

A framework to assist in the governance and management of data in the digital ecosystem

by Avirup Dasgupta

Thesis submitted in fulfilment of the requirements for
the degree of

**Doctor of Philosophy (C02029)
Information Systems**

under the supervision of:

Dr Asif Q.Gill

Dr Farookh Hussain

University of Technology Sydney
Faculty of Engineering and Information Technology
School of Computer Science
September 2021

Certificate of Authorship

I, Avirup Dasgupta, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in Information Systems in the Faculty of Engineering and IT, School of Computer Science 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.

Signature: Production Note:
Signature removed prior to publication.

Date: 18 March 2022

Acknowledgments

I have been fortunate to have Dr. Asif Gill and Dr. Farookh Hussain as my supervisors.

First of all, I want to express my sincere gratitude to my principal supervisor, Dr. Asif Gill, for giving me the opportunity to enroll as a PhD student at UTS and work with him on this research project. His highly valuable coaching, feedback, support and careful guidance helped me finish this research project. Dr Asif Gill was always there to provide support in face to face meetings, phone conferences, and by email. His valuable understanding in the field of digital ecosystem as an academic, researcher, and practitioner provided a wealth of knowledge that I used in this research project

I am especially thankful to my wife Chandni, who has stood by me all along and encouraged me for doing this study. I want to express my sincere thanks to all my colleagues for their valuable research inputs and support and my brother and my parents for his continuous encouragement.

I also wish to thank the Australian Government for its support for providing the funding for my research project for the length of my PhD study period.

I am thankful to all of the reviewers for their valuable feedback and comments.

Thank you all.

Table of Contents

TITLE	I
LIST OF FIGURES	8
LIST OF TABLES	10
LIST OF ABBREVIATIONS.....	12
PUBLICATIONS	13
ABSTRACT	14
CHAPTER 1: INTRODUCTION	16
1.1 RESEARCH BACKGROUND	18
1.2 RESEARCH PROBLEM	24
1.3 RESEARCH QUESTIONS AND OBJECTIVES.....	28
1.4 RESEARCH STAKEHOLDERS.....	30
1.5 RESEARCH SCOPE AND ASSUMPTIONS.....	31
1.6 RESEARCH STRATEGY.....	33
1.7 RESEARCH FINDINGS	35
1.8 CONTRIBUTION TO KNOWLEDGE	38
1.9 THESIS OUTLINE	40
1.10 CHAPTER SUMMARY.....	41
CHAPTER 2: LITERATURE REVIEW	42
2.1 KEY CONCEPTS	43
2.1.1 <i>Digital ecosystem</i>	43
2.1.2 <i>Digital Technology</i>	45
2.1.3 <i>Digital Data</i>	47
2.1.4 <i>Data Governance and Data Management</i>	50
2.2 DATA BREACHES.....	58

2.3	REVIEW OF FRAMEWORKS	61
2.3.1	<i>SLR Filtration Process</i>	61
2.3.2	<i>Review and analysis of existing literature</i>	65
2.4	RESEARCH GAPS AND QUESTION	76
2.5	DATA REGULATIONS	79
2.6	CHAPTER SUMMARY.....	85
CHAPTER 3: RESEARCH METHOD.....		86
3.1	RESEARCH IN INFORMATION SYSTEMS.....	87
3.1.1	<i>Case Study</i>	87
3.1.2	<i>Descriptive</i>	87
3.1.3	<i>Action Research</i>	88
3.1.4	<i>Design Research</i>	88
3.1.5	<i>Action Design Research</i>	88
3.2	CHOICE OF RESEARCH METHOD FOR THIS STUDY	89
3.3	USE OF DSR IN THIS THESIS	91
3.3.1	<i>Awareness of the Problem</i>	92
3.3.2	<i>Suggestion and Development</i>	93
3.3.3	<i>Evaluation</i>	94
3.3.4	<i>Conclusion</i>	96
3.4	RESEARCH INSTRUMENT.....	98
3.4.1	<i>Data Sources</i>	99
3.4.2	<i>Survey</i>	100
3.4.3	<i>Ethics Approval</i>	101
3.5	CHAPTER SUMMARY.....	103
CHAPTER 4: THE 4I FRAMEWORK.....		104
4.1	FRAMEWORK OVERVIEW.....	105
4.1.1	<i>Drivers</i>	108
4.1.2	<i>Elements</i>	113

4.1.3	<i>Stages of the Framework</i>	133
4.2	STAGES AND ELEMENTS	138
4.3	IMPLEMENTATION OF THE FRAMEWORK	139
4.4	CHAPTER SUMMARY.....	141
CHAPTER 5: FRAMEWORK EVALUATION.....		142
5.1	FRAMEWORK EVALUATION OVERVIEW.....	143
5.2	SCENARIO-BASED TESTING DEMONSTRATION.....	146
5.2.1	<i>Use Case: Application of the 4I Framework to support regulations</i>	147
5.2.2	<i>Use Case: Application of the 4I Framework to support data protection</i>	154
5.2.3	<i>Use Case: Application of the 4I Framework to support ethical data usage</i>	159
5.3	EMPIRICAL EVALUATION: SURVEY	163
5.3.1	<i>Survey plan</i>	163
5.3.2	<i>Design of the survey</i>	163
5.3.3	<i>Survey procedure</i>	163
5.3.4	<i>Survey respondent profile</i>	164
5.3.5	<i>Survey questions</i>	166
5.3.6	<i>Survey data collection</i>	167
5.3.7	<i>Survey data analysis</i>	167
5.3.8	<i>4I Framework enhancement (based on feedback)</i>	172
5.4	NOVELTY OF THE FRAMEWORK	176
5.5	CHAPTER SUMMARY.....	178
CHAPTER 6: DISCUSSION AND SUMMARY		179
6.1	RESEARCH TIMELINE.....	180
6.2	RESEARCH SUMMARY AND OUTPUT	181
6.3	RESEARCH LIMITATIONS AND FUTURE WORK	188
6.4	CONTRIBUTION TO RESEARCH	190
6.5	CONTRIBUTION TO PRACTICE	191

CONCLUSION.....	192
BIBLIOGRAPHY	193
APPENDIX A: ETHICS APPROVAL.....	203
APPENDIX B: SURVEY QUESTIONNAIRE	206
APPENDIX C: RECRUITMENT EMAIL	219
APPENDIX D: CONSENT FORM.....	222
APPENDIX E: VERSIONS OF THE 4I FRAMEWORK	224

List of Figures

Figure 1-1: Ecosystems.....	18
Figure 1-2: Data Management Functions (adapted from DMBOK).....	22
Figure 1-3: Cumulative number of data-related regulations (OECD).....	25
Figure 1-4: Research Question and Research Objective.....	29
Figure 1-5: Data Governance and Data Management relationship (in this thesis)	31
Figure 1-6: Design Science Research process	33
Figure 1-7: Research Strategy.....	34
Figure 1-8: The 4I Framework.....	36
Figure 1-9: Lack of governance in DE managed through cross-organisational role.....	39
Figure 2-1: Anatomy of the digital ecosystem.....	44
Figure 2-2: Digital Technologies enabling DE	45
Figure 2-3: DIKW hierarchy.....	48
Figure 2-4: DG principles (Brous, Janssen & Vilminko-Heikkinen 2016a, 2016b).....	52
Figure 2-5: Data Lifecycle	56
Figure 2-6: Three stage Systematic Literature Review filtration process	62
Figure 2-7: Data sharing among ecosystem partners in DE.....	73
Figure 2-8: Research Gap and Questions.....	76
Figure 3-1: Design Science Research process	89
Figure 3-2: Research Timelines	91
Figure 3-3: Interpretations of IS ontology	94
Figure 3-4: Design Research Guidelines	97
Figure 4-1: The framework components.....	106
Figure 4-2: Data Compliance perspective.....	110
Figure 4-3: Data protection perspective.....	110
Figure 4-4: Operational Efficiency perspective	111

Figure 4-5: Data monetisation perspective	112
Figure 4-6: Elements in the 4I Framework	113
Figure 4-7: Data Asset as a subset of data	114
Figure 4-8: Data Asset relationship to the organisational objectives.....	115
Figure 4-9: Data Risk as a subset of Operational Risk	117
Figure 4-10: Data Risk relationship to Data Asset.....	118
Figure 4-11: Relationship of Guidelines with other elements of the Framework.....	123
Figure 4-12: Key processes or procedures to support data lifecycle.....	123
Figure 4-13: Ecosystem Actors' roles and responsibilities.....	126
Figure 4-14: Actors in the 4I Framework	126
Figure 4-15: Sample governance structure of intra-organisational actors	129
Figure 4-16: Inter-organisational governance operating model.....	131
Figure 4-17: Implementation Activities	139
Figure 5-1: Criteria evaluated using scenarios.....	144
Figure 5-2: Criteria evaluated using surveys	145
Figure 5-3: Wearable smart IoT-enabled ecosystem	156
Figure 5-4: Laws related to health data.....	157
Figure 5-5: Data Collection Consent	161
Figure 5-6: Respondent designation profile.....	164
Figure 5-7: Respondents' demographic profile	165
Figure 5-8: Overall rating of the framework.....	171
Figure 6-1: Research Timeline.....	180
Figure 6-2: Research Questions, Objectives and Output	182

List of Tables

Table 1-1: Thesis Scope.....	32
Table 1-2: Organisation of the thesis chapters.....	40
Table 2-1: Commonly used definitions of DG.....	53
Table 2-2: Difference between IT and DG	55
Table 2-3: DG and DM Terminologies.....	57
Table 2-4: Data breach incidents	59
Table 2-5: Paper selection criteria	62
Table 2-6: Search results.....	62
Table 2-7: Studies used for the Literature Review.....	64
Table 2-8: Categorisation of Frameworks into Elements captured.....	68
Table 2-9: Gaps.....	75
Table 2-10: Data-related legislation in Australia.....	81
Table 2-11: Regulations mapped to activities.....	83
Table 3-1: Research Methods used in this research	91
Table 3-2: Evaluation Criteria	96
Table 3-3: Conformity of the research to Design Research Guidelines.....	97
Table 3-4: Research Instruments.....	98
Table 3-5: 14-step Procedure from Data Collection to Data Erasure	102
Table 4-1: Notations used	108
Table 4-2: Elements of the 4I Framework	113
Table 4-3: Taxonomy of the Data Asset element.....	116
Table 4-4: Key taxonomy of the Data Risk Element	118
Table 4-5: Key guidance to support data-related decision-making.....	120
Table 4-6: Key processes and procedures.....	124
Table 4-7: Key Technology related sub-elements of the 4I Framework.....	125

Table 4-8: Key internal actors.....	127
Table 4-9: Key external actors.....	128
Table 4-10: Key Steps of the Identify Stage.....	134
Table 4-11: Key Steps of the Insulate Stage.....	135
Table 4-12: Key steps of the Inspect Stage.....	136
Table 4-13: Key steps of the Improve Stage.....	137
Table 4-14: Sample mapping of the proposed elements of the framework to the stages.....	138
Table 4-15: Sample Implementation Steps.....	140
Table 5-1: Evaluation Criteria.....	143
Table 5-2: Mapping of Scenarios to Evaluation Problems.....	146
Table 5-3: Key obligations identified at the Identify stage.....	149
Table 5-4: Key activities to monitor risk from external actors.....	151
Table 5-5: Statistical Results on the stages from the Empirical Survey.....	169
Table 5-6: Statistical Results on the stages from the Empirical Survey.....	169
Table 5-7: Suggested changes and gap mitigation approach.....	172
Table 5-8: Concepts and sub-concepts derived from the survey responses.....	174
Table 5-9: High-level concepts derived from the survey responses.....	175
Table 6-1: Key components of the 4I Framework.....	185
Table 6-2: Key contributions.....	190

List of Abbreviations

Abbreviation	Description
APP	Australian Privacy Principles
DE	Digital Ecosystem
DG	Data Governance
DGM	Data governance and management
DM	Data Management
DSR	Design Science Research
FEIT	Faculty of Engineering and IT
GDPR	General Data Protection Regulation
IoT	Internet of Things
RO	Research Objective
RQ	Research Question
RSQ	Research Sub Question
SLR	Systematic Literature Review

Publications

Publication #	Reference	Source
Publication-1	Dasgupta, A. & Gill, A.Q. 2017, 'Fog Computing Challenges: A Systematic Review', paper presented to the Australasian Conference on Information Systems, Hobart.	https://opus.lib.uts.edu.au/bitstream/10453/124785/1/ACIS2017_paper_182_RIP.pdf
Publication-2	Dasgupta, A., Gill, A. & Hussain, F.K. 2019, 'A Conceptual Framework for Data Governance in IoT-enabled Digital IS Ecosystems', Data Science Technology and Applications (DATA), pp. 209-16.	https://www.scitepress.org/Papers/2019/79243/79243.pdf
Publication-3	Dasgupta, A., Gill, A.Q. & Hussain, F. 2019, 'Privacy of IoT-Enabled Smart Home Systems', IoT and Smart Home Automation, IntechOpen	https://www.intechopen.com/chapters/65738
Publication-4	Dasgupta, A., Gill, A.Q. & Hussain, F. 2020, 'A Review of General Data Protection Regulation for Supply Chain Ecosystem', Innovative Mobile and Internet Services in Ubiquitous Computing, eds L. Barolli, F. Xhafa & O.K. Hussain, Springer International Publishing, Cham, pp. 456-65.	https://link.springer.com/chapter/10.1007/978-3-030-22263-5_44

Abstract

The digital ecosystem (DE) continues to grow with the proliferation of new digital offerings every day, a trend that is expected to accelerate rapidly in the next few years. The digital ecosystem involves several players, platforms and industries that provide solutions based on advanced technologies such as the Internet of Things (IoT), cloud computing, analytics and artificial intelligence. The data-driven digital ecosystem provides organisations with the information they need to make better insightful decisions for monetary benefits. However, there are a few challenges. There is limited guidance available on how to effectively establish integrated data governance and management for the data-intensive digital ecosystem. The existing approaches focus on individual organisations rather than the ecosystem. There is a need to look beyond the boundary of a single enterprise. To address data governance and management concerns, this thesis applies the design science research (DSR) approach to develop a framework that can be utilised to create an integrated data governance and management capabilities for a focal enterprise in the DE. Rather than having a fixed one-size-fit-all approach, the framework focusses on the adaptability approach to address the changing business and regulatory landscape.

The framework has three major components: Drivers, Elements and Stages. The driver has four key purposes (e.g., Data Compliance, Data Protection, Monetisation and Operational Efficiency). Drivers provide justification to conduct data governance and data management activities. The element component comprises of six elements (e.g., Data Asset, Data Risk, Guidance, Processes and Procedures, Ecosystem Actors and Technology) and the underlying attributes. Elements provide stakeholders key toolkits to govern and manage data. There are four key stages: 1) Identify, 2) Insulate, 3)

Inspect and 4) Improve. The stages provide guidance to achieve objectives of drivers with the elements. The framework is evaluated through scenario-based testing and survey. The results indicate that the framework is reasonably suited to support integrated data governance and management activities in different organisational contexts.

Keywords: Data governance, Data Management, Framework, Compliance, Digital Ecosystem, GDPR, Australian Privacy Principles, Data Breach

Chapter 1: Introduction

Digital ecosystems (DEs) are rapidly evolving with an increasing number of organisations working together to create interconnected offerings that provide significant benefits when compared to one company's single product or service offering. Strong collaboration among companies is now the foundation that determines the success of an enterprise in the digital ecosystem. Data is the critical strategic asset that enables organisations to meet the demands of customers. However, the shift in focus from single enterprise to multi-enterprise context, has given rise to newer challenges in terms of collecting, processing, storing and accessing data in a secure and regulatory compliant manner.

Though traditional approaches to the governance and management of data seem useful; however, there is a lack of understanding as to what it should look like in the digital ecosystem context. The regulatory changes around data and the rapid growth of data have aggravated the data challenges. There is an inadequate understanding around what laws are in place when an organisation handles data and what needs to be done to comply with the regulations. Corporations are under pressure to overcome the urgent problems they face in handling data in a multi-organisational context. This has resulted in renewed attention given to data governance (DG) and data management (DM) approaches (van den Broek & van Veenstra 2015), with an increasing number of organisations exploring how they can better manage risk related to data by effectively controlling data within the company as well as across company boundaries. There are very few published studies in the literature regarding data handling in an ecosystem context. However, given the rise in data volume, data exchange between companies, privacy and security concerns, there is a necessity to explore how to assist

in the end-to-end governance and management of data in the digital ecosystem as part of the operational and business activities.

This study contributes by examining the current practices in the data landscape from the perspective of the DG and DM office of a focal enterprise in DE. The study involves: a) determining DG and DM related challenges arising due to data exchange among ecosystem players and investigating the suitability of existing frameworks for DE, b) identifying the key data-related legislative requirements in Australia and across the globe and c) the development and evaluation of a framework that provides guidance on governing and managing data for the DE.

This chapter begins by discussing the research background in Section 1.1. The research problem is discussed in Section 1.2. The research questions and objectives are discussed in Section 1.3. The target stakeholders are discussed in Section 1.4. Section 1.5 describes the scope of the research along with the assumptions. Section 1.6 outlines the research strategy. Section 1.7 summarises the findings of this thesis. Section 1.8 explains the contribution of this research while Section 1.9 provides a summary of the remaining chapters of this thesis before concluding in Section 1.10.

1.1 Research Background

The term ecosystems describes the interaction between living organisms and their non-living environment (Tansley 1935). Since its inception, the term ecosystem has been used in different areas including biology, business, management, technology and innovation, leading to more specific ecosystems such as business ecosystems (Pappas et al. 2018). The term business ecosystem was officially introduced in 1993 by Dr James F. Moore, from Harvard Law School. Moore's view was that innovative businesses are unable to grow in a vacuum and are dependent on different resources such as partners, capital and customers. Moore advocated the growth of cooperative networks where companies would co-evolve their capabilities in a business ecosystem around an innovation, both by working cooperatively and competitively, to work towards the next innovations (Moore 2006). See Figure 1-1.

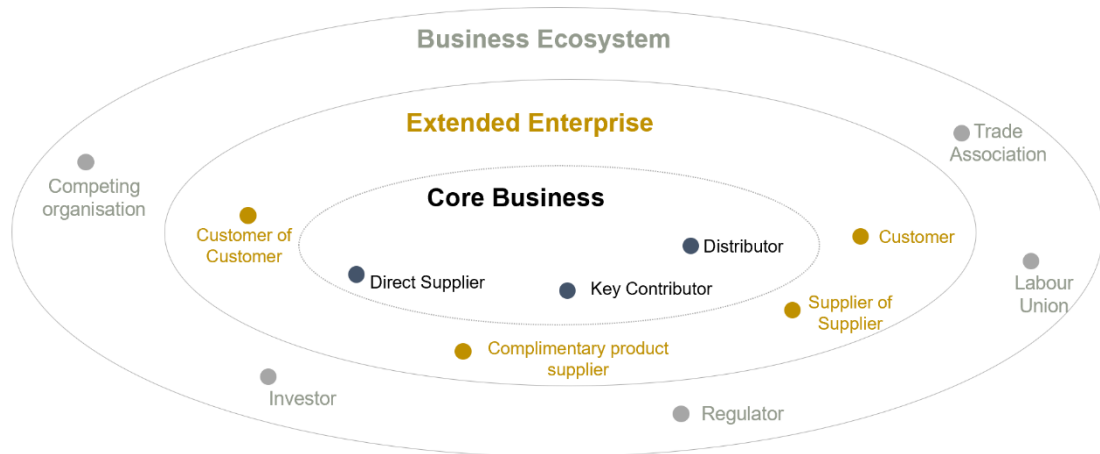


Figure 1-1: Ecosystems

In 2002, the philosophy of the business ecosystem further gained momentum when the digital business ecosystem (DBE) emerged as an extension of the business ecosystem (Li, Badr & Biennier 2012). DBE is made up of two tiers: a digital (ecosystem) and business (ecosystem) (Stanley & Briscoe 2010). While the term business ecosystem

refers to an economic community of individuals and organisations that operate outside their traditional industry boundaries, the term digital ecosystem (DE) relates to a virtual environment made up of digital entities such as hardware, software applications and processes (Nachira, Dini & Nicolai 2007). DEs analyse IT-based complex relationships among players in the IT industry, including suppliers of software, hardware, telecom operators, application developers as well as the end user (Tsujimoto et al. 2018) and non-direct ecosystem stakeholders such as local governments, communities and legislators (Lütjen et al. 2019) for value creation.

The DE has seen an increase in interest in several disciplines including information systems (Senyo, Liu & Effah 2019). The focus of the DE is to offer a solution to a customer with a wide range of needs instead of focussing on one segment at a time. Common examples of digital ecosystems are the IoT-enabled healthcare ecosystem, web or mobile applications involving payment gateways and financial services ecosystems. An instance of DE is that of Danske Bank, a Denmark-based financial institution. Danske Bank utilised its network of partners and businesses to develop an web-based system that combined bank customer data with real-estate market listings, thereby providing prospective homebuyers cost estimates on tax, heating and electricity (Bennett 2017). Another example is that of pharmaceutical companies. In addition to supplying drugs as a product, medical companies now offer wellness solutions, comprising of monitoring and real-time adjustment to the patient, and a preventive package (Wright & Jones 2018).

It is an established fact that in today's world, the most prominent businesses are those that bring together several distinct groups of entities in a network. One of the main reasons behind DE's rapid expansion is that with the right product or service provided

by third parties, organisations can lower in-house development and maintenance costs, improve productivity in addition to providing new values for customers. DE's rapid evolution has been possible primarily due to the progression of Artificial Intelligence (AI), Internet of Things (IoT), cloud computing and other technologies, in addition to the reduction in the price of hardware devices such as sensors. McKinsey, a leading management consulting firm, estimates that by 2025, DEs will be generating \$60 trillion that amounts to 30% of corporate earnings, with companies which implement an ecosystem approach having revenues higher than those without (Bughin, Catlin & Dietz 2019).

Cross-border and cross-enterprise data access, usage, storage, retention, disposal and exchange are crucial to the growth of DE. The ability of an organisation to succeed in the DE is dependent on how well the organisation is able to adapt and change their way of functioning from controlling data in silos to governing and managing it in a wider context. Therefore, it is essential that for the DE to thrive, the exchange of data throughout its lifecycle is governed and managed with a common vision.

Governance refers to the exercise of authority and control over an organisation, process or geopolitical area. It was originally tied to government and associated with the process of setting, controlling, administering and monitoring adherence to policy. Governance has been adapted into other spaces such as information technology (IT) governance for IT and office systems, corporate governance for business environments, cloud governance for cloud services, information governance and data governance (DG) to manage data.

In 1995, Horne was first to connect governance with the optimal use of assets and treated data as an asset that drives the significance of governing data in an organisation

(Horne 1995). DG can be viewed through different lenses or interpretations, as discussed by Begg & Caira (2012). Although there is no single, recognised and universally accepted definition of DG, researchers generally refer to DG as how data-related decisions are taken within an organisation, who has the authority to take actions, with what information, at what time, under what condition using what techniques (Khatri & Brown 2010; Welfare 2019). DAMA International, an association of data and information management, considers DG to be a framework for decision rights and accountabilities over the management of its data assets. On the other hand, EDM Council, a Global Association created to elevate the practice of Data Management, considers DG as the rules of engagement for data management, focused on the implementation of policies, standards, and operational procedures to ensure that stakeholders comply.

