.au/sites/default/files/media/documents/2021/FutureTransportStrategy2056.pdf>
- Transport for NSW. (2021). Future Transport Technology Roadmap 2021-2024. Retrieved from https://future.transport.nsw.gov.au/sites/default/files/media/documents/2021/NSW_Future_Transport_Technology_Roadmap_2021-2024.pdf
- Tricker, B., & Tricker, R. I. (2015). *Corporate Governance: Principles, Policies, and Practices*: Oxford University Press.
- Turner, J. R., & Cochrane, R. A. (1993). Goals-and-methods matrix: coping with projects with ill defined goals and/or methods of achieving them. *International Journal of Project Management*, 11(2), 93-102.
- University of Technology Sydney. (2022). *Graduate Research Candidature Management, Thesis Preparation and Submission Procedures* (Version 1.13 ed.).
- Van de Ven, A. H. (2017). The innovation journey: you can't control it, but you can learn to maneuver it. *Innovation*, 19(1), 39-42.
- Van de Ven, A. H., Polley, D. E., Garud, R., & Venkataraman, S. (1999). *The Innovation Journey* Oxford University Press.
- Van Maanen, J. (2015). The present of things past: Ethnography and career studies. *Human Relations*, 68(1), 35-53.
- VentureBeat. (2019). Why do 87% of data science projects never make it into production? Retrieved from <https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/>
- Viaene, S., & Van den Bunder, A. (2011). The Secrets to Managing Business Analytics Projects. *MIT Sloan Management Review*, 53(1), 65.
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.
- Ward, J. S., & Barker, A. (2013). Undefined by Data: A Survey of Big Data Definitions. *arXiv preprint arXiv:1309.5821*.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1999). *Unobtrusive Measures* (Vol. 2): Sage Publications.

- Westergren, U. H., Saarikko, T., & Blomquist, T. (2017). theIoTguide: A Business Guide to the Internet of Things. Retrieved from <https://www.diva-portal.org/smash/get/diva2:1110187/FULLTEXT02>
- Williams, T. C. (2017). Lean Project Governance. In *Filling Execution Gaps* (pp. 157-192): De Gruyter.
- Yin, R. K. (2018). *Case Study Research and Applications : Design and Methods* (6th ed.): SAGE Publications, Inc.
- Zhang, X., Le, Y., Liu, Y., & Chen, X. (2021). Fostering ambidextrous innovation strategies in large infrastructure projects: a team heterogeneity perspective. *IEEE Transactions on Engineering Management*.
- Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, 215-225.