UNIQUE CHARACTERISTICS OF DATA SCIENCE INITIATIVES

Sandeep Mathur[1], Shankar Sankaran[2], Samuel MacAulay[3], Ivor Tsang[4]

[1]Greenpeace Australia Pacific; [2]University of Technology Sydney; [3]The University of Queensland; [4]University of Technology Sydney, Australia

Abstract Data is increasingly pervasive in organisational contexts, shaping investments from inception to benefits realisation. Data Science Initiatives (DSIs) have gained traction to unlock value from data, yet their success rate remains a concern, with estimates suggesting a significant failure rate. This paper employs a multi-method approach, utilising six DSIs as case studies within Transport for NSW, supplemented by semi-structured interviews with practitioners across various Australian organisations. The study reveals that DSIs possess unique characteristics that render traditional ICT-enabled program management practices ineffective. It offers practical guidance for practitioners to enhance DSI delivery efficiency. Furthermore, the paper delineates three key stages - Exploitation, Transition, and Exploration - characterising the delivery of DSIs. Theoretical contributions extend to the exploration of uncertainty within data science projects, elucidating their impact on DSI success or failure. Future research avenues may entail validating these findings across diverse public and private sector organisations internationally.

Keywords: *Exploratory Projects, Program Management, Change Management, Data Science, Data Management, Agile Methodology*

1. INTRODUCTION

Data Science Initiatives[1] (DSIs) have unique challenges that make the application of traditional program management techniques problematic. These challenges arise primarily due to uncertainty they carry in data being ingested which has a cascading impact on scope, schedule and ultimately value creation.

In all DSIs, data from various known and unknown sources is ingested into a data store and transformed and insights are generated using this transformed data. 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

[1] We use the term Data Science Initiative (DSI) to describe investments in Data Analytics, Business Intelligence and Data Science including Machine Learning and Artificial Intelligence technologies.

Corresponding author: Sandeep Mathur, sandeep@mathur.com.au

https://doi.org/1056889/lvje9472
PAGE 79-90

neither technologies nor customer requirements are known at the start of the project. The uncertainty in DSIs makes it difficult to manage them as Exploitative Projects which focus on optimising cost-quality-time triple constraints to deliver new products and services (Lenfle, 2008). The fundamental tension between exploitation of old certainties and exploration of new possibilities identified by March (1991) is relevant to DSIs.

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

2. LITERATURE REVIEW

The focus of this research is to improve practices for delivery of DSIs 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 are:

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

Significant literature exists covering various domains of ICT initiatives delivery. However, gap exists in literature in relation to DSIs especially in acknowledgement of uncertainty, complexity and exploratory nature which influences the management and governance of DSIs.

2.1. 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 using standard project methodology which largely focusses on delivery of a defined scope, cost and schedule. We also interpret that Waterfall methods do not allow effective risk management of ICT-enabled projects thus causing high failure in delivery.

2.2. DATA SCIENCE INITIATIVES

Our case studies showed us need for domains such as Program Management, Change Management, Data Management, Data Science and Development-Operations (DevOps) to deliver DSIs. DSIs often include implementation of Artificial Intelligence (AI) and Machine Learning (ML). For the DSIs included as case studies in this research, we found that it was never about bringing in data from one source and closing the project. We see DSIs being typically implemented as a Program on a continuous spectrum rather than a single one-off Project.

We thus argue that DSIs are typically exploratory projects due to uncertainty around benefits and delivery itself. Other than setup of foundation infrastructure, the delivery work packages of DSIs cannot be clearly defined at the outset and thus the schedule cannot be prepared in detail. DSIs are unable to conform to the rational Waterfall approach of the projects in delivery of unique product, service, or a 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 further lean themselves to be delivered in a progressive elaboration using Agile methods.