Data management (DM), sometimes referred to as resource management or enterprise information management, focuses on the proper generation, storage, and retrieval of data to make good business decisions (Rob & Coronel 1997). DM emerged as a practice from the 1960s when the Association of Data Processing Service Organisations (ADPSO) began providing data management advice for professionals. DM is a much broader and older concept in comparison to DG and encompasses the entire lifecycle of the data asset. The Data Management Body of Knowledge (DAMA-DMBOK) formally defines it as the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles. DAMA identified the following 10 functions that relate to data management (Mosley et al. 2010) activities as shown in Figure 1-2.

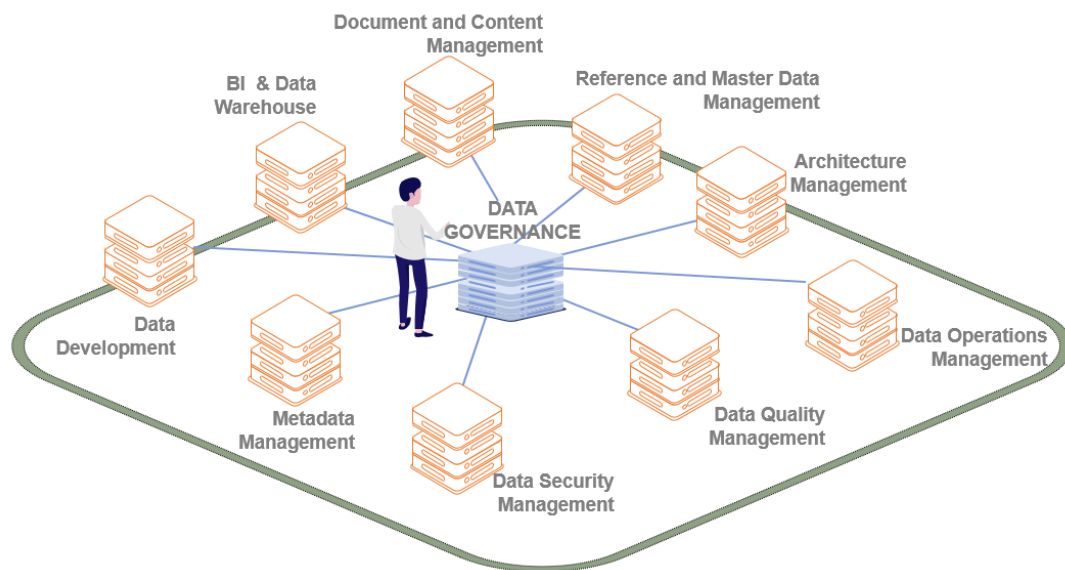


Figure 1-2: Data Management Functions (adapted from DMBOK)

Of the 10 functions, DG is a key constituent of the discipline of DM and primarily relates to an organisation's overall data management strategy. It is important to note that the main difference between the terms 'governance' and 'management' is that governance refers to the decisions that must be made and who makes these decisions in order to ensure the effective management and use of resources, whereas management involves implementing decisions (Alhassan, Sammon & Daly 2018). The differentiation is based on the proposal put forward by the International Organisation for Standardization (ISO) regarding Governance and Management (ISO/IEC 2008). Thus, all deliverables of DG such as rules and roles depend on the definition, scope and constituent components of DM (STEENBEEK 2019).

In the DE, enterprises utilise the cooperative and coordinated efforts of firms to drive meaningful customer interactions, leading to positive business results. DG and DM practices not only need to be customer and stakeholder driven, but they must also

support the various requirements related to security, privacy, transborder obligations and regulatory aspects. This means, to utilise data successfully, consideration needs to be given to enhance the overall existing knowledge from a technological and legal perspective.

Framework is a term that has been frequently used in the information system and practitioner literature. Notable frameworks are the Zachman framework (Zachman 1987) and TOGAF (Jonkers, Proper & Turner 2009). Miles et al. (1994) defined a conceptual framework as a means to explain, either graphically or in narrative form, the main things such as the main factors, constructs or variables, and the relationships between them. To manage information in DE, there is a need to have a framework which combines the governance and management fundamentals aligned with organisational change and stakeholder buy-in (TDAN.com 2015).

.

1.2 Research Problem

This section discusses the research problem. A digital ecosystem is not confined to a single institution or managed as a single database. The infrastructure required for DE spans multiple data providers, computing paradigms, data types, connections, domains and locations. All actors in DE need to interact and collaborate actively to increase value with a heavy reliance on cross-organisational data flows.

In the past, a siloed data exchange meant that there were less requirements that a firm needed to comply with. The established existing practices were adequate to satisfy the firms' data governance or management needs. However, the review of the current challenges and risks faced in relation to DG and DM reveals that with the growing importance of data and the expansion of the DE, the established ideas around data access and exchange of data, are no longer valid (Lis & Otto 2021).

Enterprises are now facing a variety of new challenges. One issue is that unknowingly and unsuspectingly, consumer data is collected by unaccountable entities with whom no obvious or direct relationship exists and is used in ways that are hidden from the data subject (Pike 2019). In light of the evolution of the ecosystem concept, organisations now need to have a holistic view of exchange of data in a multivendor-enabled business model and environment. A recurring theme of discussion among industry practitioners and academicians in the DE is how to govern and manage data exchange in a diverse ecosystem where data is shared, transformed, added, enriched and processed for different purposes by different procedures and different players.

Stakeholders in DE need to be able to address the risks arising due to incompetent DM techniques or the ungoverned capturing, storing, sharing and deletion of data. If not mitigated properly, poor DG and DM can result in the financial and reputational

decline of a business. This is evident from the sharp increase in the number of data security and privacy incidents resulting in data breaches involving high profile organisations like Amazon and Facebook (Isaak & Hanna 2018; Kuhn 2018).

Closely related to data breaches are the regulations. Recently, there has been a sharp rise in the number of updates made by government agencies to data-related mandates (see Figure 1-3) with an emphasis on protecting the privacy and security of data.

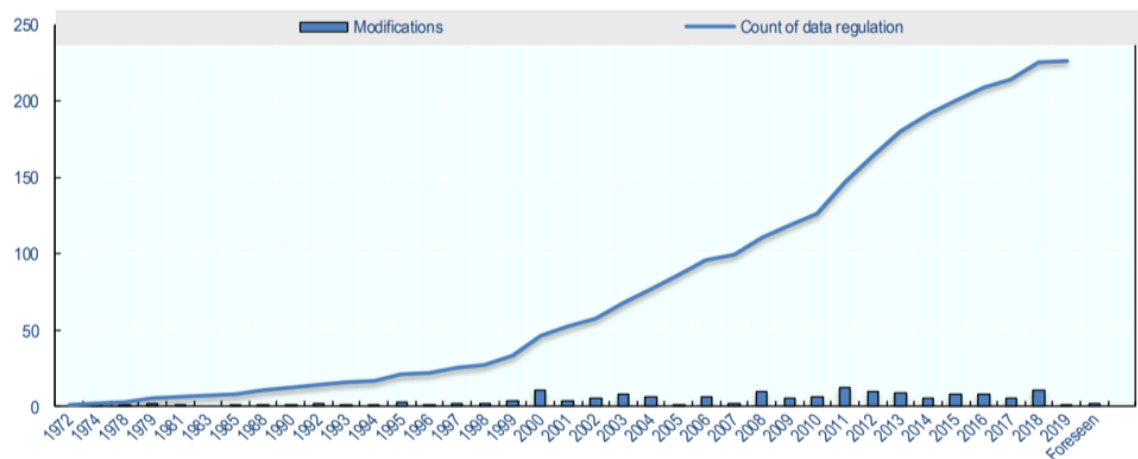


Figure 1-3: Cumulative number of data-related regulations (OECD)

Such regulatory requirements have aggravated the existing DG and DM approaches, making it increasingly difficult to fulfil compliance obligations using the existing approaches. For example, from a regulatory perspective, organisations and service providers need to adhere to country-specific data regulations, like Australian Privacy Protection (APP) law, California Consumer Privacy Act (CCPA) and General Data Protection Regulation (GDPR) (Esposito et al. 2017) with overlapping requirements across jurisdictions (DataGuidance 2018).

Enterprises acknowledge that data breach threats, data security, data privacy, data confidentiality, data retention and archiving, consents, regulatory requirements and a lack of clarity on data ownership are obstacles for data exchange among the DE actors. In summary, proper governance and management of data is an important for DE participants to become profitable and commercially successful.

To address the challenges around managing data in the DE, an increasing number of organisations are renewing their focus on redefining the strategies around DG and DM (Deloitte 2019) practices. Firms have already established a wide variety of initiatives to meet the demands of the DE (Engels 2019), but several issues remain. Surveys such as that conducted by Redgate in 2018, a software company based in England, found that 52 percent of companies lack understanding of the requirements around governing and managing data (Redgate 2018).

Information systems researchers often emphasise the importance of DG's role in the effectiveness of data-enabled platforms (Prieëlle, Reuver & Rezaei 2020). A number of frameworks exist that help us understand how firms operate today when it comes to governing and managing data. However, most of the existing frameworks (Gantait, Patra & Mukherjee 2018; Sicari 2018; Virgilio A.F. Almeida 2015; Weber 2016) are based primarily on single organisation practices. Acting in the DE demands a much more holistic ecosystem wide approach than a siloed approach. To be effective in the DE context, the requirement is to have a framework that not only encompasses an all-inclusive approach to collect, exchange, secure, store and delete data across company boundaries, but also delivers structured guidance on the rules of engagement for business and management activities among ecosystem partners (Valdez-De-Leon 2019). There is scant literature on DG and DM in an ecosystem context (Lee, Zhu &

Jeffery 2019; Lis & Otto 2020; Yannuzzi et al. 2017) with majority of the studies focussing on specific topics such as ownership, technology or industry. They do not resolve the questions related to practical implementation of DG and DM in multi-party environment. This provides the motivation and opportunity to develop a framework that extends beyond a single organisation and can assist the data-related business functions to address the several DG and DM concerns in the DE. Therefore, this thesis focuses on developing a dependable framework for the information-intensive DE, which can assist organisations in their governance and management of data, both in an intra- and inter-organisational context.

1.3 Research Questions and Objectives

The shortcomings mentioned in the previous sections provide the justification for conducting this research study on DG and DM in DE. The main goal of this research is to develop a framework that can fill the gaps in end-to-end governing and managing data in inter-organisational context.

To do this, the research intends to find the answers to the following research questions (RQs).

Research Question (RQ): *How to effectively assist in the governance and management of data in the distributed digital ecosystem?*

The requisite knowledge required to answer this RQ was gained through finding answers to several sub-questions that were derived from the main research question.

The subordinate questions investigated are:

- Research Sub-Question 1 (RSQ1): *What are the key challenges of the digital ecosystem from a data governance and management perspective?*
- Research Sub-Question 2 (RSQ2): *What are the existing regulatory requirements from a data governance and management perspective?*
- Research Sub-Question 3 (RSQ3): *How can we develop an integrated and adaptive data governance and management framework that addresses the challenges and regulatory requirements relevant to digital ecosystems?*
- Research Sub-Question 4 (RSQ4): *How can the framework be validated using the DSR evaluation criteria?*

To answer the RQs, this study pursued the following research objectives:

- a) Research Objective 1 (RO1): *Conduct a literature review of the existing data governance and management frameworks used and assess the associated challenges from a DE perspective*
- b) Research Objective 2 (RO2): *Conduct a literature review of the existing data-related regulations*
- c) Research Objective 3 (RO3): *Develop a framework based on regulations and industry standards to assist in governance and management of data in DE*
- d) Research Objective 4 (RO4): *Evaluate the proposed framework developed in RO3.*

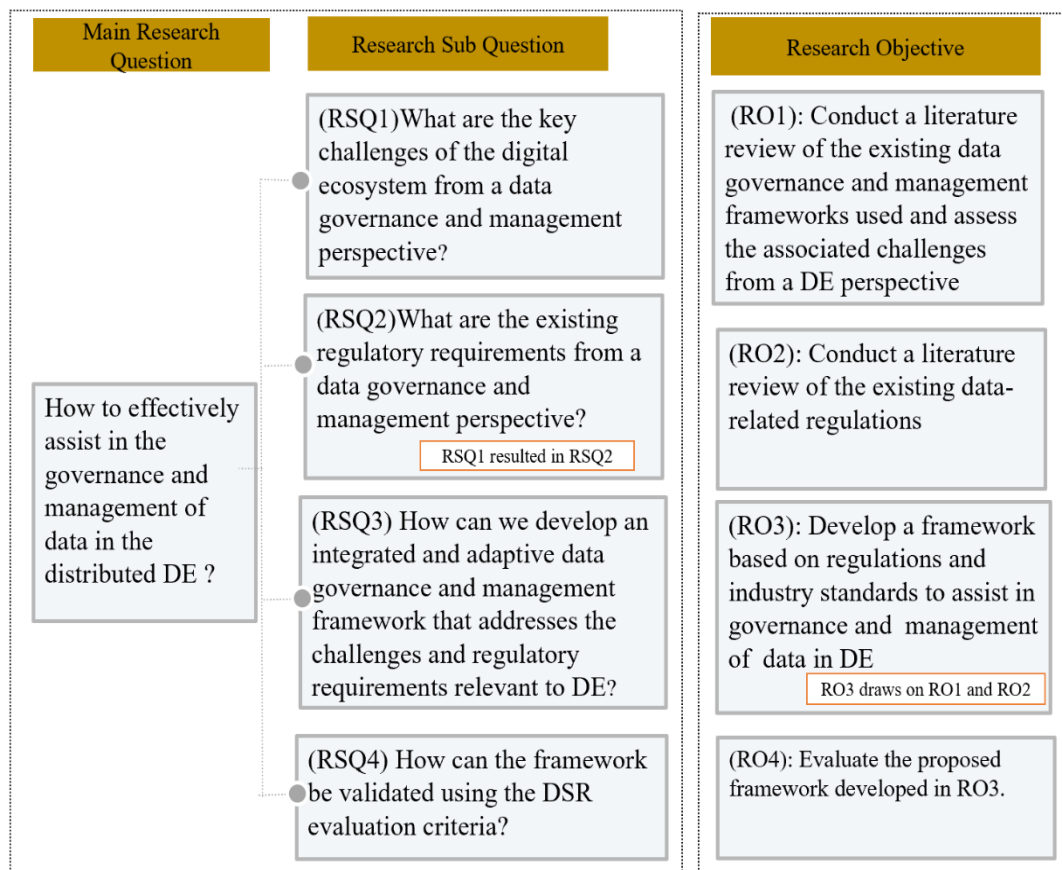


Figure 1-4: Research Question and Research Objective

1.4 Research Stakeholders

The primary stakeholders of this dissertation are those who are involved in performing, supporting and improving the data management capability and practice in an organisation. Practitioners, mainly those working for the Chief Data Officer (CDO) and Data Protection office (DPO), will find the developed framework useful when performing data governance and management functions in an organisation. The proposed research can also assist data stewards, data engineers, developers and data custodians through all stages of acquiring, sharing and controlling data effectively. Using the adaptive framework in this research, practitioners can ensure that processes and systems keep pace with frequent legislative changes such as the Australian Privacy Principle (APP) and the EU General Data Protection Regulation (GDPR). For the academics, the framework provides guidance to further enhance and evolve DE-oriented research.

1.5 Research Scope and Assumptions

Data governance and management is a vast topic. However, the scope of this research is limited to the collaborative distributed ecosystem context and covers the rules of engagement among the internal and external stakeholders. It includes the people or actors and the legal or compliance aspects (such as privacy and security) that need to be considered to manage and govern data at a macro-level in DE. This thesis chose to review the following regulations: GDPR and APP.

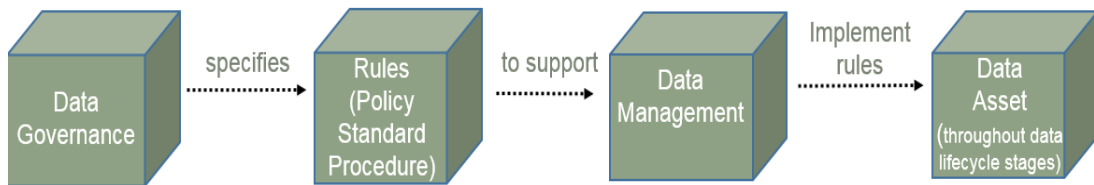


Figure 1-5: Data Governance and Data Management relationship (in this thesis)

Actual technical solutions, data architecture, data quality, meta-data, master-data management and reporting are excluded from the scope of this thesis. This study also excludes studies related to data curation and data modelling (Ladley 2019). Although extremely important, these items are already well-researched and mature. These dimensions are part of the holistic proposed framework. However, individual level of details under each topic such as data architecture, data quality, meta-data, data modelling, reporting master-data are beyond the scope of this PhD and they can be included in further research by other doctoral or research projects.

It may however be noted that this thesis briefly discusses the technical implementation details as part of three scenario-based case studies used to evaluate the artefact framework developed as part of this study.

The scope of this thesis is summarised in Table 1-1.

Table 1-1: Thesis Scope

In Scope (DE context)	Out of Scope
<p>The 4I Framework provides guidance to govern and manage data and includes the following aspects:</p> <ul style="list-style-type: none"> • Data Regulations and Risk • Actor and their interactions <p>The intended user of the framework is a focal enterprise within DE.</p>	<p>The 4I Framework will not provide guidance on technical solutions and implementations on topics related to</p> <ul style="list-style-type: none"> • Data Quality • Data Warehousing & Business Intelligence • Meta Data • Master Data • Data Modelling & Design • Data Architecture • Data Storage & Operations • Data Integration & Interoperability • Data Volume <p>The framework does not elaborate on the data monetisation aspects of DE</p>

A few assumptions were made when conducting this study. Firstly, the information mined from the available literature is accurate and represents the current state of practices. Secondly, the interpretations of the legislative mandates were satisfactorily translated into DG and DM requirements. Thirdly, the survey questions created to validate the proposed framework were adequately aligned to the research focus. Finally, the level of granularity of the developed framework was appropriate for this type of research.

1.6 Research Strategy

This research focussed on designing and building a framework to assist in solving the DG and DM challenges of DEs. Hence, Design Science Research (DSR) (Prat, Comyn-Wattiau & Akoka 2014) was used to develop and evaluate the framework. DSR is a problem-focused (Kuechler & Vaishnavi 2008) research methodology and aims to design and test an innovative product or artefact that could be used to solve a real-world problem. In DSR, the artefacts generally comprise of technical elements such as concepts, models, methods, frameworks or instantiations alongside social factors such as humans, people, work processes, teams and groups in an organisation (March & Smith 1995). The DSR research stages used in this study are outlined in Figure 1-6 and elaborated in Chapter 3.

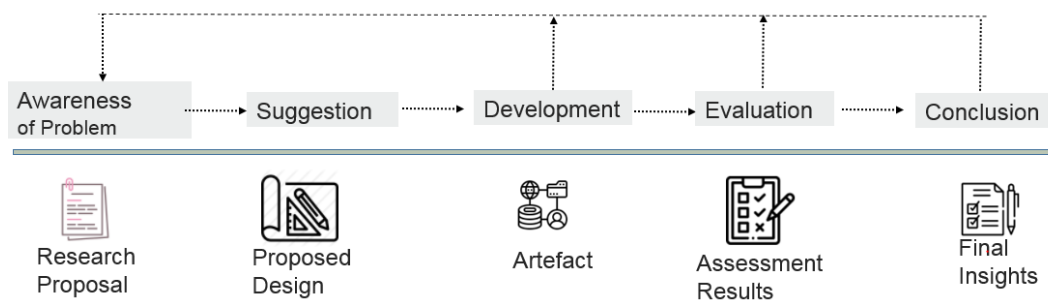


Figure 1-6: Design Science Research process

In addition, the method involved in obtaining the UTS Engineering and Information Technology Faculty Human Research Ethics Committee (FEIT HREC) panel's approval prior to engagement with human subjects with regards to the interview questions is detailed in Chapter 3.

Figure 1-7 shows the overall research strategy used in this study.

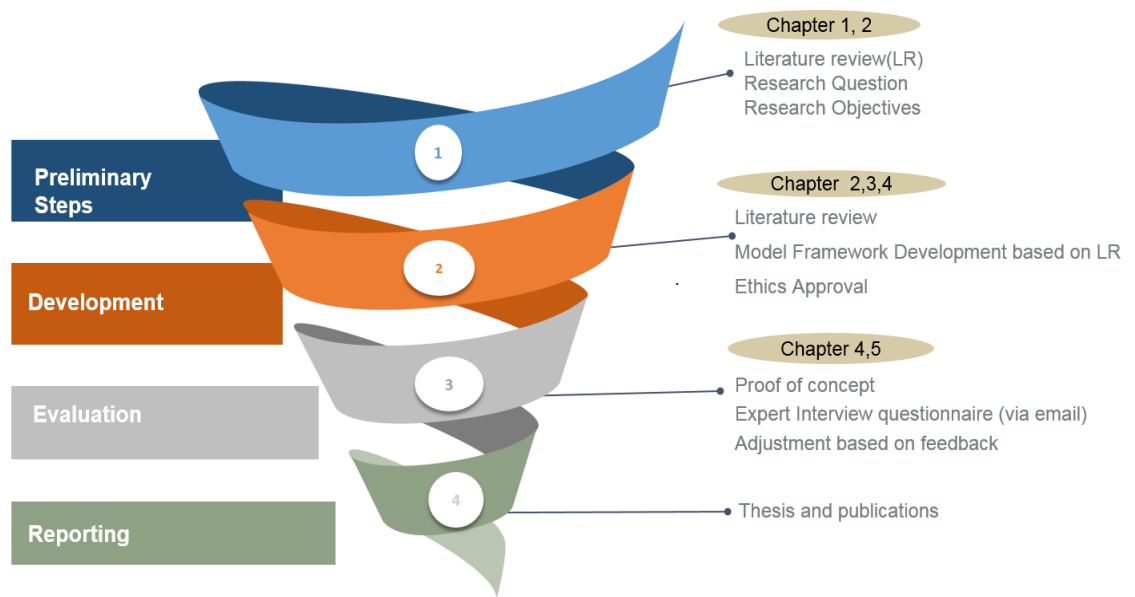


Figure 1-7: Research Strategy

1.7 Research Findings

The successful adoption of DE requires a thorough understanding of DE, DG and DM, its challenges and possible solutions.

Firstly, for RSQ1, the research reviewed existing literature on DG, DM in DE context to get a clear understanding of the role of DG and DM. This included investigating the complexities of governing and managing data in DE. Systematic literature review (SLR) was conducted to identify the existing academic and industry DG and DM related frameworks and their salient features and assess their suitability to address the industry needs in DE.

For RSQ2, the research reviewed the existing data-related regulatory requirements, particularly the General Data Protection Regulations (GDPR) and Australian Privacy Principles (APP), using literature review.

To address RSQ3, the elements identified in RQS1 and RSQ2 provided the basis to create a framework called the 4I Framework (see Figure 1-8) that can support regulations such as GDPR and Australian Privacy Principles and can be tailored for other regulations. This was supplemented with the additional data collected using the survey instrument.

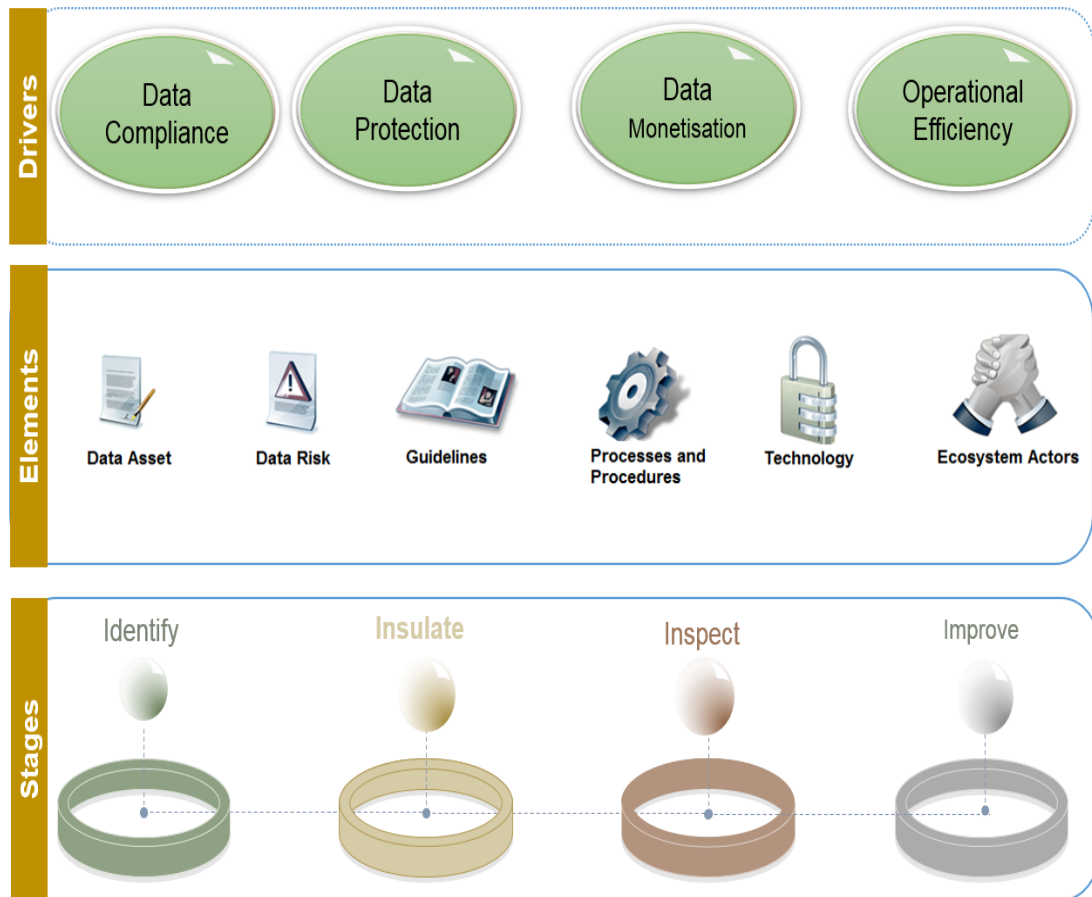


Figure 1-8: The 4I Framework

The study initially identified the key strategic initiatives that drive the need for DG and DM in the DE. The fundamental reasons to undertake a DG program in an organisation include: i) a strategy around data monetisation, ii) demonstrating compliance, iii) efficient management and monitoring of data and associated risks and iv) increased data protection resulting in better customer confidence. The study next detected six core non-overlapping high-level elements to develop the data governance and management (DGM) framework: (a) data asset, (b) data risk, (c) guidelines, (d) processes and procedures, (e) technology and (f) ecosystem actors. The first element was about the data, the benefits associated with it, and the salient features. Data risk was the second element that emerged as a crucial factor. This element was concerned

about risks faced by a focal firm from a security, privacy and external organisation perspective. The third element guidelines dealt with the actions necessary to plan and control data. Guidelines were delivered through processes and procedures. Advanced systems, tools and technologies made up the fifth element that emphasised the utility tools to manage, control and audit data. The final element was oriented towards human factors including roles, responsibilities, accountabilities, communication-collaboration plans and organisation structure.

.

1.8 Contribution to Knowledge

The contributions of this thesis are as follows:

- *The development of a framework that will provide guidance to help organisations to govern and manage data effectively in the inter-organisational and intra-organisational context in DE.*
- *The proposed framework is applied to scenarios, which shows the applicability of the proposed methods. Using the adaptive framework developed in this research, practitioners can ensure that systems adhere to laws such as the General Data Protection Principles and Australian Privacy Principles.*
- *This study synthesised the scattered knowledge of governing and managing data in DE. This is a contribution to the existing body of DGM knowledge and provides future researchers the basis to elaborate on the notion of collaborative DGM.*
- *As a part of the framework, the new co-ordinating role of the data referee was proposed (see Figure 1-9)*

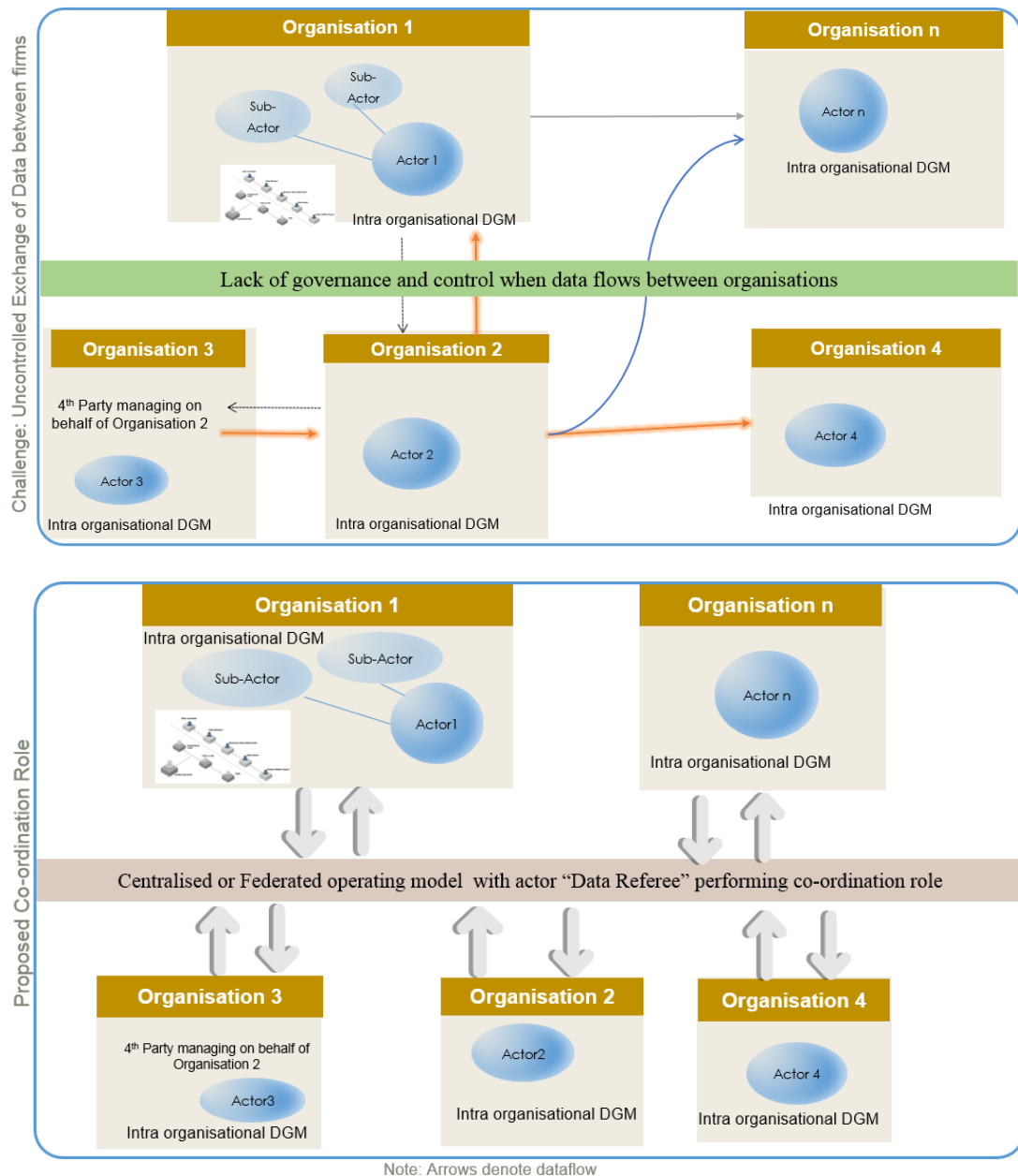


Figure 1-9: Lack of governance in DE managed through cross-organisational role

1.9 Thesis Outline

This research comprises six chapters namely: Introduction; Literature Review; Research Method; the 4I Framework; Framework Evaluation; and Discussion Summary and Conclusions.

Table 1-2: Organisation of the thesis chapters

Chapter	Description
Chapter 1 - Introduction	This chapter introduces the thesis, outlines the focus of study, and highlights its overall significance. It describes the key areas of the research to provide the research background and context of this research.
Chapter 2 - Literature Review	This chapter starts by elaborating the key concepts of the digital ecosystem, data governance and management. Approaches, techniques or practices used in governing and managing data is critically reviewed alongside the review of key data-related regulations that impacts DE.
Chapter 3 – Research Method	This chapter discusses the development of the research methods, ethical considerations and data analysis techniques.
Chapter 4 - The 4I Framework	This chapter introduces the proposed framework developed using the literature review and industry expert opinion.
Chapter 5 - Framework Evaluation	This chapter demonstrates the usage of the framework through three scenario-based examples before evaluating it based on feedback from a group of industry practitioners.
Chapter 6 - Discussion and Summary	This chapter summarises the contributions, implications of the research, limitations and outlines future research directions and final comments of this topic.

1.10 Chapter Summary

Chapter 1 introduced the purpose and perspective of this study and highlighted the problem and issues. An overview of the key topics of the research were presented to provide the research background and context. The main topics include the digital ecosystem, DG and DM, and regulations and frameworks. The research problem was identified by explaining the DE-specific challenges around DG and DM. Based on the problem at hand, the research questions and objectives were formulated. The research strategy that most effectively addressed the research problem was described along with the description of the target audience of this thesis. Chapter 1 further acknowledged the scope of the research and assumptions made. The research findings that led to the development of the 4I Framework were discussed. Finally, the contribution of this study was described before concluding the chapter by providing an overview of the thesis structure.

In the next chapter, data governance, data management and digital ecosystem-related concepts are introduced. Also, the existing literature on DG, DM and the regulations are reviewed before identifying the critical research gaps.

Chapter 2: Literature Review

The intent of this chapter is to review the relevant background literature to obtain a clear understanding of the current state of the role of DG and DM in the context of DE and arrive at the specific research questions. To support the discussion, this chapter is organised in three sections.

Firstly, this chapter starts by elaborating the core concepts and working definition of the DE, DG and DM to provide better insights into the research context.

Secondly, it reviews the existing literature to identify the existing academic and industry frameworks to govern and manage data, their salient features and assess their suitability for enterprises in the DE. Additionally, it identifies the requirements from the regulators that needs to be considered when managing data in DE.

Finally, the chapter concludes with the identification of the crucial research gaps related to DE centric DG and DM and potential research areas for investigation in this thesis.

2.1 Key Concepts

2.1.1 Digital ecosystem

A digital ecosystem (DE) is a complex data-driven mesh that relies on the transmission and exchange of data and the use of advanced big data analytics to provide essential services to end-users and involves several players. Fu (2006, p. 143) defines a digital ecosystem as a digital environment composed of digital species or digital components. These species or components can be software applications, services, knowledge, business processes, business models, training modules, contractual frameworks or law. According to Gartner reports (Gartner 2018), it comprises of multiple independent community of interacting organisms (human and nonhuman, single and organised) and their virtual and physical environments. The focus of the DE is to offer a solution to a customer with a wide range of needs instead of focussing on one segment at a time. The key reason behind the speedy adoption of DE is that with the right product or service provided by third parties, organisations can lower the in-house development and maintenance cost, improve productivity in addition to providing new value for customers (Lis & Otto 2021).

The term digital ecosystem is very broad and comprises of actors, purpose, preconditions, usage scenario, processes, things and post-conditions as illustrated in Figure 2-1.

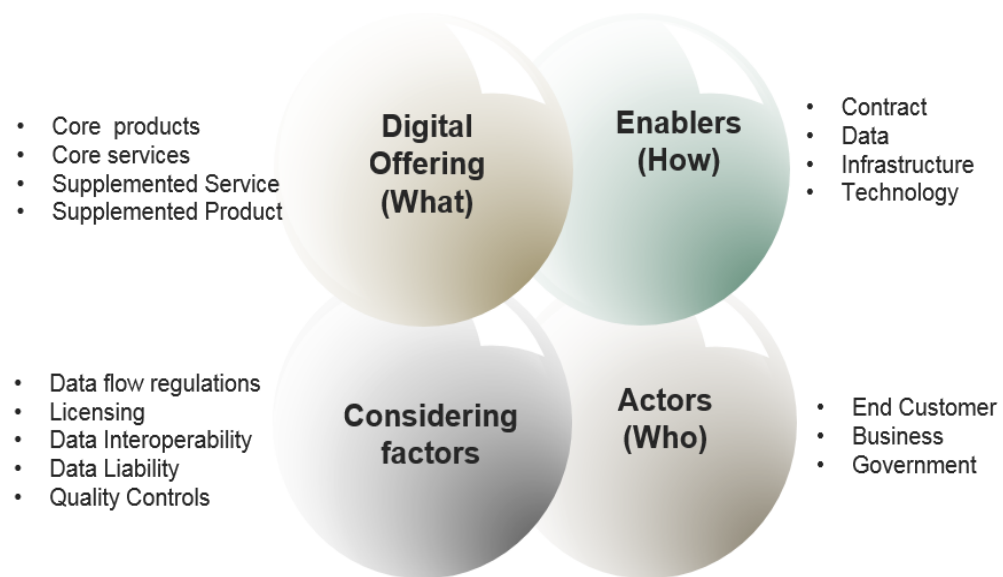


Figure 2-1: Anatomy of the digital ecosystem

Actors and Interactions

DE involves interactions among the participants and requires the formulation of a proper co-ordination effort in terms of rules of engagement and processes.

Each actor in the DE, whether a supplier, a central firm, a complementor or the customer, has different attributes, decision-making approaches and objectives (Tsujimoto et al. 2018). Each actor participates in the DE by collaborating with organisations with varied skillsets with a common purpose. These actors or partners span multiple industries and countries, linked through the Statement of Works. For example, IBM Watson has 50+ partners with whom they have business relationships, which allows them to provide a product or service that its customers need.

2.1.2 Digital Technology

The increased penetration of DE has been possible mainly due to the advancement and emergence of a diverse range of technologies such as contemporary cloud, IoT, mobile, social and big data.



Figure 2-2: Digital Technologies enabling DE

Digital technologies, either standalone or as a part of combination of technologies, has resulted in the creation of a number of innovative offerings in the diversified areas such as agriculture, energy, finance, health, medical and transport.

A. Cloud enabled digital ecosystem

Cloud computing is a computing model involving a network of interconnected servers. National Institute of Standards and Technology (NIST) of USA defines it as a model to enable ubiquitous, suitable, on-demand network access to a shared computing resource (Mell & Grance 2009). Cloud offers users to store data, deploy software applications, and manage infrastructures with minimal management effort or service provider interaction. Example of a Cloud enabled DE is that of the Digital Financial Ecosystem where digital wallet and payment gateways enable consumers to purchase retail product.

B. IoT-enabled digital ecosystem

Internet of Things (IoT) is a concept which involves a number of ‘things’ such as embedded sensors, actuators and other physical devices which are “connected” through a common network and transmits data over the network. The value of IoT is derived by aggregating the collected data from different sources and the insights provided by quality data from the connected “things” (Sullivan 2015). Example of an IoT enabled Digital Healthcare Ecosystem is that of the Medtronic’s digital meter. The meter sends an alert before a patient’s blood-glucose level reaches a threshold (Kees et al. 2015).

C. Social Technology enabled Digital Ecosystem

Social technologies can be described as technologies that support communication, engagement and association among people (Weinman 2015). Social technology is not a standalone paradigm, but a combination of technologies and practices. Airbnb, that matches room vacancy for people requiring accommodation, is an instance of the Digital Hospitality Ecosystem powered by social, mobile and cloud technology.

D. Mobile Technology enabled Digital Ecosystem

IBM defines mobile technology as a technology that goes where the user goes. That is, it allows users access to content irrespective of its location. Global Positioning System, or GPS, cell-phone towers, bluetooth beacons, and wi-fi networks are used frequently to provide the physical location with varied levels of precision. Mobile technology is characterised by internet-enabled devices like smartphones, tablets and watches. In Digital Automobile ecosystem, mobile devices are used by firms to

determine the insurance premium for car drivers (Handel et al. 2014) using a car-fitted device named Snapshot, the sends driving habits of driver in real time. In entertainment venues, users' mobile devices can be used to access advance ride reservations, obtain customer services and pay for visitor purchases.

E. Big Data Analytics Technology enabled Digital Ecosystem

Big Data is a term used to refer to data that is large, fast or difficult to process using existing tools. The process involved in capturing, storing, analyzing, enhancing and visualizing big data is a complex process (Anwar et al. 2021). Big data analytics uses techniques against diverse set of data to manage data in real time. For example, in the manufacturing industry, big data analytics are used to perform real time predictive maintenance. Rio Tinto, a mining firm belonging to the Digital Manufacturing Ecosystem, uses big data analytics to capture, clean and process large quantities of equipment data to predict the health of the tractors.

2.1.3 Digital Data

Data provides information used to make informed decisions. The benefit to stakeholders using data is best explained using the Data, Information, Knowledge and Wisdom hierarchy (see Figure 2-3 based on (Jifa & Lingling 2014)).

Each layer in DIKW layer adds certain characteristics over and above the previous one, resulting in an offering that enables stakeholders to make informed decision (Dasgupta, Gill & Hussain 2019b).

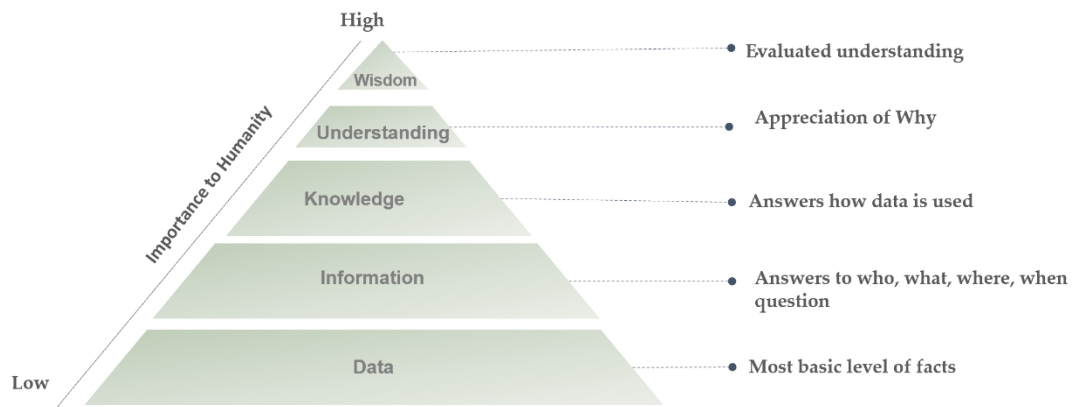


Figure 2-3: DIKW hierarchy

Digital Data is essentially a term used to represent physical things or activities in binary or discrete line code format. Digital data is the backbone of DE and is used to benefit the stakeholders through creation, collection, sharing, processing and aggregation of data using digital technologies (Gill 2021). There exists a wide variety of digital data that are produced and shared between multiple source systems spread across different organisations. Digital data can be classified differently, depending on the context such as sensitivity (personal or sensitive information), format (structured or unstructured) or production source (human or machine generated).

Examples of digital data can be finance, health, location or maps, images or videos. If we take the case of the connected car, the digital data can be about the location, road, weather, traffic conditions, the driving behavior of the driver or data about the use of entertainment, navigation and other associated services used by the car users (Kerber & L. 2018). The data collectively assists the driver in solving safety and security problems. In this instance, to achieve data protection and data privacy compliance, two important principles need to be considered:- i) the right of vehicle users to decide if data can be shared and with whom (consent), and ii) all service providers should be in

an equal, fair, reasonable and non-discriminatory position to offer services (Kerber 2019).

Throughout this thesis, the term data refers to digital data. The different attributes of data are elaborated further in Chapter 4.

2.1.4 Data Governance and Data Management

The success criteria of DE is heavily reliant on how efficiently digital data flows between firms. Traditionally, all firms perform governance and management of data to a certain extent. However, for DE, the presence of newer digital technologies like IoT, cloud, big data, regulations, multiorganisation involvement has added an extra layer of complexity. Good data, DG and DM are now a pre-requisite to ensure players involved in DE benefit from better data practices. It has now become critical that practitioners involved in DG and DM navigate the DE paradigm. In this section, the concepts of DG and DM are discussed.

As mentioned in Chapter 1, data governance (DG) is not a new concept and has been a research focus for over two decades (Kontzer 2006). DG has been an established discipline for the past two decades (Ibrahim Alhassan 2016). However, even today, one of the major complications in studying DG is the lack of consensus regarding the definition of DG. There have been several discussions on the definition of DG since its inception, as shown in Table 2-1, with none considered by industry or academia to be the official definition (Yebenes & Zorrilla 2019).

For example, according to Khatri & Brown (2010), DG involves a decision that has to be made to ensure effective data management and who makes these decisions and how, while management encompasses the implementation of the decisions. Weber, Otto & Österle (2009) considered DG to refer to decision rights, roles and accountability as part of an enterprise-wide DQM. Otto (2011) believed that ensuring compliance and improving business productivity and supporting business integration are the key goals of DG. On the other hand, Janne J. Korhonen et al. (2013) considered DG as an

organisational approach to data and information management that formalizes a set of policies and procedures to encompass the full life cycle of data, from acquisition to use and disposal. Gartner had a similar concrete view. They defined it as the process of setting decision rights and answerability for an asset, establishing policies consistent with business objectives, investing in assets that support business objectives, establishing steps to ensure compliance with corporate policies, and ensuring adequate corporate risk management (Gartner 2016b).

Several studies have linked governance, risk and compliance (GRC) or the legal framework with DG. Janssen et al. (2020) considered DG as a mechanism to provide guidelines and rules to plan and control to achieve objectives in an annual basis. DAMA's definition focused on processes, roles and formal goals; for better decision-making, assuring compliance, increasing efficiency and business integration. Likewise, Brous, Janssen & Vilminko-Heikkinen (2016a), in their systematic review of DG principles, identified four main principles: the organisation of data management, ensuring alignment with business requirements, ensuring compliance, and ensuring a common understanding of data. See Figure 2-4.