2.3. SUMMARY AND IMPLICATIONS

A review of related literature shows a gap exists in how the DSIs are defined in portfolios and how they are executed as programs. This research addresses the gap through the 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 rich literature and proven delivery frameworks which have matured over the past three decades (Axelos, 2022; Project Management Institute, 2016, 2017, 2019, 2021b). However, the failure rate in delivery of DSIs point to a gap in their ability 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

literature and practitioners of the emerging data-science domain. Without the proposed body of work, there will be more failed programs, dissatisfied sponsors and delay much needed investment in this emerging domain as well as delay the benefits that will flow from harnessing the data and the nuggets in it.

3. RESEARCH SETTING AND METHODS

Taking a practice lens on delivery of DSIs guided us to focus on full life cycle of DSIs. 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, we chose to study delivery of DSIs within a single organisation (Transport) between January 2017 to December 2020 where the primary author of this paper was employed and delivered DSIs.

3.1 RESEARCH METHODS

A multi-methods approach 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 to analyse six DSIs in Transport, each of which provided a unique scope and opportunity to understand characteristics of DSIs. Our interest was to understand characteristics of DSIs as experienced by the organisation's participants themselves and identify uniqueness with this class of initiatives to bring in improvements within the organisation. The six DSIs chosen as case studies reflect the chronology and the maturity of Transport in delivering DSIs (**Error! Reference source not found.**).

Using an interpretive research tradition associated with case-studies, ontological and epistemological assumptions on DSI characteristics emerged which were externally validated with practitioners from five organisations delivering DSIs using semi-structured interviews. The interviews used open-ended questions to gain lived experienced of interviewees. Interpretive approach (Sandberg, 2005) to justify knowledge produced was adopted by analysing interview transcripts leading to coherent interpretations of DSI characteristics.

3.2 RESEARCH SETTING

Our 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 NSW in Australia. The research method uses participant-observation technique and multiple case studies over full program life cycle covering a period of five years collecting DSI data. The research has used documentation, archival records, direct observations, participant-observation, and physical artifacts as source of data.

4. DATA COLLECTION AND ANALYSIS

The six DSIs chosen as case studies represent contemporary phenomenon in depth and within its real-world context at Transport that was particularly useful for our research question because the organisation needs to better understand the unique characteristics of DSIs. **Error! Reference source not found.** provides a summary of the six DSIs.

Table 1. Summary of six Transport DSIs

Program	**Description**	**Period**	**Business Case Benefits**	**Benefits Delivered**	**Budget (AUD)**
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	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.	$5.14m
Ferry	Implement evidence-based Ferry Contract Management & improved customer experience.	Apr 2018 - Dec 2020	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 dependency on Ferry Operator.	$4.8m
CTABS	Enable data analytics and verification of Provider self-reporting.	Oct 2017 - Mar 2018	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.	$289k
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	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.	$357k

Ligh Rail Priority	Provide priority to Light Rail at traffic intersections shared with other road users.	Feb 2019 - Mar 2020	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.	$1.49m
MPR	Ensure data management and architectural consistency of Operational Data Lake (ODL) across multiple performance reporting business cases.	Jan 2020 - Jun 2021	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.	$4.4m

The chronology of six DSIs has been bracketed into three stages: Exploration, Transition and Exploitation that Transport went through as the six DSIs were delivered. When the author commenced delivering his first DSI (Vanguard) as traditional ICT Program, he faced challenges in managing the schedule. The planned milestones were not met. In hindsight, the organisation was not aware of the exploratory nature of DSIs. However, as we progressed, we started acknowledging the unique characteristics and making changes to the delivery processes. At macro level, we map this initial stage to Transport's "exploration" stage. Transport's "Transition" stage maps to the organisation accepting the uniqueness of DSIs, adapting to delivery processes, and building skills to deliver DSIs successfully. "Exploitation" stage refers to a mature state where organisation accepts that datasets come with uncertainty; agile methods are practiced, and management accepts DSI business cases without measurable benefits. **Error! Reference source not found.** shows the timeline and highlights of the six DSIs indicating author's journey from uncertainty and frustration of not being able to deliver program outcomes as per the schedule to acceptance of exploratory nature of DSIs and ability to plan for the uncertainty and engage the stakeholders effectively.

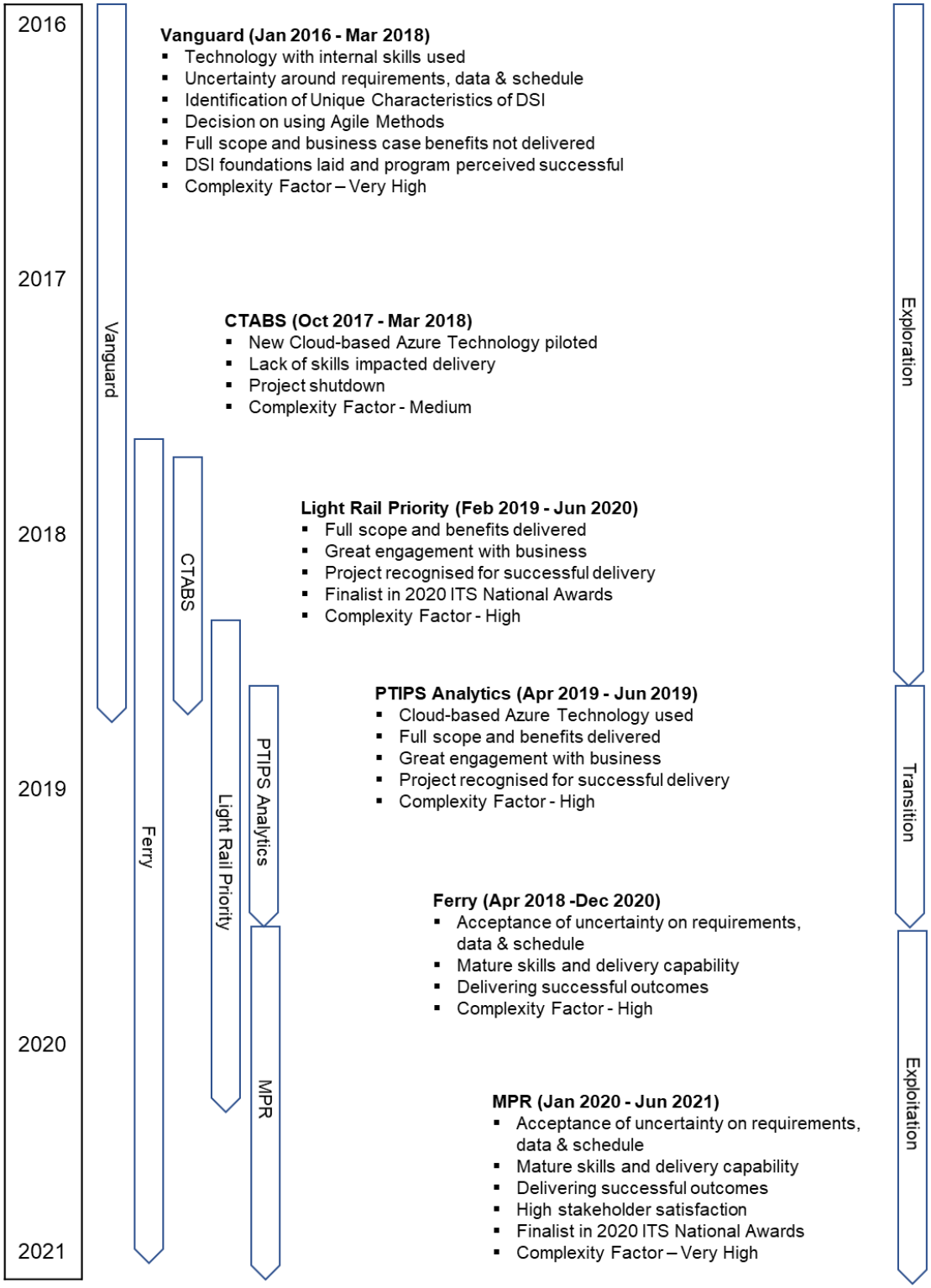

Figure 1. Overview of six Transport for NSW DSIs

While the scale of the DSIs is different, together they have allowed us to identify characteristics of DSIs which brought in uncertainty in their management and governance.