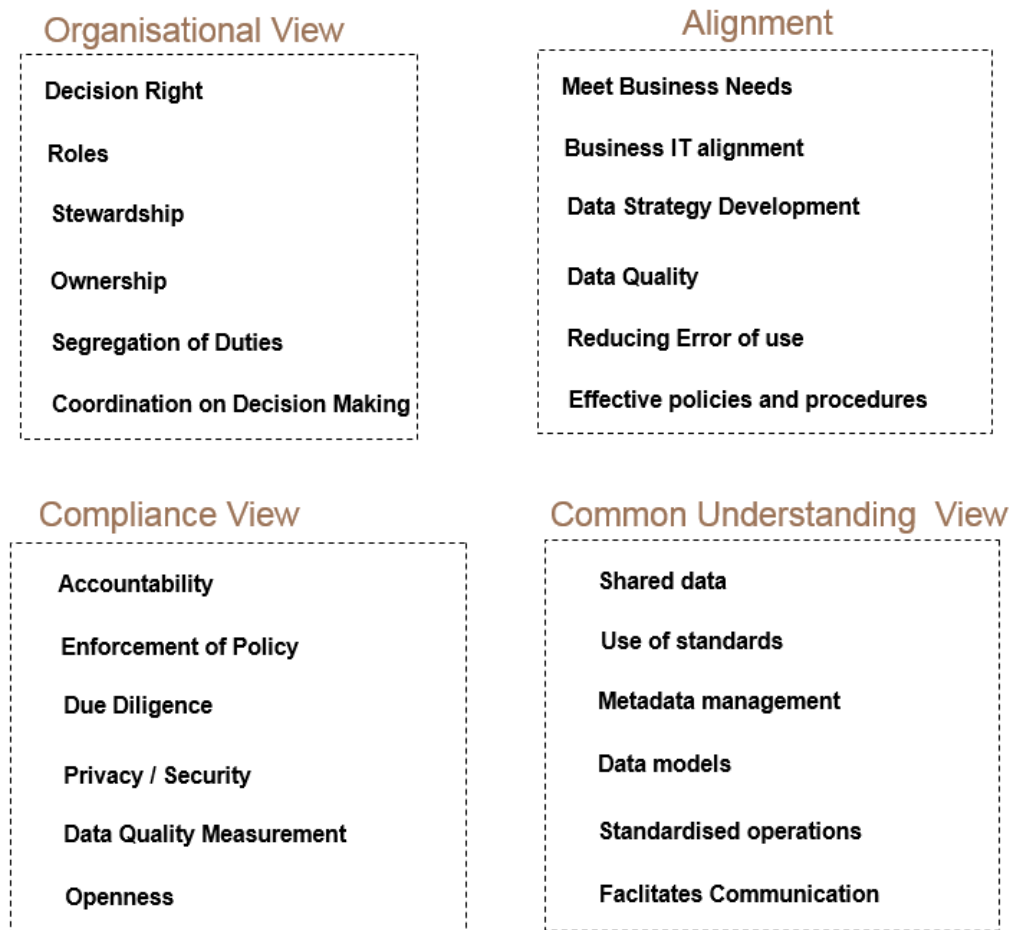


Figure 2-4: DG principles (Brous, Janssen & Vilminko-Heikkinen 2016a, 2016b)

There is no uncertainty that DG is essential for the existence of an organisation (Begg & Cairn 2012). When properly implemented, DG results in profitable data usage in an organisation. With appropriate DG in place, businesses can make insightful business decisions by setting the data in context and transforming the information to knowledge. This includes ensuring the necessary quality, completeness, availability, confidentiality, integrity and security of data throughout its lifecycle (Al-Ruithe, Benkhelifa & Hameed 2018; Dasgupta & Gill 2017).

Table 2-1: Commonly used definitions of DG

Definition	Description	Source
Ladley	DG is a required business capability which enables value to be obtained from data. DG is the organisation and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets.	(Ladley 2019)
Dyche and Level	DG is a combination of strategy and execution. It's an approach that requires one to be both holistic and pragmatic: Holistic. All aspects of data usage and maintenance are taken into account in developing the vision Pragmatic. Political challenges and cross-departmental struggles are part of the equation. So, the tactical deployment needs to be phased in to ensure quick "wins" and prevent organisational fatigue from larger, more monolithic exercises.	(Dyché & Levy 2011)
IBM	The quality control approach for adding new rigor and discipline to the process of monitoring, using, improving and protecting organisational information.	(Wróbel, Komnata & Rudek 2017)
DG Institute	The system of decision rights and accountabilities for information-related processes.	(Otto 2011)
Gartner	Specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, consumption and control of data and analytics.	(Gartner 2016a)
National Security Advisor USA	A set of processes that ensures that data assets are formally managed throughout the enterprise. A DG model establishes authority and management and decision-making parameters related to the data produced or managed by the enterprise.	(Ladley 2019)
Enlels	DG is the framework that forms the basis for dealing with and managing data for all internal and external stakeholders of an organisation.	(Engels 2019)
Janne	An organisational approach to data and information management that formalizes a set of policies and procedures to encompass the full life cycle of data, from acquisition to use and to disposal.	(Janne J. Korhonen et al. 2013)
Forrester	An agile approach to DG which focuses on just enough controls for managing risk, which enables the broader and more insightful use of data required by the evolving needs of an expanding business ecosystem.	(Brous, Janssen & Vilminko-Heikkinen 2016a)
Abraham	DG specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, DG specifies decision rights and accountabilities for an organisation's decisionmaking about its data. Furthermore, DG formalizes data policies, standards, and procedures and monitors compliance.	(Abraham, Schneider & vom Brocke 2019)

For the purpose of this study, the working definition of DG used was based on the definition put forward by Abraham and Enlels in Table 2-1. In other words, in this study *DG specifies a framework for managing data as a strategic enterprise asset and formalises data policies, standards, and procedures and monitors compliance. It forms the basis for dealing with and managing data for all internal and external stakeholders of an organisation.*

Some practitioners and researchers confuse IT governance with DG. While DG deals in the management of data assets to improve business results for business stakeholders, the focus of IT governance is primarily on IT systems and how they interoperate even though there are some overlapping areas between IT and DG (see Table 2-2). IT governance is concerned specifically with the management of IT resources such as servers, computer networks, and software applications through risk monitoring and control (Peterson 2004) generally in alignment with company goals and strategies (Weill & Ross 2004). Traditionally, financial assets were administered using governance but in the last few decades, it has been expanded to include IT and data assets (Robert C. Rickards 2012). There are numerous standard IT governance frameworks established by the International Organisation for Standardization (ISO), the International Electromechanical Commission and the IT Governance Institute such as ISO/IEC 7799, Control Objectives for Information and Related Technology (COBIT) and the Information Technology Infrastructure Library (ITIL). While COBIT is considered the de facto framework for the governance of IT assets with a special focus on ensuring IT processes and activities align with corporate goals (Egelstaff & Wells 2013).

ITIL was developed to provide the best practice for the services that IT provides to its customers.

Table 2-2: Difference between IT and DG

IT Governance	DG
IT-driven	Business-Driven
Led by CIO	Led by Business Unit stewardship
Policy and process ensuring <ul style="list-style-type: none"> • effective evaluation • selection • prioritisation • funding of competing IT assets and investments. 	Policies, processes and practices that address <ul style="list-style-type: none"> • validity • accuracy • timeliness • data integrity • completeness data integrity
Overseeing implementation of IT policies and processes and extracting measurable business benefits	Operational focus

Adapted from (Dimick 2013)

Data Management, as stated earlier in Chapter 1, is a broad multidisciplinary topic that involves dealing with several applications, systems, and complex functions. DM supports cyber security, marketing, sales, corporate strategy, finance, compliance and audit. From an operational perspective, it supports business intelligence, reporting and analytics (Pearce 2017b) with DG ensuring data is well managed by an enterprise (Merkus 2015b). The management of data to generate value follows a general cycle starting from creation to disposal. Figure 2-5 below provides an example of the general data flow cycle.

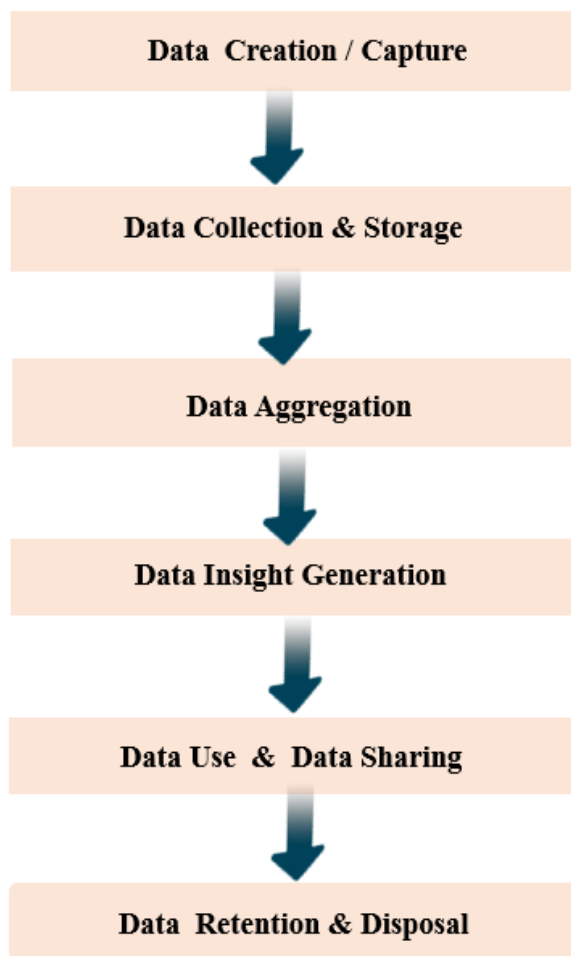


Figure 2-5: Data Lifecycle

The day-to-day execution of data governance policies are executed through DM. Thus, DG and DM are not mutually exclusive and complement each other to support organisations to reach their objective.

DG and DM formalise and accounts for almost every part of an organisation's data-related activities such as data replication, archival, security, backup, policies as well as individual technologies such as meta-data management, data lineage-business glossary mapping, governance council, change management and, master data.

The terminologies used in the thesis (adapted from Gill (2015); R. Minerva (2015); Rose, Eldridge & Chapin (2015) are shown in Table 2-3.

Table 2-3: DG and DM Terminologies

Term	Definition
Data	Facts and statistics collected together for reference or analysis.
Data Compliance	Act of adhering to, and demonstrating adherence to, external laws, regulations, corporate policies and procedures.
Data Dictionary	Centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format. It is a structure that stores metadata.
Data Steward	The data steward function is sometimes called the DG Board, the Data Stewardship Council, or the DG Office. The data stewards have business understanding to describe, establish, declare and enforce business rules around data in addition to ensuring adherence to policy.
Data Warehouse	The storage area for the purpose of reporting and analysis. The data warehouse consists of data that has its schema applied “On Write” as it is placed in well-structured databases for use in analytics.
Metadata/ System Catalog	Definitions, lineage (where does this data come from?), business definitions, technical metadata i.e., it is information that describes and provides information about other data.
Metadata Management	Metadata captures the attributes of data such as the type, length, timestamp, source, owner as well as relationships in data (semantics). Once defined and documented, these attributes can be used to search, link, aggregate and grant access to the associated dataset and helps with data traceability and lineage. The use of uniform methods and tools for defining, collecting, and managing information metadata ensures that data is identified consistently across the enterprise .
Data Quality Management	DQM refers to the planning, implementation and control activities that apply quality management techniques to measure, assess, improve and ensure the fitness of data for use (Mosley et al. 2010) .

2.2 Data Breaches

NIST defines a data breach as the unauthorized access, transfer, change, or deletion of sensitive or proprietary information. According to Cannon et al. (2007), data breach is the procurement or unauthorised access of data containing sensitive information in any format which compromises the confidentiality or security of such information and creates a reasonable risk of its misuse. A privacy breach also comes under the gamut of a data breach (Chatterjee & Sokol 2019). The impact of data breaches usually results in the financial and reputational decline of businesses. The below table lists recent data breaches resulting of either mishandling, misuse or mis-sharing of data by a DE player in multi-organisation environment.

Table 2-4: Data breach incidents

Event Involving	Event description	Event impact	Event cause
British Airways	Private information of 50000 users was exposed and harvested by cyber-criminals (Gromenko).	Compliance penalty	Incompetent governance to protect data
Global Transcription Services (GTS)	1600 medical letters were found in a Sydney bin by a sub-contractor for a company tasked with transcribing medical letters sent from specialists to general practitioners (Commissioner for Privacy and Data Protection 2016)	Compliance penalty	Insider threat caused by third-party vendor
Landmark White	Home loan details of customers hacked in major data breach (AFR 2019) (Herald 2020)	Loss of business with major banks like ANZ	Cybersecurity incident caused by exposed API accessed and misconfigured by third party
Flight Centre	Australian information and Privacy Commissioner confirmed that Flight Center disclosed 7000 customers passport and credit card information without consent to third party (Commissioner 2020)	Reputational Harm Financial loss	Consent not sought from data subject
Travelex	GDPR Possible fine due to non-reporting of data breach incident (Guardian 2020)	Financial Loss	Willful non-adherence to regulation on data breach disclosure mandated by regulators
Australian National University	Details of bank account, passport, tax stolen (Rashid & Ahmed 2020)	Privacy breach	Cybersecurity incident
Commonwealth Bank of Australia	In May 2016, two magnetic storage tapes containing 15 years of bank statements belonging to approximately 20 million customers disappeared (Tonkin 2019)	Reputational harm	Physical loss of digital data arising due to mismanagement by third-party ecosystem partner
HealthEngine	The Australian Competition and Consumer Commission (ACCC) alleged that from 30 April 2014 to 30 June 2018, HealthEngine gave information such as names, phone numbers, email addresses and date of birth of over 135,000 patients to private health insurance brokers for a fee without adequately disclosing to consumers it would do so (McGrath, Blumer & Carter 2018). Data retention and archiving (e.g., Cousins, 2016)	Unethical usage of data	Non-adherence to privacy laws arising due to sharing of data without a lawful basis to ecosystem partner

These breaches underline the necessity to have clarity around an ecosystem-oriented approach including the rules of engagement for data sharing. Prior studies indicate that poor governance and environmental and community policies increase the probability of a data breach, especially when the data is breached by hackers (Lending, Minnick & Schorno 2018).

2.3 Review of Frameworks

This research conducted an extensive literature review to identify the existing data governance-related frameworks that provide an organisation with an all-inclusive approach to collect, manage access to, secure, store and exchange data. First, this study applied the SLR guidelines which have been addressed in the paper by B. Kitchenham (2007)) for systematically searching, selecting, reviewing and synthesizing the DG-related framework from relevant academic and industry publications. This study included papers written in English, selected from five well-known electronic databases (Table 2-7). Further, the work on DG by leading industry vendors like SAS, Deloitte, Oracle as well similar studies from government organisations were included. The aim was to review to synthesize the literature of DG and DM from the DE perspective.

2.3.1 SLR Filtration Process

Based on our research aim, a search string was constructed using the Boolean “OR” and “AND” operator: (“data governance framework”) AND (“digital ecosystem”). A lack of results led to the refinement of the search criteria to include only “DG framework”. For the AIS electronic library database, a variant of the above string was constructed to ensure important studies were not omitted. The preliminary search resulted in a total of 88 hits across five chosen databases (see Table 2-6) with 80 of these being different. Figure 2-6 presents the three-stage selection procedure involving the identification, filtering and selection of a paper. This approach was taken to ensure that only studies relevant to the RQ were selected. Table 2-5 shows the filtration criteria used (e.g., keyword search in title, keyword search in abstract, exploration of paper contents).

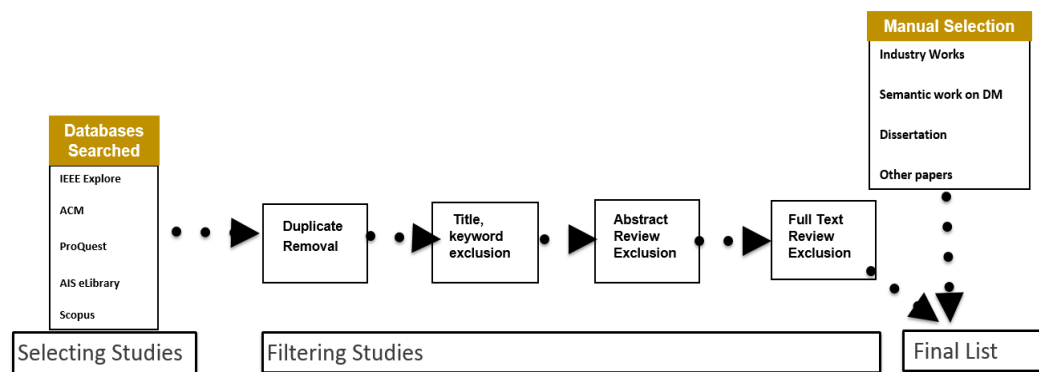


Figure 2-6: Three stage Systematic Literature Review filtration process

Table 2-5: Paper selection criteria

Filtration	Method	Assessment Criteria
Stage 1	Identify relevant studies from data sources	Keyword search published after 2016 Exclude dissertations Remove duplicates
Stage 2	Exclude studies based on titles	Title matches DG search term
Stage 3	Exclude studies based on abstract review	Abstract matches RQ on DG Remove duplicates
Final	Exclude studies based on full-text review	The article addresses the RQ

Table 2-6 gives a breakdown of the database-specific search results.

Table 2-6: Search results

Database	1st Filtration	2nd Filtration	3rd Filtration	Final Count
IEEE	5	4	4	4
ACM	8	1	1	1
Scopus	15	4	4	4
AISeL	23	3	3	3
Proquest	38	7	3	4
Others – Dissertation				5
Others – Government frameworks				3
Other- Practice-oriented frameworks (including DM frameworks)				10
Other – Academic papers obtained from existing SLR and latest papers				9
Total				43

16 unique results also included comprehensive SLRs conducted by Alhassan, Sammon & Daly (2018); (Ibrahim Alhassan 2016) and (Abraham, Schneider & vom Brocke

2019). These reviews examined the DG challenges and associated frameworks with the majority of these studies using a combination of terms such as (data governance) or (data governance organisation) or (governance data) or (data governance in cloud computing) or (data governance for cloud computing) or (cloud data governance or ("information governance")) as part of the search criteria and left out the practice-oriented literature on data breaches and compliance.

In addition to the above 16 studies, this study selected five dissertations and six additional academic papers by consulting the references listed in the aforementioned SLRs.

As mentioned in the previous sections, DG and DM are interdependent and complement each other, with DG setting the rules that are executed through DM. Hence, alongside the frameworks, the review exercise incorporated established literature on DM such as DCAM, ISACA and TDAN. In total, ten such studies from leading industry practitioners and the three frameworks used in Australian government departments were identified through a manual search in the search engines. The majority of the studies chosen contained all the aspects of the RQ in scope of this thesis. Finally, three recent studies conducted by researchers who have examined data or platform ecosystem based approaches for DG and DM were included in the review.

Table 2-7: Studies used for the Literature Review

<p>Generic (A1-A25)</p> <p>(Abraham, Schneider & vom Brocke 2019; Al-Badi, Tarhini & Khan 2018; Al-Ruithe & Benkhelifa 2017; Al-Ruithe, Benkhelifa & Hameed 2016a, 2018; Engels 2019; Heredia-Vizcaino & Nieto 2019; Kim & Cho 2017; Lee, Zhu & Jeffery 2019; Li et al. 2019; Lis & Otto 2020; Mahanti 2018; Nokkala, Salmela & Toivonen 2019; Otto & Jarke 2019; Paskaleva et al. 2017; Reis, Viterbo & Bernardini 2018; Thammaboosadee & Dumthanasarn 2018; Wimmer, Boneva & Giacomo 2018; Yang et al. 2019; Yebenes & Zorrilla 2019)</p> <p>(Curry & Sheth 2018; Perscheid, Ostern & Moormann 2020)</p> <p>(Ender 2021; Geisler et al. 2021; Jagals & Karger 2021)</p> <p>Thesis (D1-D5)</p> <p>(Brickson 2016; Cave 2017; Merkus 2015a; Paananen 2020; STEFANO 2016)</p>	<p>Industry Body. Trade association, Product Company, Managed Service Provider and Consulting Firms (P1-P10)</p> <p>(Council 2015; Deloitte ; Gartner 2017; Hitachi 2019; Infotech 2016; KPMG 2016; Mosley et al. 2010; PWC 2019; TDAN.com 2015; Wróbel, Komnata & Rudek 2017)</p> <p>.</p> <p>Government (G1-G3)</p> <p>(Government 2019; Norbib & Bakar 2021; Security 2018)</p>
--	--

A1-A25: Academic papers, D1-D5: Thesis in order of listing, G1-G2-Government papers, P1-P10: Professional papers (in order of listing)

2.3.2 Review and analysis of existing literature

The first serious work on the DG framework was published in 2007 (Niemi 2011). Since then, several studies have examined this topic. The studies can be broadly categorised into two groups: a) a DG framework in a single organisation context, and b) a DG framework in an ecosystem context. The majority of the studies belonged to the first category (Nokkala, Salmela & Toivonen 2019) with a limited number of studies considered data sharing in the DE context.

In their big data governance framework, (Al-Badi, Tarhini & Khan 2018) identified eight core principles :1) stakeholder selection, 2) big data scope determination, 3) policies and standard settings, 4) measure and monitor quality, 5) data storage, 6) communication and data management, 7) optimize and compute and 8) identify organisational structure.

Aspects of DG related to cloud governance was proposed by Al-Ruithe, Benkhelifa & Hameed (2018) in their cloud framework for public sector.

Wimmer, Boneva & Giacomo (2018) focused on interoperability among different governance functions that can assist in promoting interoperability among e-Government developments.

In their study, the conceptual framework of DG in a sustainable urban data smart city was proposed by the authors (Paskaleva et al. 2017). It comprised of six pillars: 1) project context, 2) data collection, 3) data generation, 4) data identification, 5) data use and legacy and 6) data sharing and management.

Otto & Jarke (2019) investigated the use case of an alliance driven platform and noted that working in such ecosystem scenarios requires co-ordination involving task-forces and working groups, and do not follow traditional approach.

Abraham, Schneider & vom Brocke (2019) expressed concerns over trust, risk and security and suggested a conceptual framework with antecedents as part of the DG framework.

Young et al. (2019) suggested considering legal-technical framework taking data trust under consideration when sharing data among ecosystem players. Mahanti (2018) identified the key critical success factors of Data Governance implementation framework from a single enterprise context.

A healthcare specific framework comprising of 3 domains and 12 components were proposed by the authors in Li et al. (2019). Lee, Zhu & Jeffery (2019) identified data ownership, access and usage as key decision domains influencing DG. The use of insurance to minimize data security breach cost and include breaches caused by suppliers was advocated by Franke (2017). The importance of evidence-based policy making was emphasised in the framework proposed by Parycek & Pereira (2017). From the open data perspective, Thammaboosadee & Dumthanasarn (2018) suggested changes to Thailand's Public Information Act to move towards a DG framework.

In the context of compliance, the authors recommended identifying cyber risk and data breaches as part of the compliance function (Chatterjee & Sokol 2019). In one of the early papers from 2007, Daring et al. (2007) analysed and characterised digital ecosystem governance into licensing and regulation, technology, organisation, communication, trust and balance of interests

Recently in 2019, the United Nations Educational, Scientific and Cultural Organisation (UNESCO) recommended the role of judicial operators and the rule of law in the DE should be taken into account to enable an environment for DG that is aligned with international human rights standards.