5. CONCLUSION AND RECOMMENDATION

In this section, we 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 our conclusion.

5.1 CONCLUSION

We conclude that current literature does not adequately cover unique characteristics of DSIs, and the business managers and practitioners 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.

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 of exploitative initiatives. We identify six unique characteristics of DSIs to be used in delivery of DSIs and complement domains identified in literature review and practice - PMI's The Standard for Program Management (Project Management Institute, 2017) for program management; 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. The practitioners should consider integrating these domains in any DSI delivery framework they are developing to ensure that envisaged value is delivered and address some of the challenges primary author faced over five years in delivering DSIs. While each of the highlighted domain is rich in information and mature, the lack of integration will cause continued failure of DSIs.

We also conclude that organisation and teams go through stages of Exploitation, Transition and Exploration in delivery of DSIs. The process of delivering DSIs becomes efficient as they deliver more of them. With exception of data from well-defined and structured source, every new dataset carries uncertainty in scope and quality. This brings in framing of an underpinning exploration component to a DSI combined with a shift to exploitative as the organisation and teams mature.

5.2 LIMITATIONS AND IMPLICATIONS OF RESEARCH

This research has used six DSIs from one public sector organisation in Australia as case studies to identify unique characteristics and validated with semi-structured interviews with practitioners from five organisations. Sandberg (2005) notes that truth is always something unfinished within the interpretive tradition, the criteria proposed do not enable researchers to generate absolute truth claims. We believe that the DSI characteristics we 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 currency of the research work being done as some of the characteristics will change as the maturity DSIs changes from being exploratory to exploitative.

5.3 RECOMMENDATIONS

The six unique characteristics are summarised in Table 1:

Table 1. Unique Characteristics of DSIs

No	Description
(i)	DSIs carry high degree of uncertainty right from initiation through to closing phases except for when the data is from a well-defined and structured source
(ii)	DSIs are often enablers for decision making & 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 act as an enabler to next one
(v)	Skills required to deliver a DSI are different to those required for a typical ICT program
(vi)	DSIs do not end and after initial delivery convert into managing the product, model and data

We suggest additional research to validate the characteristics with other public and private sector organisations delivering DSIs. As the field is evolving rapidly, the authors believe that the six characteristics identified in this paper will also evolve. These characteristics 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, 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 improving data-driven decisioning capability.

REFERENCES

1. Asay, M. (2017). 85% of big data projects fail, but your developers can help yours succeed. *Big Data Tech. Rebuplic, 11*, 1-5.
2. Axelos. (2022). *Managing successful programmes* (5th ed.): PeopleCert.
3. 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.
4. Earley, S. E. (2017). DAMA-DMBOK: Data management body of knowledge. *Technics Publications*.
5. Hiatt, J. (2006). *ADKAR: a model for change in business, government, and our community*: Prosci.
6. Lenfle, S. (2008). Exploration and project management. *International Journal of Project Management, 26*(5), 469-478.
7. March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science 2*(1), 71-87.
8. 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-itil-software-development-7573
9. Project Management Institute. (2016). *Governance of Portfolios, Programs, and ProjectsGlobal standards*.
10. Project Management Institute. (2017). *The standard for program management*(4th ed.).
11. Project Management Institute. (2019). *Benefits Realization Management : A Practice Guide*.
12. 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
13. 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.
14. Sandberg, J. (2005). How do we justify knowledge produced within interpretive approaches? *Organizational research methods, 8*(1), 41-68.
15. Scaled Agile. (2023). Scaled Agile Framework (SAFe) 6.0. Retrieved from https://www.scaledagileframework.com/