The industry associations addressed their concerns over data sharing with updated versions of DAMABOK, DCAM, ISACA and TDAN. For example, ISACA (Pearce 2017a) suggests that DG and Enterprise Data Management are critical for the future sustainability of enterprises. In DAMABOK published by the Data Management Association, practitioners identified 9 core management principles with data at the core. On the other hand, the data administrator network association's (TDAN) Non-Invasive DGTM approach comprised six key components, namely data, roles with formal accountability, processes, communications, metrics and tools. The DG Institute (DGI) provided in-depth, vendor-neutral DG best practice and guidance. COBIT 5 viewed data governance as supervising information custodianship and security.

The software engineering, product development, consulting and service provider firms have devised new frameworks aimed at addressing the impact of new requirements derived from the use of a cloud, the highly distributed nature of IoT data, data-centered architecture and big data technologies. These utilize a combination of governance mechanisms but focus on aligning the vendor solution offerings that help in data control activities. The DG framework by SAS (2019) has the following five pillars: 1) corporate drivers channels 2) DG, 3) solutions, 4) data management engagement 5) assets and the Internet of Things. The Deloitte model is based on four key segments: a) processes, policies, standards, and procedures b) organisation, roles, and responsibilities c) technology and tool capabilities and d) metadata content (or data catalog). KPMG's DG Framework (KPMG 2016) is similar consisting of people and organisations, governance, tools and processes and controls. PriceWaterHouse Coopers has data usage governance which is comparable to the existing framework from industry. The investigation into frameworks developed by government agencies

such as the Workplace Gender Equality Agency (Agency) of the Australian Government revealed that it was primarily based on industry association frameworks like DAMA.

The SLR results are summarised and interpreted to provide useful insights on existing frameworks. In terms of the constituents of the existing solutions, it was interesting to find that there was an overlap of elements that are essential to support DG and DM, regardless of whether they focussed on inter and intra organisational. Table 2-8 presents the identified major constituents related to DG and DM in the existing literature: (1) technology, (2) policy, guidelines and procedures, (3) processes, (4) structure and roles, (5) risk management and (6) data.

While these generic elements made up the building blocks of several frameworks at a high level, there were varied levels of coverage of the different elements at the detail level. The consideration of the DE context was included in few of the frameworks as indicated in the last column of Table 2-8. For example, a strategy for DG in the cloud was discussed in A3.

Table 2-8: Categorisation of Frameworks into Elements captured

	Technology	Guideline	Process	Roles	Risk Management	DE perspective discussed
Study						
G1	Y	Y	Y	Y	Y	N
G2	Y	Y	Y	Y	Y	N
G3	Y	Y	Y	Y	Y	Briefly discusses external agencies
P1	Y	Y	Y	Y	Y	Y
P2	Y	Y	Y	Y	Y	N
P3	Y	Y	Y	Y	N	Y.
P4	Y	Y	Y	Y	Y	N
P5	Y	Y		Y	Y	N
P6	Y	Y	Y	Y	Y	N
P7	Y	Y	Y	Y	Y	N
P8	Y	Y	Y	Y	Y	N
P9	Y	Y	Y	Y	Y	Y. IoT focus
P10	Y	Y	Y	Y	N	N
D1	Y	Y	Y	Y	Y	N
D2	Y	Y	Y	Y	Y	N
D3	Y	Y	Y	Y	Y	N

D4	Y	Y	Y	Y	Y	N
A1	Y	Y	Y	Y	Y	Y
A2	Y	Y	Y	Y	Y	N. Big Data specific
A3	Y	Y	Y	Y	Y	Y. Cloud Specific
A4	Y	Y	Y	Y	Y	Y. Cloud Specific
A5		Y			Y	Y. Cloud Specific
A6	Y	Y	Y	Y	Y	N
A7	Y	Y	Y	Y	Y	N
A8	Y	Y	Y	Y	Y	N
A9	Y	Y	Y	Y	Y	Yes, Focus on platforms
A10	Y	Y	Y	Y	Y	Y. Healthcare specific cross institutional data exchange
A11	Y	Y	Y	Y	N	Focus on Data ecosystem taxonomy
A12						N Focus on defining key success factors of DG
A13						N.
A14	Y	Y	Y	Y	N	Y. Focus of Platform Ecosystem designing, design Phase
A15	Y	Y	Y	Y	Y	Y, Discuss briefly in IoT context
A16	Y	Y	N	Y	N	N. Briefly discusses data licensing when sharing data
A17	Y	Y	Y	Y	Y	N
A18						N, Focus on interoperability governance
A19	Y	Y	Y	Y	Y	N
A20	Y	Y	Y	Y	Y	Y-Focus on platforms
A21	Y					Y-Discuss need of DG in data ecosystem

A22						Focus on Platform incentivisation, trust accessibility
A23	Y			Y	Y	Y-Focus on platforms DG taxonomy
A24	Y			Y	Y	Y-Discuss data ecosystem taxonomy
A25	Y	Y	Y	Y	Y	Y.

Actors, their roles, responsibilities and accountabilities were a common element found in all the frameworks. What is interesting to note are that policies were another common category found in most of the frameworks. A high percentage of studies indicated the elevated importance organisations attach to policies, though what policies need to be incorporated varied between frameworks irrespective of whether it was academic or practice-oriented literature. Processes aligned to policies were the next most recognized element of the frameworks. Despite most researchers discussing the need to control and monitor data flow and services implemented by the ecosystem service providers, the level of granularity with regards to what processes should be implemented was low. Technology, comprising tools and systems involving computation and architecture was another area which emphasised by the studies

With the emergence of DE, data is considered as critical organisational asset and needs to be governed and managed appropriately. Participating actors in DE derives benefit by integrating data, but the interoperability challenges exist. The risks affecting any one ecosystem actor can potentially affect other actors in DE. In other words, in DE, there may be a single organisation accountability, but the responsibility lies with everyone in the organisation along with the partners involved. This means that DE requires mass collaboration and co-operation to meet business demands. This SLR

studied a number of papers that provided insights into the suitability of the existing solutions to manage and govern data in DE.

The studies reviewed found that several frameworks exist for governing and managing data. However, majority of the existing practices and frameworks are focussed on single organisation. Ecosystem-oriented research is not a widely studied topic with a limited number of papers considering ecosystem (Nokkala, Salmela & Toivonen 2019) perspective. As evident from Table 2-8 (refer to column “DE perspective discussed”), very few publications investigated the ecosystem aspect. The frameworks that specifically discussed DE were scarce and mostly restricted to specific principles such as bilateral agreements, incentivising partners, technology (platforms or cloud), data type (IoT or BigData) or industry (health). For example, Lee, Zhu & Jeffery (2019) discussed practical ways to implement their framework through platform based case studies. However, negligible information was provided on the details of operationalising the framework and the focus was primarily on ownership of data. Similarly, the framework proposed by Abraham, Schneider & vom Brocke (2019) was conceptual in nature and did not validate the practical applicability of the framework. The authors noted that no consensus exists on how DG should work in multi organisational context. This view was supported by Ender (2021). In their study, the authors pointed out that governing and managing data exchanged between multiple actors are fundamentally different from traditional DG practices. While the existing intra-institutional hierarchical roles within a focal firm in DE are still relevant, the co-ordination between firms warrants considerations given to stakeholder engagement models, data interoperability standardisations and data retention needs.

The notion of privacy, misuse, mistrust and confidentiality has changed in DE where multiple organisations participate, produce, share and consume data throughout data lifecycle. The analysed papers acknowledged the growing importance of privacy and security in this research field. Most practitioner-oriented frameworks discussed the topics related to confidentiality, compliance, protection against threats under risk management category. As previously stated, proper data handling is fast becoming an enforcement focus for regulators all across the globe. The obligations that a firm along with its ecosystem partners needs to comply are growing day by day with newer regulations being enforced. Therefore, understanding the changing regulatory landscape is critical component of DG and DM and requires further research (Lis & Otto 2021). A noticeable observation was that though legal instruments were mentioned in the recent frameworks, the precise nature of the mandates around data protection were either absent or addressed insufficiently in the literature. This is particularly important, since regulators attempt to reduce the trust gap by continuing to create new laws to address data privacy and security concerns. The laws are clear about what the end goal is, however it is a very complex process. The regulators do not offer any framework to guide IT practitioners on how to process the data that can result in data rule violation. From the analysis of the key requirements of popular legislations, it can be concluded that very few frameworks mapped the key needs into their DGM frameworks. Furthermore, correlation between DG activities performed and data breach incidents, particularly the causes of data breaches arising due to ecosystem partners' accidental data leaks, was not explored in details. The increase in the number of data breaches with each passing day are an issue of serious concern and raises questions about the effectiveness of the existing DG and DM techniques.

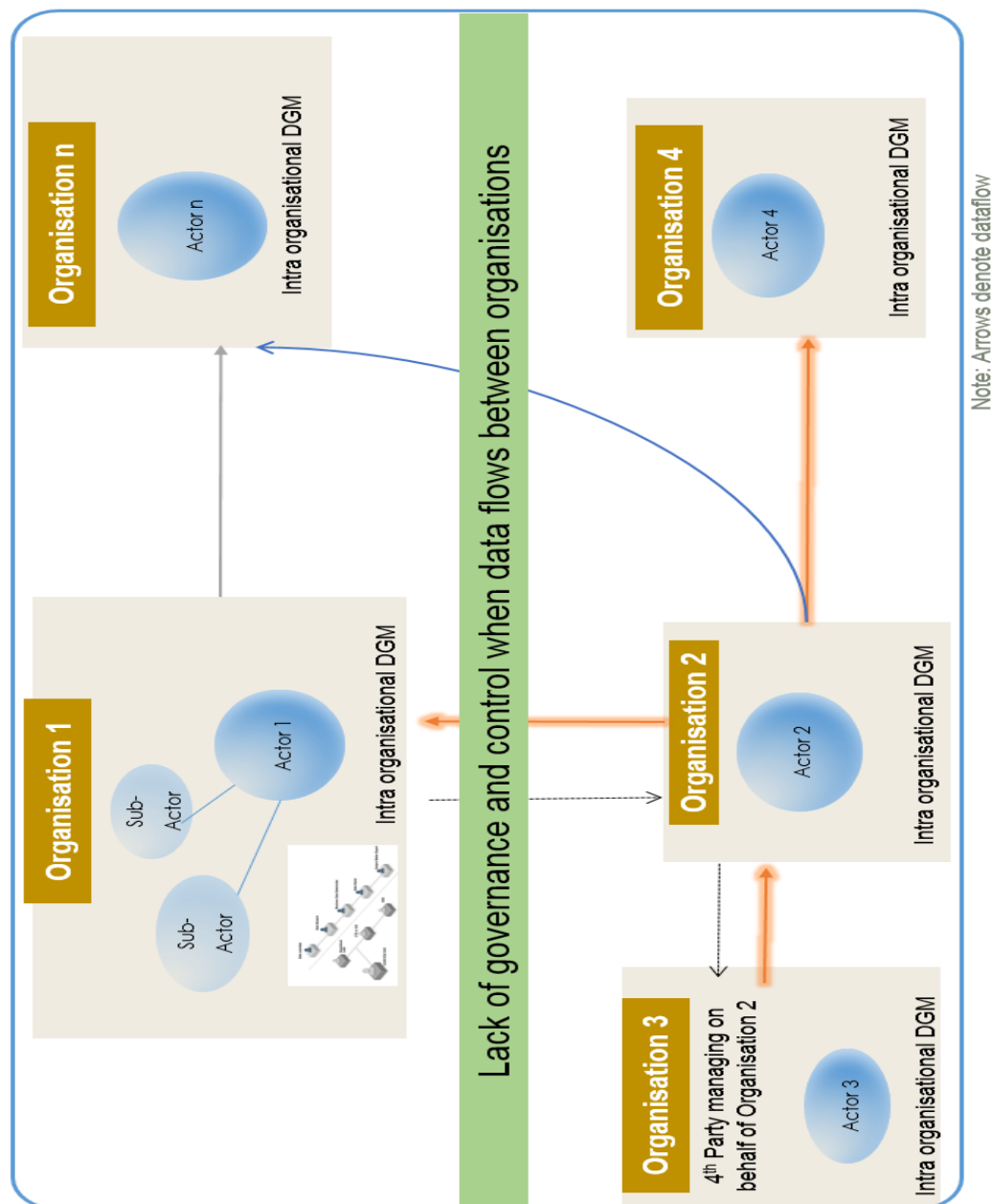


Figure 2-7: Data sharing among ecosystem partners in DE

Another observation that though there were numerous suggestions to improve DG and DM practices, irrespective of inter or intra context, the level of abstraction was very high. The existing studies failed to provide practical guidance on the concrete use of a framework to tackle practical business requirements. Furthermore, most of the papers

studied emphasised defining the principles to govern and manage data, with only a few of the selected papers highlighting the implementation and monitoring challenges involving distributed data (Nielsen 2017).

Therefore, there is an urgent need to have a solution in the form of a framework that can provide guidance on regulatory needs, privacy, security and data sharing engagement models in DE .

Table 2-9: Gaps

Gaps in existing literature	Relevant Paper
Emerging field with limited research available from DE perspective	Ender 2021
Existing frameworks focus on defining areas of governance with a lack of researched implementation approach	(Lis & Otto 2020) (Jagals & Karger 2021) (Nielsen 2017)
Existing Frameworks lack engagement model in ecosystem context	(Ender 2021; Lee, Zhu & Jeffery 2019; Nielsen 2017; Society 2017)
Limited understanding of regulatory needs	(Curry & Sheth 2018) (Al-Ruithe, Benkhelifa & Hameed 2018)

This dissertation aims to fill the gaps in the literature by prescribing through a framework (see Chapter 4) the activities necessary to govern and manage data in a consistent manner to address DE.

2.4 Research gaps and question

The research background and literature review conducted in this study identified challenges in governing and managing data in DE. The review indicated that the existing frameworks overlook inter-organisational complexities. The research gaps (RGs) and research questions based on the analysis of the existing literature are summarised in Figure 2-8.

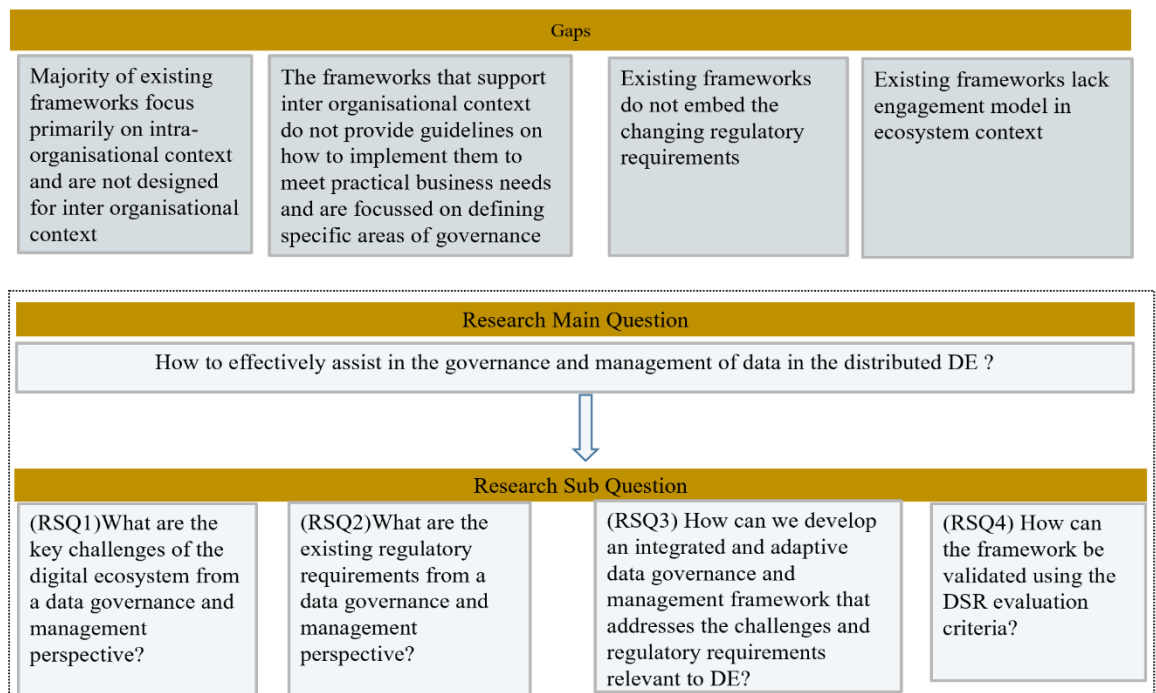


Figure 2-8: Research Gap and Questions

The key gaps identified are as follows:

RG1: Majority of existing frameworks focus primarily on intra- organisational context and are not designed for inter-organisational context.

RG2: The frameworks that support inter-organisational context do not provide guidance on how to implement them to meet practical business needs, and are focussed on defining specific challenges and areas of governance.

RG3: Existing frameworks do not embed the changing regulatory requirements needs.

RG4: No framework exists that considers engagement models between ecosystem actors in DE.

The RG's concerning the governance and management of data in DE using existing practices, resulted in the formulation of the following RQ.

RQ) "How to effectively assist in the governance and management of data in the distributed digital ecosystem?"

To tackle the broad RQ question, it was broken down into the following RSQs .

Research Sub-Question 1 (RSQ1): *What are the key challenges of the digital ecosystem from a data governance and management perspective?*

This study reviewed existing literature on DG, DM and data breach incidents to highlight the issues faced by organisations participating in DE to govern and manage data and suitability of existing mechanisms to meet ecosystem-centric industry demands.

Research Sub-Question 2 (RSQ2): *What are the existing regulatory requirements from a data governance and management perspective?*

The study reviewed existing data related regulations (Section 2.5, Section 5.2.1) such as GDPR and APP to list the key requirements any organisation in DE needs to take into account.

Research Sub-Question 3 (RSQ3): *How can we develop an integrated and adaptive*

data governance and management framework that addresses the challenges and regulatory requirements relevant to digital ecosystems?

The findings of SLR conducted (RSQ1) suggested the development of a practical framework to address the requirements of DE. The existing solutions provide deep insights to understand the dimensions that are necessary to assist in DE-oriented DG and DM. SLR indicated that the existing studies could be drawn upon to provide guidance (Al-Ruithe, Benkhelifa & Hameed 2016b) to counter the shortcomings. Thus, SLR results provide the input for the DSR methods “awareness of problem” and “suggestion” stages to develop the 4I Framework (see Chapter 3). The study used the findings from the RSQ1 and RSQ2 to develop the initial version of the framework. The main aim of RSQ3 is to fill the gaps in the literature by prescribing through the 4I Framework (see Chapter 4) the activities necessary to consistently govern and manage data. The proposed framework considers regulations and risk-based structured approaches to govern and manage data in DE.

Research Sub-Question 4 (RSQ4): *How can the framework be validated using the DSR evaluation criteria?*

The proposed solution is validated using survey and scenario-based examples (see Chapter 5) using the evaluation criteria outlined in Chapter 3.

2.5 Data Regulations

Regulations related to data has grown since the 2000s. However, the last three years have witnessed an increase in the creation and amendment of data related regulations. One of the main reasons for that is the sharp increase in the number of incidents arising from data breaches (see Section 2.2).

In the past, data transfer between organisations were a diadic relation, but in the DE dataflows in multidirection. For example, data related to a home loan application can be uploaded in cloud hosted in USA based server. The data can be enriched with the help of credit score data supplied by another Australian company before being sold to another third party organisation. to send target real-estate furnishing products to sell to the customer. Thus, to manage multiple systems and applications, it is critical to understand how they will interact to achieve the business goal. For that reason, new regulations provide the requirements that needs to be considered through which this use or misuse of information can be controlled.

Since the beginning, complying with regulations has been a challenging task due to the prescriptive nature of the laws. DE inherits the existing problems faced when complying with regulations.

The most prominent legislative requirements are HIPAA (Health Insurance Portability and Accountability Act) for healthcare, GDPR (The General Data Protection Regulation) for EU citizens, MA Risk (Mindestanforderungen an das Risikomanagement) for supply chain management and Australian Privacy Protection (APP) law for Australian Businesses. The compliance requirements varies from country to country and are industry-domain driven. It should be noted that when it comes to the collection, use and sharing of PI to external third parties without notice

or consent of consumers, legislations are consistent across jurisdictions with 128 out of 194 countries having a law in place for data protection and privacy (Development 2021).

Data-related regulations can be classified into overlapping categories of data security, data rights and privacy, responsible data handling and data protection.

An intrinsic part of data laws are data privacy issues. Digital privacy is defined as the attention that individuals have in sustaining a personal space, free from interference by other people and organisations (Clarke 1999). Sensitive personal identifiable information is the core of data privacy. Personal Identifiable Information (PII) includes details such as Title, First name, Surname, Date of Birth, Residential address and Phone number. Financial, health, and the geophysical location of an IoT user is also considered as sensitive personal information (SPI).

In Europe, the General Data Protection Regulation (GDPR) is a regulation constituted by the EU parliament, Council and European Commission. It is one of the most stringent data laws with the primary purpose to safeguard the personal data rights of citizens and residents of the European Union in this era of new technological advancements (Chaudhuri 2016). GDPR directives aim to streamline businesses within the EU from a regulatory perspective. IoT, whose value lies in the data coming from the sensors, comes under the purview of these laws. The data can result in the identification of a person who directly violates the GDPR law of processing data without consent (Mansfield-Devine 2016). There is general agreement among industry experts that the GDPR, approved by the European Union parliament, Council and European Commission is one of the most rigid regulations. GDPR has been in effect from May 25, 2018 and it mandates organisations to develop and implement

technological and organisational procedures to ensure compliance to legal needs. Regulators can penalise organisations for GDPR non-compliance by up to 2-4% of global income or 20 million Euros (Commission 2016). GDPR applies to any company that offers goods and services to European citizens and handle their data, regardless of their geographic location. The accountability of protecting an EU citizen's personal data falls on the shoulders of organisations. GDPR aims to ensure that individuals have better control over usage of data involving the individual. In a recent study Gartner (2018) reported that as of 25th May 2018, less than 20% of worldwide organisations were compliant with EU's GDPR. According to this report, at a minimum, 20% of organisations which fall under the jurisdiction of GDPR will be accused of non-compliance by 2021. Moreover, regulations such as GDPR affect and influence the governance of organisations, not only those in the EU but also transborder. In Australia, a number of regulations related to data, data protection, database protection, privacy and security laws exist as shown in Table 2-10.

Table 2-10: Data-related legislation in Australia

Legislations
Mandatory Data Breach Notifications (MDBN)
Do Not Call Register Act 2006 (Cth)
Spam Act 2003
Information Privacy Act 2014 (ACT)
Privacy and Personal Information Protection Act 1998 (NSW)
Information Privacy Act 2009 (Qld)
Privacy and Data Protection Act 2014 (Vic)
Personal Information and Protection Act 2004 (Tas),
My Health Records Act 2012 (Cth)
ACCC Consumer Data Right (CDR)
Australian Privacy Principles (APP)
Evidence Act 1995
Electronic Transactions Act 1999
Electronic Transactions Regulation 2000
Digital transition and enterprise content management (ECM) policy

In addition to legislations, guidance to handle data is provided such as the Australian Prudential Regulation Authority CPG 235 Managing Data Risk for the financial sector that considers data risk as a sub-risk of operational risk (APRA 2013). Operational risk is formally defined as the risk that causes the decline or breakdown of services and results from the deficiencies in IS, internal processes and workforces, or disruptions from external events. In the USA, California Senate Bill 327 (California 2018) was announced in 2018. The goal was to empower the State of California the right to file law enforcement complaints against companies that did not implement adequate security protections for the IoT devices they manufacture. The bill gives the California state government the right to hold manufacturers of IoT devices accountable for protecting consumers' data. The IoT Cybersecurity Improvement Act of 2017 requires: (i) that IoT devices are patchable, (ii) that devices do not have known vulnerabilities, (iii) that devices rely on standard protocols, (iv) that devices do not contain hard-coded passwords and (v) that certain technical aspects of privacy are upheld in the IoT era. Reports from government bodies such as the US Federal Trade Commission (FTC) and EU (Michael S. Smith 2015) acknowledge the risks associated with the realm of privacy and security.

There is no doubt regulations are one of the most critical element to consider in DE. Formal regulations provide rules that must be followed by participating organisations of the DE. The list of regulations are extensive (Aljeraisy et al. 2021) and beyond the scope of this thesis. Table 2-11 lists maps the requirements from few well-known regulations that impacts data, into activities that needs to be considered engaging in data exchange. The list is confined only a few legislations and can be extended to other regulations.

Table 2-11: Regulations mapped to activities

Regulation	Data Subject Rights	Data Handling rights	Governance	Data Security
GDPR	<p>Transparency around data collection and processing</p> <p>Right of access</p> <p>Right to data correction</p> <p>Right to data erasure and data retention</p> <p>Right not to be subject to profiling</p> <p>Data portability</p>	<p>Purpose limitation</p> <p>Data minimization</p> <p>Data quality</p> <p>Legal basis of data processing</p> <p>Sensitive data classification</p> <p>Third-party relationships</p> <p>International data transfers (cross-border)</p>	<p>Appointment of data protection officer (DPO)</p> <p>Written records of processing activities</p>	<p>Security of processing through data controls</p> <p>Breach notification to authority</p> <p>Breach notification to data subject</p> <p>Data protection by default design</p>
APP	<p>Right to object to data collection</p> <p>Right to data access</p> <p>Right to data correction</p> <p>Right to anonymity</p> <p>Right to object to marketing</p> <p>Right to notification of data collection</p>	<p>Unsolicited collection handling</p> <p>Quality</p> <p>Cross-border disclosure</p> <p>Transparent management</p> <p>Purpose limitation</p> <p>Use, adoption or disclosure of government related identifiers</p> <p>Data minimization</p> <p>Retention</p>	<p>Promote good privacy governance within agencies by treating personal information as a valuable asset</p>	<p>Security</p>
ACCC Consumer Data Right (CDR)	<p>Consumer consent</p> <p>Safe and controlled use of data</p>			<p>Data protection and privacy</p>

CCPA	<p>Right to request to have data deleted</p> <p>Right to stop the sale of their information.</p> <p>Right to sue in case of data breach</p>	Data mapping	Due diligence to check data obtained from third parties are from legitimate source.	
APRA CPG235		<p>Data lifecycle management</p> <p>Implementing controls and validations</p>	<p>Staff awareness</p> <p>Data risk assurance</p>	

2.6 Chapter Summary

In this chapter, we first went through the basics of the DE, DG and DM. We then conducted a systematic literature review of the existing DG frameworks and DM literature and discussed the results including the operating models, common elements as well as shortcomings of these frameworks from a DE perspective. The results revealed the gaps that exist in the literature and resulted in the formulation of the research questions. The reviews suggested the development of a framework that addressed the gaps. This chapter also highlighted the requirements from the regulators that influence DG and DM in multi-party environment. The results of the SLR conducted in this chapter addressed the RSQ1 (challenges), RSQ2 (regulatory requirements) and provided the foundational information to develop the 4I Framework.

In the following chapter, a few popular research methods used in IS research are discussed before the best-fit method for this study is chosen and discussed in detail.

Chapter 3: Research Method

This chapter presents the research method adopted by this study to seek answers to the research questions and achieve the research objectives, as stated in Chapter 1. It describes how the study was conceived and conducted to fulfill the research objectives of the dissertation. The study examined different research approaches in information systems (IS) at the beginning of this chapter. Given the nature of the research, this study develops the research design by adopting the design science research (DSR) methodology guidelines proposed by Peffers et al. (2007). This chapter presents how the iterative DSR is used in performing this research across the phases such as awareness of the problem, suggestions and developments, evaluation, data collection and analysis and conclusion and communication.

3.1 Research in Information Systems

A research method is a technique or procedure that can be used to gather research data and analyse it against a research hypothesis (Crotty, 1998). Methods commonly used are sampling, measurement and scaling, questionnaires, non/participant observations, interviews, focus groups, case studies, life history, narratives, visual ethnographic methods, statistical analysis, data reduction, theme identification, comparative analysis, cognitive mapping, interpretative methods, document analysis, content analysis and conversation analysis (Adida et al. 1998). In information systems research, a strong emphasis is given to research methods and paradigms. The choice of research method, whether qualitative and quantitative, influences the way in which the researcher collects data to achieve the research outcomes. A plethora of diverse research approaches are employed in IS (Niglas 2001). Notable IS research approaches are discussed in the following.

3.1.1 Case Study

Case study research is the most common qualitative method used in information systems. Case studies investigate a phenomenon in its natural environment by employing multiple approaches of data collection to gather information from one or a few entities, particularly when the boundaries between phenomenon and context are not clearly evident.

3.1.2 Descriptive

Descriptive research relies on observation as a means of collecting data as against examining records or artefacts. It attempts to examine situations in order to establish

what is the normal and what can be predicted to happen again under the same circumstances.

3.1.3 Action Research

Action research (AR) aims to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework (Rapoport 1970). According to this definition, AR has the twofold goal of contributing to practice and research at the same time. The definition furthermore assumes that there is a concrete client involved. Therefore, AR is highly context-dependent while attempting to address the specific client's concerns.

3.1.4 Design Research

The design Science approach in information system (IS) research has been widely adopted (Gregor and Jones 2007; Hevner et al. 2004; March and Smith 1995). It is problem focused and seeks to design an innovative product or artefact that addresses unsolved problems within an organisation. DR combines both qualitative and quantitative methods and is established as an important and legitimate IS research paradigm.

3.1.5 Action Design Research

The action design research (ADR) method combines action research (AR) and DSR. ADR is considered an appropriate research method to design, develop and evaluate artefacts that aim to address a practical phenomenon in a practical context.

3.2 Choice of Research Method for this study

The selection of a research approach depends on several factors, including the nature of the research problem and the RQ in hand. The landscape of this research is discovering problems and creation of -solutions to address the problems. It is about finding the concepts, relationships and processes that help to design and build a framework to manage and govern data in the DE. For this reason, this research uses the iteration-based design science research (DSR) approach (see Figure 3-1) which fits well with the ‘build and evaluate’ cycle of this information systems-related research (Niu, Lu & Zhang 2009). DSR provides the essential steps (build and evaluate) that support the creative designs to solve problems in the application environment and to grow the application knowledge bases.

In this study, the planned primary outcome is the framework artefact. The framework is not a tool in itself, but a step-by-step procedure that can serve as a guide. The overall objective of this artefact is to develop and evaluate productive approaches to govern and manage data which are the key to a successful collaborative data-compliant ecosystem.

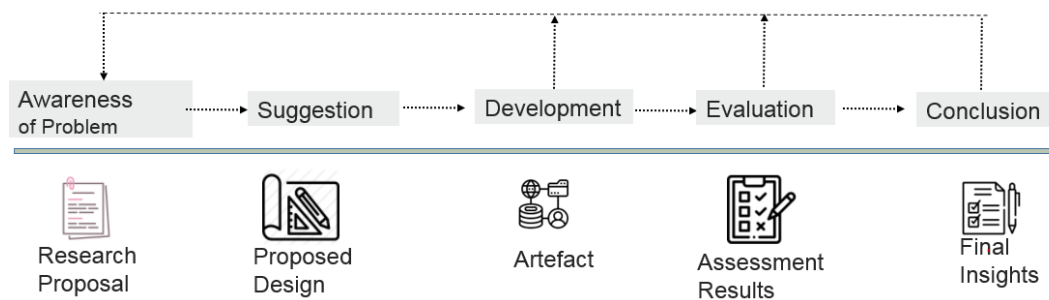


Figure 3-1: Design Science Research process

As illustrated in Figure 3-1, the first action of the DSR is problem identification in which the motive behind the research is formulated. In this stage, the research problem and the importance and relevance of the research is substantiated at the same time ascertaining the value and capability of the proposed solution. The awareness of the problem is attained by conducting a thorough review of the literature. The second stage of the DSR method seeks to find an answer to the question: “How to solve the problem?” In order to suggest a feasible solution, the researchers need to gain a deep and up-to-date understanding and knowledge of the emerging technologies (Geerts 2011). The third stage is the development stage. It is concerned with the design, development and implementation of the artefact. The effective execution of this activity results in the creation of viable artefacts that support problem solving as well as demonstrates how the designed artefact works through a prototype implementation. Ostrowski & Helfert (2012) stated that this stage can include simulation, case studies, experimentation, proof or other appropriate activities. The last stage in DSR is evaluation. The perceived effectiveness of the artefact in solving the identified problem needs to be observed and measured against the criteria set in the second stage of the DSR. Effective evaluation requires knowledge of the relevant metrics and analysis techniques. The final activity of the DSRM enables researchers to communicate the research problem and its importance, the artefact as the proposed solution and the efficacy, novelty, and clarity of the solution to other researchers in the field as well as interested parties. The communication in the DSRM can be in the form of a dissertation, journal article, research report, or conference presentation (Offermann et al. 2009).

3.3 Use of DSR in this Thesis

The DSR process (adapted from (Hevner & Chatterjee 2010; Kuechler & Vaishnavi 2008; Peffers et al. 2007)) comprises of several steps with each step comprising of one or more activities and corresponding outcomes. Table 3-1 gives an overview of how each step in the DSR methodology is covered by this thesis.

Table 3-1: Research Methods used in this research

DSR Step	Activity	Outcome
1. Awareness of Problem	The DSR uses literature review to identify what is known about the research topic and identify the key research gaps.	Research Question Research Objectives (Chapter 1, 2)
2. Suggestion & Development	The DSR uses a literature review, explanatory proof-of-principle prototyping and industry expert feedback using surveys to develop the framework.	4I Framework (Chapter 2,4,5)
3. Evaluation	This DSR step includes an evaluation method to evaluate the proposed framework and determine if the framework is fit for its purpose in relation to the research question. The evaluation step includes: <ul style="list-style-type: none"> Engagement of experts from industry to examine the artefacts using the survey Illustrative scenario-based case-studies. The results and the feedback from the participants are constantly incorporated in the proposed framework.	Evaluation using survey Explanatory scenario-based validation (Chapter 5)
4. Conclusion	This DSR step concludes whether the artefact developed presents any contributions. It outlines the outputs, discusses the proposed 4I Framework limitations and possible future research directions.	Thesis Results Publications

Steps from 2 and 3 are done iteratively. The research timeline is show in Figure 3-2.

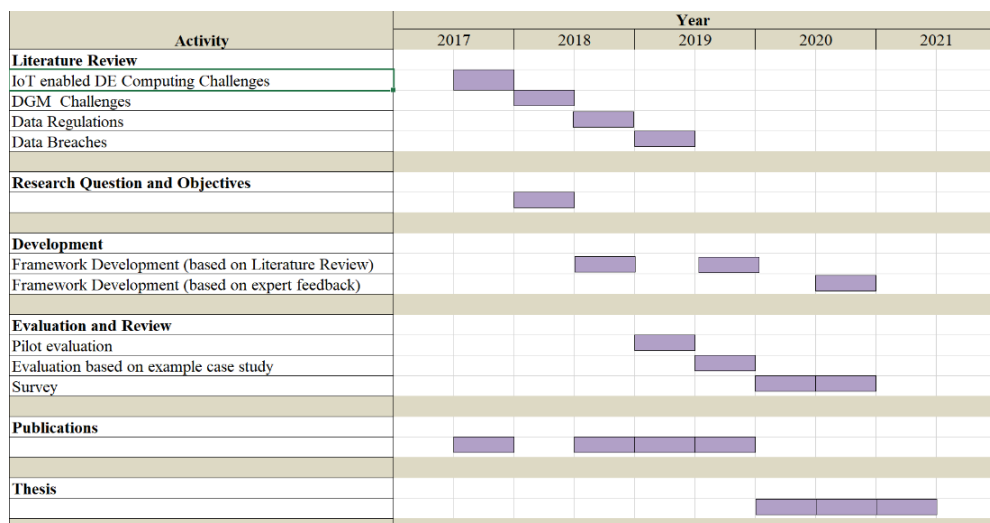


Figure 3-2: Research Timelines

3.3.1 Awareness of the Problem

Researchers have typically approached DSR from one of the four following entry points (Peffer, Tuunanen & Niehaves 2018):

- a) Problem-centred initiation occurs when stakeholders encounter problems that need to be analysed prior to resolving them. The main activities in problem-driven investigation includes describing problems, formulating and testing hypotheses about their causes and prioritising the problems to be solved.
- b) Objective-centred initiation is triggered by an industry or research requirement need that can be addressed by developing an artefact. It involves describing and prioritising stakeholder goals and operationalising them.
- c) Design and development-centred initiation where technology is in search of problems that can be solved. Problem investigation in this case consists of making an inventory of goals and of current technology, and in identifying functionality and performance requirements for new technology.
- d) Client content initiation, which focusses on solving a specific research problem for a specific client.

For this study, the entry point was problem centred initiation. Chapter 1 and 2 discussed the “awareness of the problem”. The literature review assisted in identifying the existing problems, formulating RQ and defining RO. The main RQ formulated was *‘How to effectively assist in the governance and management of data in the distributed DE?’*.

ROs defined were as follows:

- a) Research Objective 1 (RO1): *Conduct a literature review of the existing data governance and management frameworks used and assess the associated challenges from a DE perspective*
- b) Research Objective 2 (RO2): *Conduct a literature review of the existing data-related regulations*
- c) Research Objective 3 (RO3): *Develop a framework based on regulations and industry standards to assist in governance and management of data in DE*
- d) Research Objective 4 (RO4): *Evaluate the proposed framework developed in RO3.*

3.3.2 Suggestion and Development

The consolidation and synthesis of the literature on existing practices in Chapter 2 suggested the creation of a solution in the form of DE-oriented framework to address the research gaps. The analysis of the scattered literature on existing practices in the previous DSR stage provided guidance to fulfil the industry demands of DE. Thus, development of the 4I Framework started with the inputs provided in Chapter 1 and 2. The framework (see Chapter 4) helped in meeting RO3. The final version of the framework evolved based on industry feedback.

In recent years, ontologies, a kind of agreement on a domain representation, has been applied in information systems (Wand & Weber 2004). It is foundational in representing IS and provides a vocabulary to describe properties, classes and attributes of the components of a conceptual domain under consideration (See Figure 3-2). In addition, the relationships among the different elements involved are also described in

ontology (Viinikkala 2004) and implicit new knowledge can be inferred through logical reasoners or inference engines.

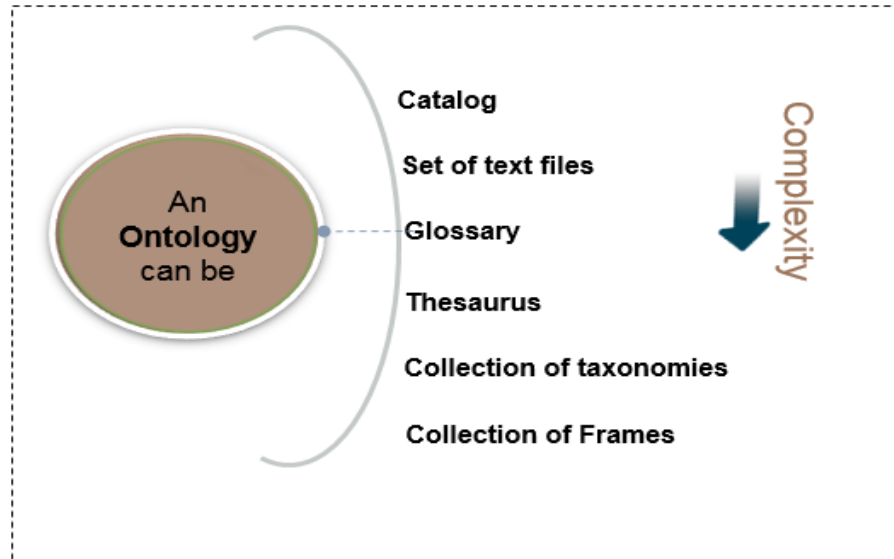


Figure 3-3: Interpretations of IS ontology

There are several languages available to construct ontologies such as RDF (Resource Description Framework), RDFS (RDF Schema), KIF (Knowledge Interchange Format) and OWL (Web Ontology Language) (Kalibatiene & Vasilecas 2011). In addition, UML (unified modelling language) has long been used as an ontology modelling language since it provides a mainstream commercial mechanism (Cranefield & Purvis 1999) for knowledge representation (KR). The 4I Framework research uses relationship diagrams to show the association among the different constituents of the artefact.

3.3.3 Evaluation

In this phase, the focus of the study was to evaluate 4I Framework (RO4) using two stage iteration.

Iteration1: *Illustrative scenario-based examples with reasonable arguments provided*

The 4I Framework evaluation was aimed to ensure that the developed framework can be implemented in the DGM space, particularly in the DE, in different contexts. In this study, to complement the survey assessment, the 4I Framework concept was validated through three real-life scenarios based on three different contexts, Smart Homes, Supply Chains and Wearable device. The applicability of the framework was discussed through an exploratory study with each of the three proof-of-concept cases published in peer-reviewed publications. This confirmed that the proposed framework can be used in real-life scenarios. The multiple case studies also allowed this study to generalize the findings. The final reviewed and improved version of the framework incorporated the advice from the surveys and the conference participants.

Iteration2: *Expert opinions were sought through survey offered to professionals from the industry*

The study focuses on capturing collective feedback and recommendations about implementing data governance and management practices from expert professionals from the ICT industry. This research collects feedback and opinions using an online survey. The assessment was done against the evaluation criteria (EC) adopted from the study by Prat, Comyn-Wattiau & Akoka (2014).

Table 3-2: Evaluation Criteria

Characteristic	Description
Coverage	The aim of this characteristic is to evaluate if the framework includes sufficient components required to implement DG and DM
Generality	The aim of characteristic is to assess how well the framework supported different businesses spread across different domains, as well as how easy it is to tailor the 4I Framework to different environments and circumstances
Clarity	The aim of this characteristic is to appraise whether the 4I Framework is clearly explainable and how easy it is to understand the framework
Efficacy	The aim of the efficacy characteristic is to determine the degree to which the artefact meets its desired goal of managing DG and DM activities in the context of DE
Importance and Relevance	The aim of this characteristic is to determine whether the elements included in the framework are relevant and important to address the DG and DM problems and challenges in an organisation
Novelty	This characteristic aims at demonstrating the framework's novelty (new knowledge)

3.3.4 Conclusion

The outputs of the DSR cycle are composed of the following:

- 4I Framework consisting of Drivers, Elements and Stages.
- Publications

The above phases on DSR cycle was performed in a manner (see Table 3-3) which conforms to the guidelines (see Figure 3-3) set out by Hevner in terms of contributing and rigorous IS artefacts.

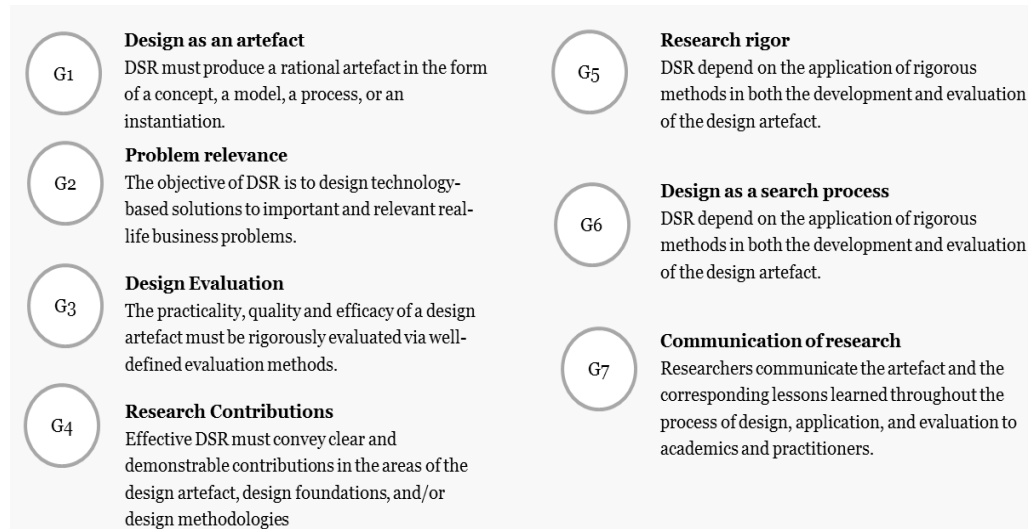


Figure 3-4: Design Research Guidelines

Table 3-3: Conformity of the research to Design Research Guidelines

Hevener et al.'s (2004) Guidelines	Conformity of the thesis To Hevener's guidelines
G1	The artefact developed in this thesis is a framework and questionnaires.
G2	The literature review and data breach-related activities on DGM show that the problem space being addressed is important and relevant.
G3	The artefact (in this case the framework) is evaluated by establishing a use case instantiation which is validated using expert feedback and descriptively assessed in terms of 'fit-for-purpose'.
G4	This research identifies the gap in the present knowledge base with respect to a DG in DE. A design artefact is delivered in the form of a framework.
G5	This research design follows well-founded prescriptions from the IS literature concerning the use of design science involving qualitative and quantitative methods. The design science construction component was guided by use-case development, the literature survey and evaluation phases involving iterative development through feedback through assessment of functionality. The original version of the framework and the final improved version went through several revisions resulting in the incremental enhancements.
G6	The evaluation of the 4I Framework (see Chapter 5) aims to establish if the framework addresses the research gaps explained
G7	Aspects of the research pertaining to the framework have already been peer reviewed and published. Material from this thesis has also been presented at practitioner (community) conferences and will also be published in the literature.

3.4 Research Instrument

The necessary resources used for the design, development and evaluation of the proposed framework are shown in Table 3-4.

Table 3-4: Research Instruments

Resource	Description
Research Process	Iterative DSR method to design, develop and evaluate
Data	Literature review based on <ul style="list-style-type: none">• Papers with references to DG and DM (industry practice-oriented)• Papers with reference to DG framework (academic) and DM• Reports on data breaches• Standards• Regulations
Feedback from industry experts	Survey questionnaires using <ul style="list-style-type: none">• Microsoft Excel• Google Forms Scenario-based use case study and publication feedback
Research ethics	Formal approval from UTS Compliance

3.4.1 Data Sources

The data sources used to identify the problem for this research consists of published academic papers, government reports, industry whitepapers and standards, organisation papers, consortium and media reports. Expert opinions from industry specialists (from Australia, India, Germany, Spain and Netherlands) and researchers (from across the world) are also incorporated to enhance the artefact.

3.4.2 Survey

Tools: A questionnaire using **Microsoft Excel** was piloted with three participants for clarity and applicability. This enabled the validation of the questionnaire by rewording and deleting ambiguous statements. Based on the feedback with regard to the ease of use, the Excel questionnaire was replaced by an online survey using **Google Forms**. The Google form used Likert ratings based on the commonly used structure (Hyndman 2008). For few of the respondents, face-to-face meetings or telephonic meetings were held to collect the survey answers.

Questionnaire development: The survey questionnaire was developed based on the analysis of existing relevant DGM theories, methods and components of the 4I Framework. All the questions in the questionnaire are based on the scenarios obtained from literature review, which can cause risk and potential loss in real-world implementation. The questions developed for the survey were peer-reviewed.

Survey feedback assessment: The survey collected feedback about the framework, both at an individual component level as well as an overall level (see Appendix). Thirty qualified participants involved in projects (with data and analytics responsibilities) were involved in the evaluation of the artefact. Both qualitative and quantitative data were generated from the participants' feedback.

Survey feedback analysis: Statistical formulas are suited to provide an analysis of a survey's numerical data (Hyndman 2008). This research uses statistical formulas such

as mean, standard deviation, skewness (measure of the asymmetry) and kurtosis (measure of the peakedness) to understand the survey's numerical data.

For the quantitative data, the feedback was examined against the EC by reviewing the responses provided by the respondents and provided a good indication of the practicality, relevance, coverage and clarity of the 4I Framework

3.4.3 Ethics Approval

Ethics are an integral part of IS research (Creswell 2009) where data is collected from people. It governs the moral responsibilities of the researcher and his/her obligation on what is morally acceptable behaviour (Kimmel 1988) during a research study to safeguard the rights of the respondents.

As part of the University's commitment to the Australian Code for the Responsible Conduct of Research and the National Statement on Ethical Conduct in Human Research, UTS policy requires any research conducted by UTS staff or students that involves humans must receive ethical approval from the UTS Human Research Ethics Committee (HREC) before proceeding. As part of the data collection requirement, an application for ethics approval was submitted to the UTS FEIT HREC panel. The research did not present any ethical concerns. The approval document can be found in the Appendix. A formal consent letter was sent to each of the participants. The letter included details about the project, the survey questionnaires, the anonymity of the data collection, and how the data would be stored. Once the participants gave consent via email, information about the online survey and the 4I Framework was also sent to willing participants.

Table 3-5 summarises the procedure from question design to gathering responses for the validation.

Table 3-5: 14-step Procedure from Data Collection to Data Erasure

Step Number	Description
1	Construct the Questionnaire
2	Apply for and obtain UTS ethics approval
3	Prepare consent form for survey
4	Prepare Participant Information Sheet for survey
5	Identify participants for survey from LinkedIn and Industry contacts
6	Send email to participants for consent
7	Send out survey form in Microsoft Excel format to 3 candidates who agreed to participate (with approved consent form signed.)
8	Based on feedback via Communication tools (LinkedIn messenger), replace MS Excel survey with online Google Form
9	Send out survey form to other respondents who provided a signed consent form using the Google Form
10	Met a few respondents face to face (at international conference and local locations) and filled out the survey
11	Concluded the survey after receiving 30 completed responses
12	Copied survey related records at UTS Cloudstor. This included: Scanning the copies of the physically signed consent form Copying the consent forms sent via email Copying the survey data from Google Forms and stored it (including the participant's country of work manually)
13	Deactivated Google Form survey (will be deleted)
14	Performed data analysis

3.5 Chapter Summary

This research was conducted to develop a new framework, the 4I Framework, for the DE. Initially, a few research theories that are relevant to answer the research questions were discussed. Keeping the RQ in mind, DSR was considered the most appropriate approach for this research. An overview and the activities in each phase were discussed, and the instruments used to conduct the phases were outlined. A proof-of-concept demonstration using the scenarios was recognised as an appropriate way to examine the feasibility of the framework. This was supplemented by the evaluation conducted through the survey involving practitioners and experts from industry to acquire information regarding the practicality, relevance and usefulness of the 4I Framework. In the next chapter, the 4I Framework is discussed in detail.

Chapter 4: The 4I Framework

This chapter presents the final revised version of the 4I Framework (see Figure 4-1), which addresses RQ3 (see Table 1-2). The 4I Framework has been designed and developed using a well-known DSR method as discussed in Chapter 3 and addresses the research gaps identified in Chapter 2. The 4I Framework has been iteratively refined based on the industry feedback using the DSR method. The framework is made up of three key components: (a) drivers, (b) elements and (c) stages. A detailed view of the 4I Framework components and their relationships are presented in the following sub-sections.

4.1 Framework Overview

The final version of the 4I Framework evolved through different iterations. The initial version of the framework was developed in 2018 based on the insights derived from literature review. It comprised of four stages and seven key elements (see Figure F-1). It is important to note that this version did not incorporate any feedback from the industry. The second iteration of the artefact involved enhancement and refinements based on further literature review of events related to data breach in 2019 (see Figure F-2). The third iteration was developed in 2020 was based on the feedback provided by domain experts. There were some major changes in this version such as the introduction of the driver component. The final version of the framework, as illustrated in Figure 4-1, is made up of three main components: 1) drivers, 2) elements, and 3) stages that collectively help in addressing the DG and DM for an organisation in the DE context. Each of the components has a different purpose and focus as follows:

Drivers: Drivers focus on the organisational strategy and explain the main reason why an organisation needs to undertake DGM initiatives.

Elements: Elements focus on the core items such as people, processes, collaboration, and the essential toolkits required for the framework to work.

Stages: Stages focus on summarising the activities that need to be undertaken systematically by an organisation to achieve the objectives of the drivers with the elements. The 4I Framework comprises four stages. The implementation success of the framework to realise the requisite outcome is dependent on the proper implementation of the stages.

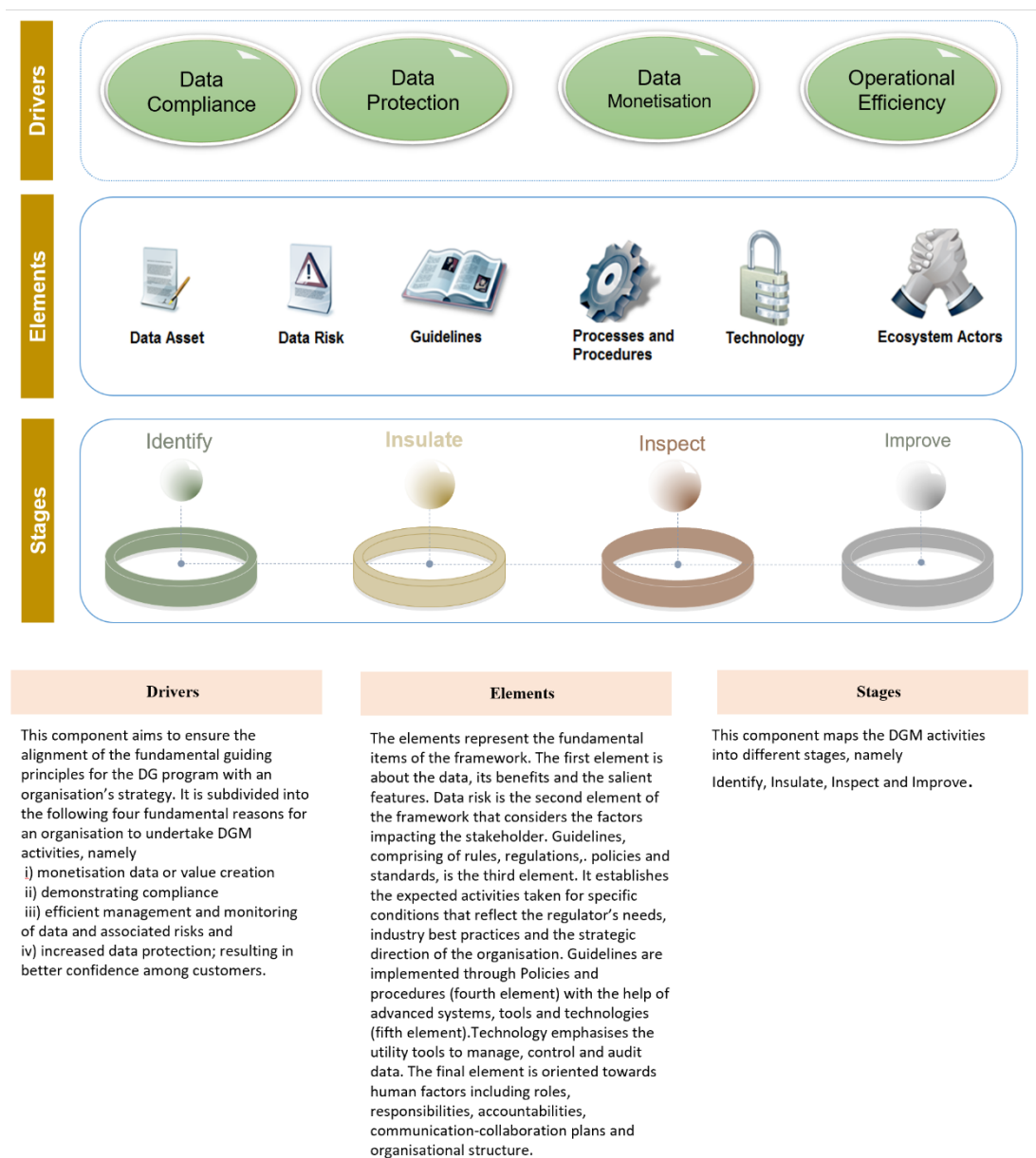


Figure 4-1: The framework components

While the framework includes several elements, the emphasis is given to the newer challenges arising that are particularly relevant in the DE, such as data related risks, regulatory compliance and data actors and their interactions. The framework also encompasses the well-established operational data governance and management items

such as metadata management, data quality and value and performance optimization. However, as mentioned in Section 1.5, the level of details provided for the aforementioned topics are at a high level.

4.1.1 Drivers

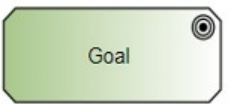
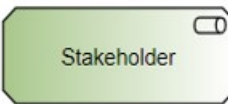
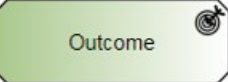
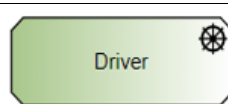
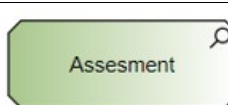
This component aims to ensure the alignment of the fundamental guiding principles for the DG program with the organisation's strategy. The driver's central idea is to ensure the participants or stakeholders understand the strategic objectives, outcomes, and critical success factors needed to establish the requirements and processes to govern.

Drivers comprise a set of strategic initiatives such as

- i) Data compliance
- ii) Data protection consisting of data security and privacy activities
- iii) Data monetisation or value creation
- iv) Operational efficiencies including optimized data curation and control processes linked to organisational productivity

The associations between the drivers and the goals, outcomes and stakeholders are explained below. Table 4-1 gives an overview of the notations used and the definitions.

Table 4-1: Notations used

	Represents a high-level statement of the desired target state for the stakeholders of an organisation in DE
	Represents the role of an individual, team or organisation that represents their interest in DE
	Represents the end result
	Represents the reason that motivates an organisation to define its goal and implement the changes to achieve them
	Represents the result of the anlysis of the current state of the organisation with respect to certain driver

4.1.1.1 Data Compliance

Data compliance refers to the act of obeying and demonstrating adherence to corporate policies, external laws, regulations, and procedures. In the face of growing technology dependency and cyber risk, a lack of data compliance is one of the key obstacles for digital businesses. A regulatory agency is empowered to oversee and enforce compliance within a given industry sector through various measures such as products, antitrust rulings, and compliance laws. From an organisation's perspective, the biggest challenge is how to decode the legal terms into the regulatory agency's relevant compliance intentions and ensure the same interpretation between agencies and organisations in both inter- and intra-organisational settings of digital IS. Data compliance requirements are highly volatile and involves various levels of complications that can create challenges for organisations. With non-compliance resulting in organisations being fined heavily, data compliance is a key motivating factor for stakeholders to undertake data compliance conformance activities. As illustrated in Figure 4-2, the main accountability is to ensure that compliance changes are met from the CEO's or management's perspective. By being compliant, a business can satisfy regulators and safeguard itself from paying fines or other financial liabilities. Being compliant is critical in improving the reputation of the organisation.

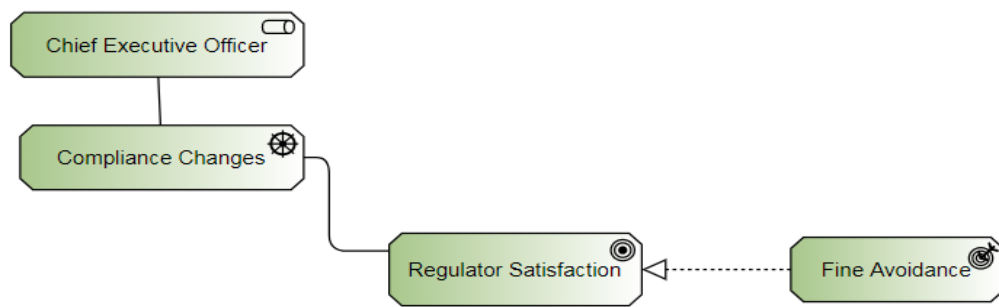


Figure 4-2: Data Compliance perspective

4.1.1.2 Data Protection

With the rise in technology dependence and cyber security incidents, safeguarding data has become a high priority for organisations. Data protection revolves around security, privacy, resourcing, the operating model and consent concepts and aims to ensure threats to external and internal data are controlled through control mechanisms. A data loss event can have a severe impact on the organisation's reputation.

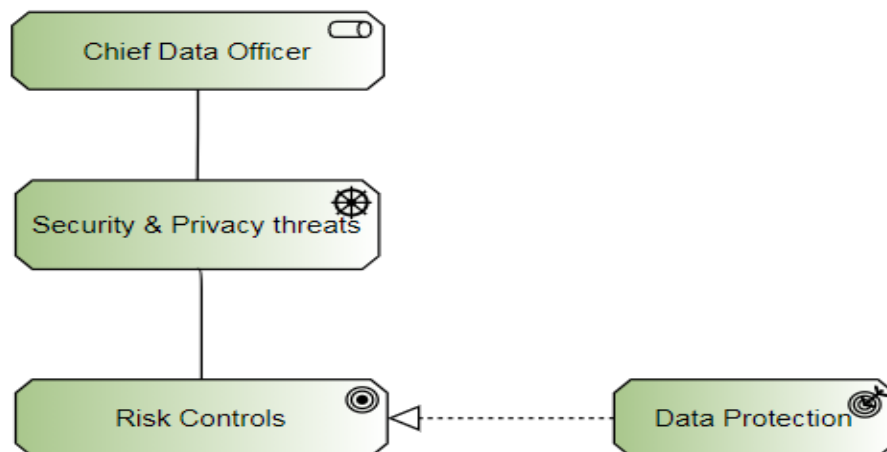


Figure 4-3: Data protection perspective

4.1.1.3 Operational Efficiency

Business organisations recognise the criticality of data and the challenges faced in managing data spread across various disparate source systems in the DE. But data-

driven decision making is not able to achieve reliability unless the errors in the data are rectified. Furthermore, the absence of standard processes and clearly defined roles and responsibilities for DG leads to unnecessary decision-making delays. The operational efficiency of the business function drives the tactical and strategic approach to be executed on a regular basis to govern and manage data using processes and technology proactively. Operational effectiveness results in the removal of ambiguity and reduces duplicated efforts. As illustrated in Figure 4-4, it aims to improve efficiencies in the way data is handled and combines a number of concepts covering the entire data management life cycle including understanding data flow, data quality, data lineage and data consumption, data duplications, etc. It also establishes the groundwork to meet the future demands of the regulators and customers.

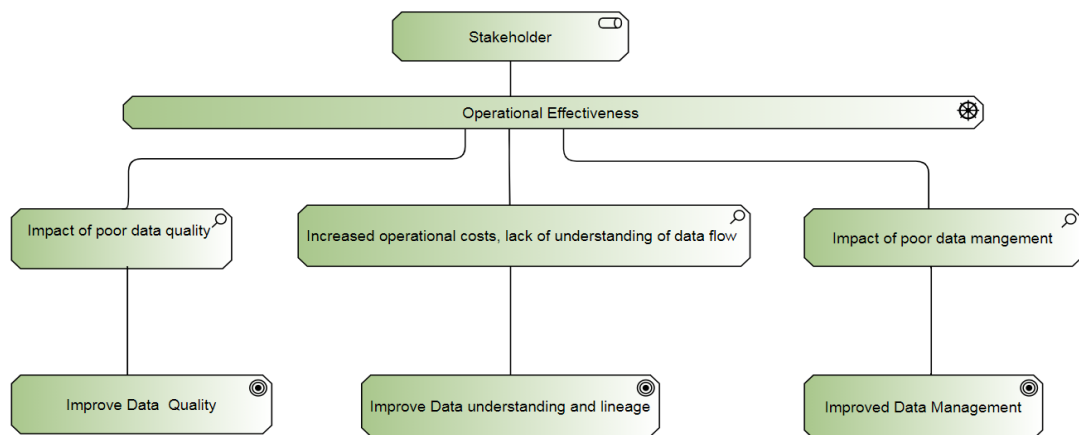


Figure 4-4: Operational Efficiency perspective

4.1.1.4 Data Monetisation

The value of data is a characteristic that refers to the usefulness of data for making a better decision. Through a thorough understanding of customer-centric needs, organisations aim to derive value from clear and timely insights provided to the stakeholders and customers. Thus, by harnessing the power of the data and extracting

knowledge from data, newer solutions result in increased customer satisfaction. In the same vein, from the CEO's point of view, this increased revenue maximises profitability and provides a competitive advantage.

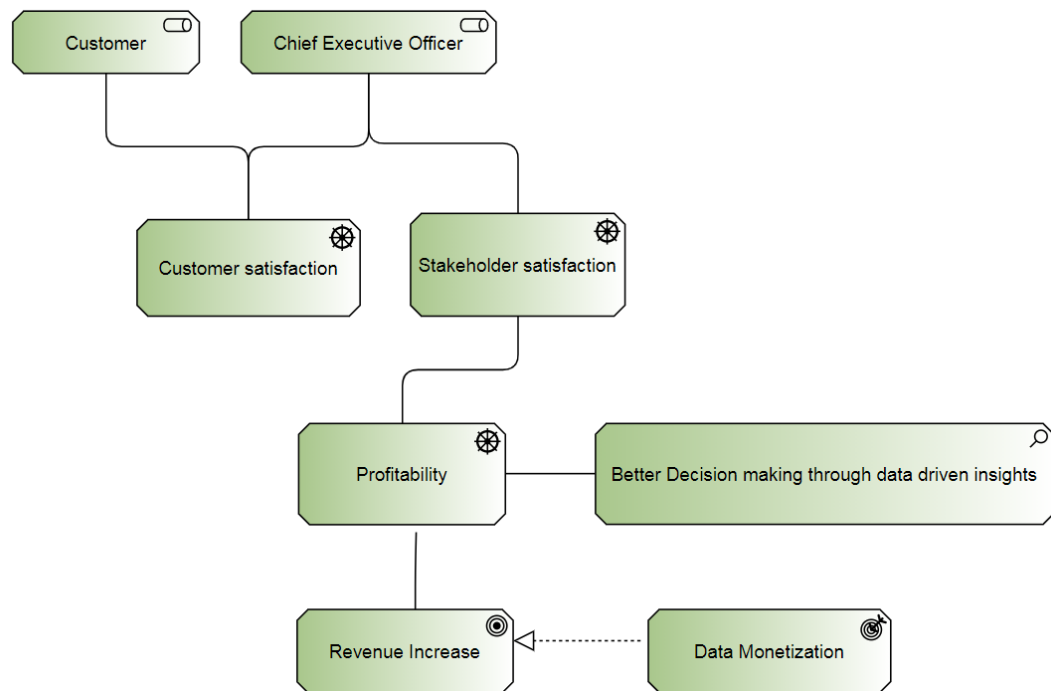


Figure 4-5: Data monetisation perspective

4.1.2 Elements

The elements represent the fundamental elements of the framework. As illustrated in Figure 4-6, the framework includes (1) data asset, (2) data risk, (3) guidelines, (4) processes and procedures, (5) technology and (6) ecosystem actors. Collectively, the elements discussed in this section equip the stakeholders to identify and implement appropriate controls and remediation activities to manage and utilize data effectively.



Figure 4-6: Elements in the 4I Framework

Table 4-2: Elements of the 4I Framework

Element	Description
Data Asset	Data Asset is concerned with what actual physical data needs to be governed in the DE. The benefits of the data asset and the prominent characteristics of the data asset are covered in this element
Data Risk	Data risk is caused by exposure to disruptions, threats or vulnerabilities resulting in financial or reputational loss to the organisation. The data risk element concentrates on determining and compiling the factors that affect the stakeholders and provides input in the development of the risk control mechanism.
Guidelines	The aim of this element is to establish boundaries and expected actions taken for specific circumstances, settings, situations, or conditions that reflect the regulator's needs, industry best practices and the strategic direction of the organisation. Guidelines comprise rules and regulations, policies and standards.
Processes and Procedures	Processes and Procedures defines how the different interfaces and functionalities work together to deliver an operational solution based on the guidelines.
Technology	Technology is the key enabler in the framework. It includes both hardware and software agents, applications, platforms and infrastructure technologies that are required to support the digital ecosystem.
Ecosystem Actors	The focus of this element is on the different parties participating in the data-related decisions and activities in the ecosystem. This includes leadership, organisation structures, roles and responsibilities.

4.2.1.1 Data Asset

A data asset is the foundational element in the 4I Framework. It was observed in Chapter 1 (background) and Chapter 2 (SLR analysis) that data identified as an asset, if not governed properly, can result in a failure to accommodate the needs of a firm. A data asset is concerned with “what” actual data needs to be governed.

As shown in Figure 4-7, a data asset is considered a subclass of the data, having properties such as data structure, storage, interface and location.

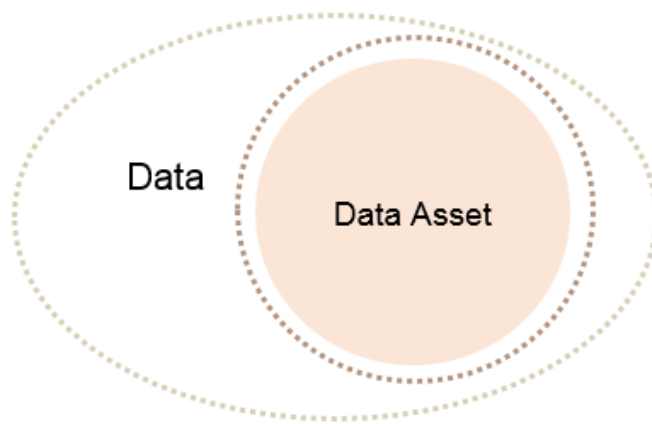


Figure 4-7: Data Asset as a subset of data

A data asset is typically composed of the collection of data elements from which value or innovation can be derived to contribute to the organisation’s business growth with the help of decision making. The selection criteria of data as an asset are organisation-specific and are identified based on the input from data consumers or business stakeholders. The benefits of data can be used to provide insights to make decisions, reduce operational costs, improve or monetise productivity through data sharing arrangements.

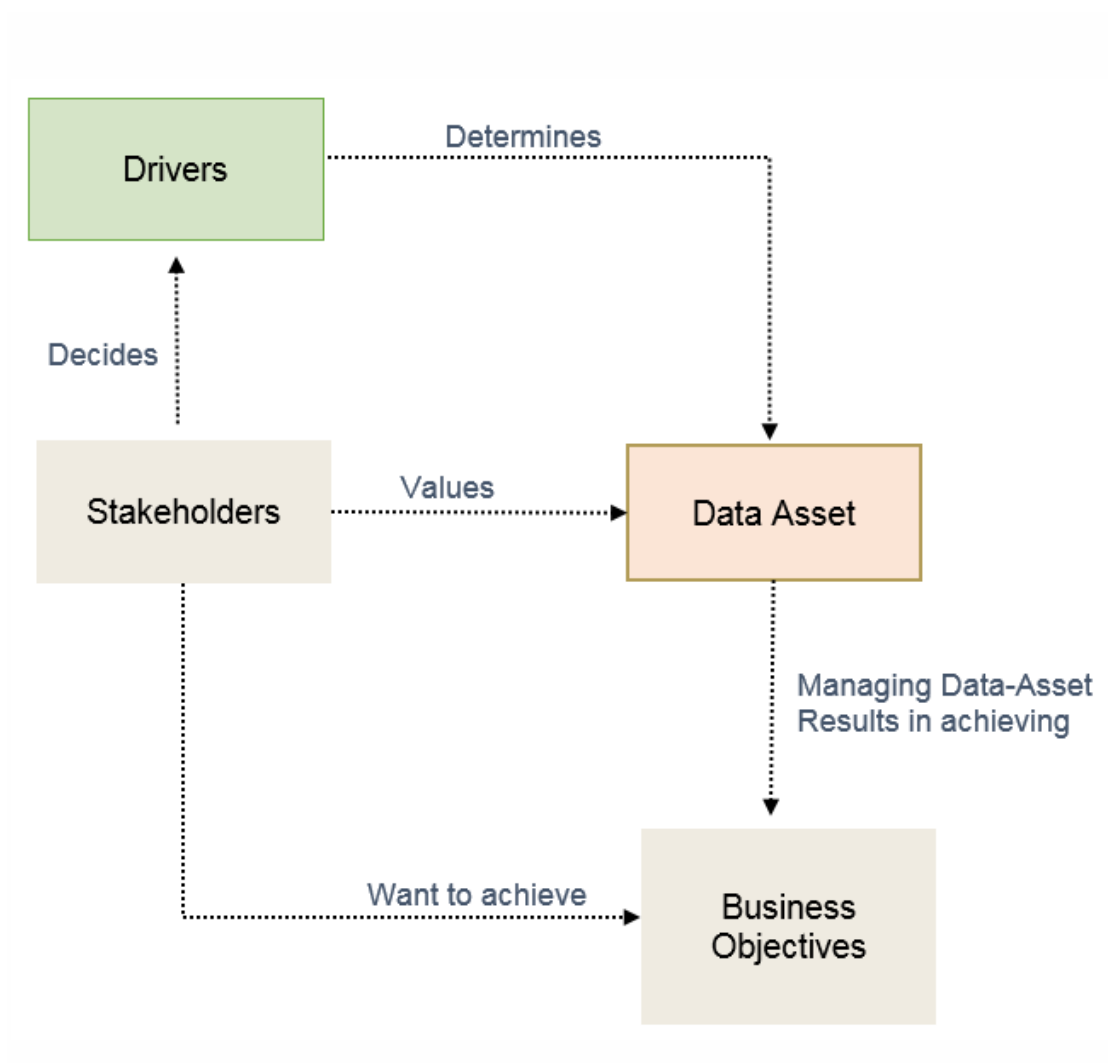


Figure 4-8: Data Asset relationship to the organisational objectives

A data asset is often mandated by law or policy recognised by an organisation to be governed depending on its contents, particularly when exchanged with an external party. For example, the *Basel Committee on Banking Supervision's Standard Number 239 (BCBS239)* standard requires financial institutions to scrutinize the contents of data.

Any challenges concerning data quality (DQ) or data breaches can have an adverse impact on the reputation of an organisation, its customer privacy, and trust and user experience. Therefore, an inventory of data assets are the basis to assess health and

risks associated with data that enables the transparency and auditability of the data assets.

Table 4-3 summarises the key taxonomy of this element.

Table 4-3: Taxonomy of the Data Asset element

Sub-element	Description
Data Value	Describes the business value of the data to the organisation.
Data Location	Describes the storage medium of the data asset including if it is in transit or at rest.
Data Classification	Describes the categorisation of the data asset from a privacy perspective such as confidentiality, sensitivity, etc.
Data Availability	Describes the availability of the data asset from a real-time perspective.
Data Structure	Describes the technical details of the data asset such as unstructured, semi-structured, data volume rate of growth, etc.
Data Integrity	Describes the completeness and accuracy of the data.

4.2.1.2 Data Risk

The data risk element of the 4I Framework is linked to the data compliance and data protection stream discussed in the drivers section of the framework. As observed in Chapter 2, regulators consider data risk as a subset of the operational risk.

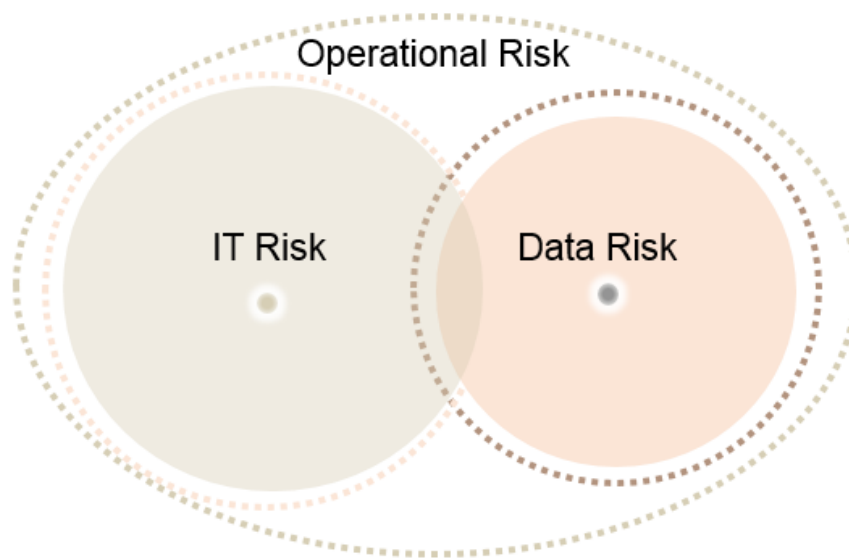


Figure 4-9: Data Risk as a subset of Operational Risk

A data risk can be described as uncertain future events that might lead to the failure to address multiple business goals. The risk event in the 4I Framework can be **caused** either through a) the failure of people, processes, solution design, development and technology or b) cybersecurity incidents including malicious activities. Regardless of the risk being posed by malicious activities or inadequate controls, it is vital to have a thorough understanding and documentation of the risk event, its cause, and impacts as the **impact or consequences** can be both financial and reputational. However, the appetite to live with the potential consequence is specific to the **risk appetite** of the firm.

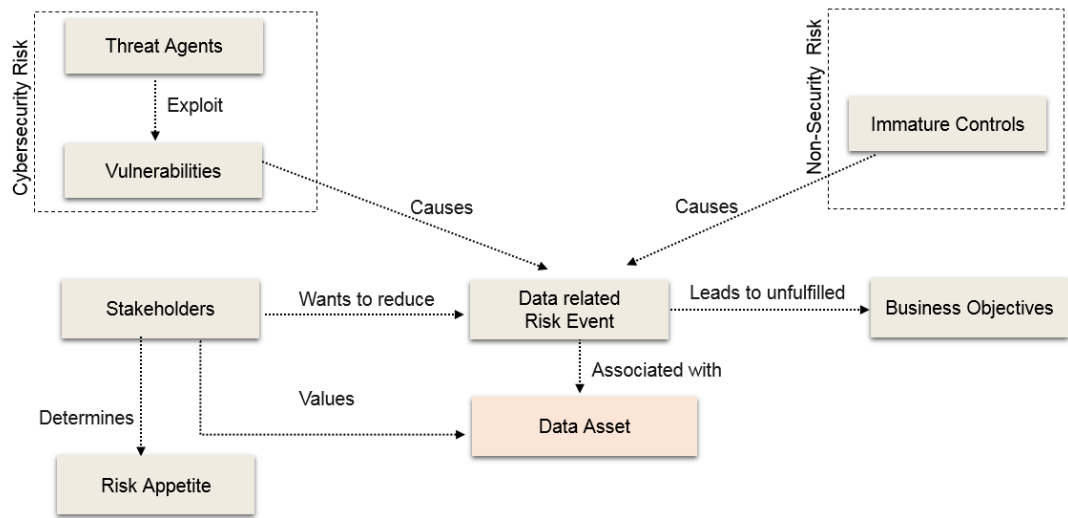


Figure 4-10: Data Risk relationship to Data Asset

By documenting a risk event, its cause and impacts throughout the data lifecycle, this element is able to determine the risk priority and identify and implement the appropriate guidelines and additional corrective actions required to manage the risk effectively. There can be several types of risk factors such as the unauthorised loss of data, breaches by an external actor due to consent violation, the unexpected loss of service due to a service level agreement (SLA) violation. In addition, the risks associated with security, upgradability and interoperability need to be considered.

Table 4-4 summarises the key taxonomy of this element.

Table 4-4: Key taxonomy of the Data Risk Element

Sub-element	Description
Stakeholders	Represents the role of an individual, team, or organisation that represents their interest in the effective data management
Data Risk Event	An unforeseen event that may potentially threaten the achievement of the business objectives. There can be several types of risk events such as service level agreement violation, privacy abuse.
Data Risk Cause	Key factors that contribute to the occurrence of a risk event
Data Risk Impact	The impacts that could result if a risk event occurs

Non-Security Risk	Risk associated with the use, ownership, operation, involvement and adoption of technology in the ecosystem.
Cybersecurity Risk	Risk associated with the threat posed by a threat agent exploiting vulnerabilities, data breaches or unintentional events involving data resulting in a financial or reputational loss
Vulnerability	Weakness found in software and hardware components that, when exploited, results in an undesirable impact related to confidentiality, reliability, or availability.
Threat Agent	Methods and things used to exploit a vulnerability
Risk Appetite	Extent to which an organisation is willing to accept risk in pursuit of its business objectives

4.2.1.3 Guidelines

Guidelines are one of the core elements within the 4I Framework that provides the basis to apply a consistent approach to enforce conformity and compliance. It spans the entire lifecycle of the data and provides guidance on collecting, processing, documenting, measuring, tracking, and communicating the activities involving data assets. The guidelines in the 4I Framework aim to establish boundaries and the expected actions to be taken for specific circumstances, settings, situations, or conditions.

Guidelines are made up of rules and regulations, policies and standards.

Rules and Regulations: This component is directly linked to the data compliance strategy of the organisation. An essential outcome of this element is to provide clarity on the legal requirements (for example, user consent) and controls the requirements. Its intent is to provide baselines to be operationally compliant and avoid the steep penalties associated with non-compliance. It comprises the i) domain specific legislative needs, ii) contractual obligations related to the capture, use and disclosure of the data or intellectual property and iii) corporate compliance as its subcomponents.

One of the key assumptions or expectations while working on this component is that it is highly volatile and frequently changes change over time. Moreover, restrictions on data locality are also expected to change from time to time and cannot be assumed to remain static. Hence, concerns such as trans-border or third-party data transfer mechanisms, consents, breach notification management, end user data usage consent management and complaints need to be investigated periodically.

Policy: Policy indicates the intents, rules and assertions that specify the high-level expected conduct of an entity to achieve its organisational goals within the context of DGM. It is crucial that the policies are well-defined and understandable. It is vital that policies reflect regulator needs, best practice and the strategic direction of the organisation. For instance, an authorisation policy can stipulate the precise access control rules for a software application. Policies such as acceptable data usage, data quality, data privacy, data sharing agreement, data classification and data storage requirements and compliance (e.g., GDPR, HIPAA, SOX, etc.) must be adhered to to ensure the organisation meets its DG and DM objectives.

Standards: Standards play an important role in defining the specific rules necessary to ensure policies are properly implemented to manage data. Several standards exist that provide specifications and directives for different areas of DG and DM such as ISO-8000, ISO25012 (data quality), CPS 231(data outsourcing), ISO-27001 (data security), ISO-11179, CWM, DCMI (metadata management), record management (ISO-15489), and payment card security management (PCI DSS). Standards provide the specific rules, expectations or criteria which must be met to comply with policy. The details of the different sub-elements are described in Table 4-5.

Table 4-5: Key guidance to support data-related decision-making

Sub-element	Description
Data access	Describes who can access the data, how it is managed and the entitlements involved.
Data risk management	The coordinated management of business operations, technology, cyber defense, people resources, and intelligence to effectively mitigate threats, vulnerabilities, and consequences that could arise in case the data asset is not protected properly.
Data outsourcing	Describes how the data is exchanged with third parties along with its purpose as well as ensuring organisations preserve control over the data in inter-organisational settings. Data center or infrastructure outsourcing is also part of this.
Vendor engagement	Describes the rights and obligations of stakeholders which are established through mutual contractual agreements which maintain their relationship.
Consent	Describes how consent is explicitly gathered from the data producer and how ensuring the information of users who have “unsubscribed” from the suppliers’ systems are passed on to the user.
Data retention or archival	Describes how the data is retained and finally disposed of at the end of the lifecycle with advice obtained from the CDO or records management team prior to disposal of data or digital content.
Data monitoring	Describes how to ensure governance standards are applied and controls such as minimization are considered and implemented. Ensuring that the processes which are in place to support the implementation of the policy are adequate.
Data acquisition	Describes how the data access rules, mappings as well as data integration between systems and organisations occur.
Data usage	Describes how the data assets are consumed, both internally and externally. The groundwork for privacy and data impact assessment is also covered here.
Data ethics	Describes policy to achieve ethical data usage. It covers ethical matters concerning the entire data lifecycle, including the codes, standards and responsible innovation
Social media	Describes how the employees’ use of social media both on the job and for private activities.
Data sovereignty or localization policy	Describes the data residency constraints.
Data classification	Describes how the data assets are classified into different groups depending on the criticality and protection requirements. It is also the pre-requisite for data exchange requests in DE and the basis of monitoring data for privacy, security, and legislative requirements.
Data hosting policy	Describes how the data is stored physically.
Data security	Describes the practices to safeguard data using techniques such as patching, encryption, obfuscation and minimization.

Data interoperability	The rules of interoperability between all devices, while remaining unique and secure.
Communication policy	Describes how policy must be communicated and available to the relevant stakeholders including data breach reporting and the response plan.
Data provenance	Describes the origin of a piece of data and the processes and methodology used to produce it.

Through well-defined DGM guidelines, firms can implement a successful data-driven organisational culture where both business and IT departments understand and share equal responsibility. Moreover, by broadcasting the guidelines on definitions, practices, methodologies and expectations, the awareness and understanding of the data practitioners and consumers in the ecosystem of data management activities can be increased.

4.2.1.4 Processes and Procedures

Guidelines do not express the details of specific rules or standards. However, the processes and procedures provide the specifics about how the guidelines are implemented. It comprises several sequential states and outlines how the different interfaces work together to implement the desired solution where all processes and procedures are clearly linked to at least one guideline. Thus, processes or procedures are derived directly from the guidelines and stipulate who does what, how and when to implement the guidelines with the help of technical or non-technical measures in a controlled manner. The activities in a process or procedure can be interrelated and may interact with one another. However, it should be noted that processes and procedures must be supported by a range of tools to enable the execution of the activities.

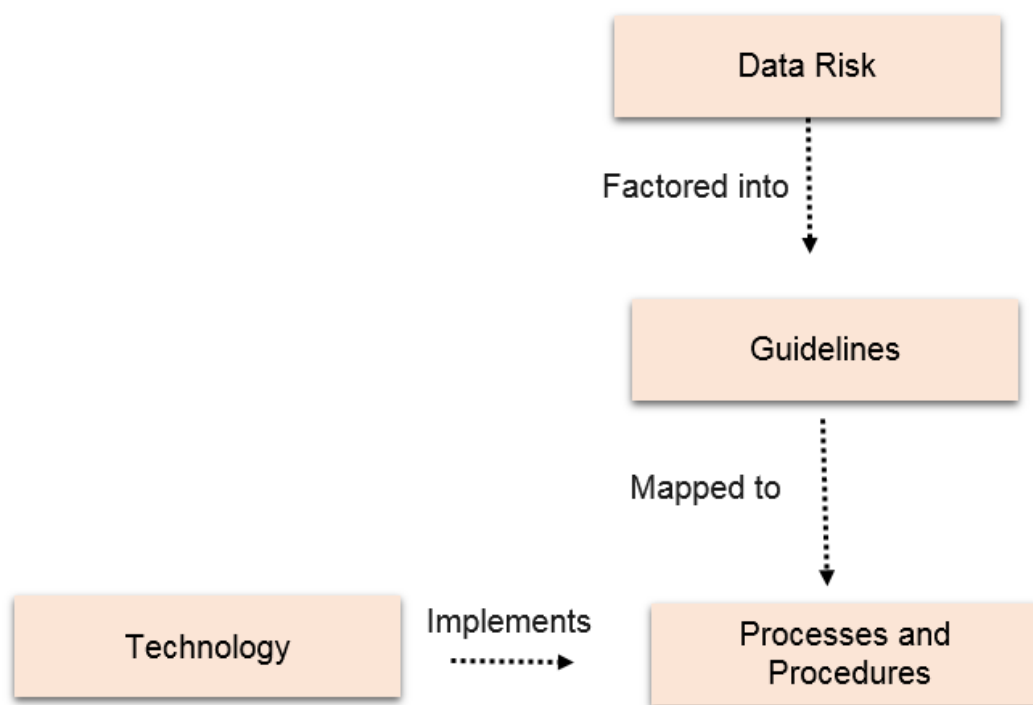


Figure 4-11: Relationship of Guidelines with other elements of the Framework

Processes can be either external or internal and are anticipated to be triggered by events which can be external to an organisation or internal. It is critical that the processes and procedures encompasses the entire data lifecycle (see Figure 4-12 for illustration) from inception to destruction.

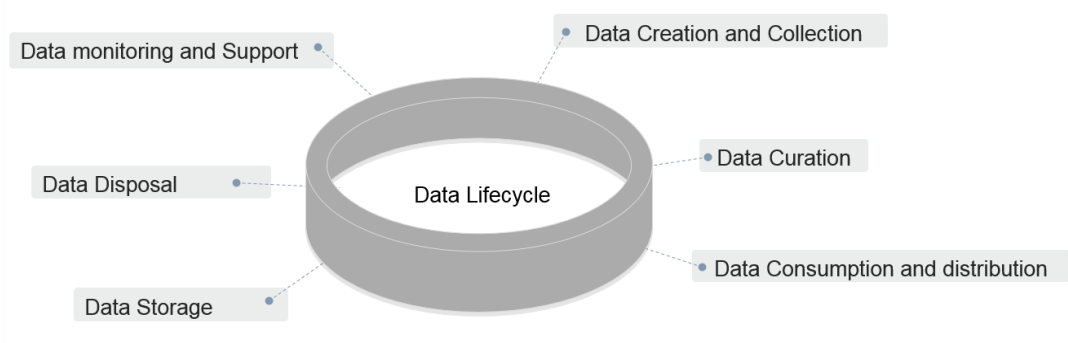


Figure 4-12: Key processes or procedures to support data lifecycle

The key taxonomy of this element is shown in Table 4-6.

Table 4-6: Key processes and procedures

Sub-element	Description
Data creation and Collection	Describes the process of data acquisition either through automated processes or manually created
Data curation	Describes how the data is managed. It covers topics such as the data interfaces, data migration, data conversion, data processing, data-deidentification, etc.
Data consumption	Describes how the data is used either directly or through sharing both external and internal to the organisation.
Data storage	Describes the medium where the data is being stored.
Data disposal	Describes the removal of data at the end of its lifecycle
Data monitoring and support	Describes the monitoring and support for day-to-day activities

4.2.1.5 Technology

Technology is the penultimate element in the framework that acts as a key enabler to support the DGM activities. It includes information technology-related services, and comprises of hardware and software agents, applications, tools, platforms, collaboration tools and the infrastructure technologies required to support the ecosystem. It provides the ecosystem actors implementing day-to-day data management activities with the necessary toolkits to perform the job effectively and encapsulates data-related systems.

The 4I Framework is not fixed to any specific set of technology. Each organisation or team may use different technology to support the DGM activities based on the requirement, complexity and budget. However, regardless of what technology is used or how diverse it is, it is important to ensure that it is secure and supports evolving needs such as real-time processing, fog computing, blockchain and IoT. In addition, organisations must ensure that they have actors with the expertise to support the use of the technology. As indicated in Figure 4-11, technology is linked to the processes and procedures element and is implemented by ecosystem actors and includes control objectives. Table 4-7 outlines the fundamental sub-elements of technology.

Table 4-7: Key Technology related sub-elements of the 4I Framework

Sub-element	Description
Application landscape	Describes the universe of the applications that are involved in handling data. Applications can be cloud-enabled SaaS, web or mobile applications, an application that exists on a virtual machine or a standalone application running on an on-premises server.
Data technology	Data technology comprises tools (platforms, data stores) to support the data lifecycle process. This can vary from spreadsheets, documents to sophisticated automated tools that support the management and maintenance of functionalities such as meta-data, DQ, security master data and the business glossary. It can include cloud services such as platform-as-a-service or software-as-a-service.
Infrastructure	This describes the data infrastructure which is made up of physical hardware and virtual resources. Its purpose is to support the flow, storage, processing and analysis of data.

4.2.1.6 Ecosystem Actors

An ecosystem actor is the core element of the framework and consists of two parts, an individual or organisation and the roles they play. As pointed out in the beginning of this thesis, DE is a complex network of stakeholders that encompasses several industry verticals. These stakeholders comprise data producers, data suppliers, data consumers or customers, data processors, and platform orchestrators connected through digital platforms and they interact in ways that creates value for the organisations. Data subject is the heart of the ecosystem. With the ever increasing need for data sharing collaboration among stakeholders, DG initiatives have become transboundary and inter-organisational, with the need to engage multiple participating groups to manage and supply data. This requires an organisation to have not just an internal viewpoint, but to take care of the obligations beyond the organisation boundary.

As shown in Figure 4-13, there are several parties involved who can be categorized as either external or internal organisational governance, each with different roles and

accountability depending on the implementation and context (van den Broek & van Veenstra, 2018). Depending on the organisation, the governance structure can be centralized, federated or decentralized. Smaller organisations may have employees who take responsibility for several roles, while larger firms may have multiple employees for one role or add additional positions.

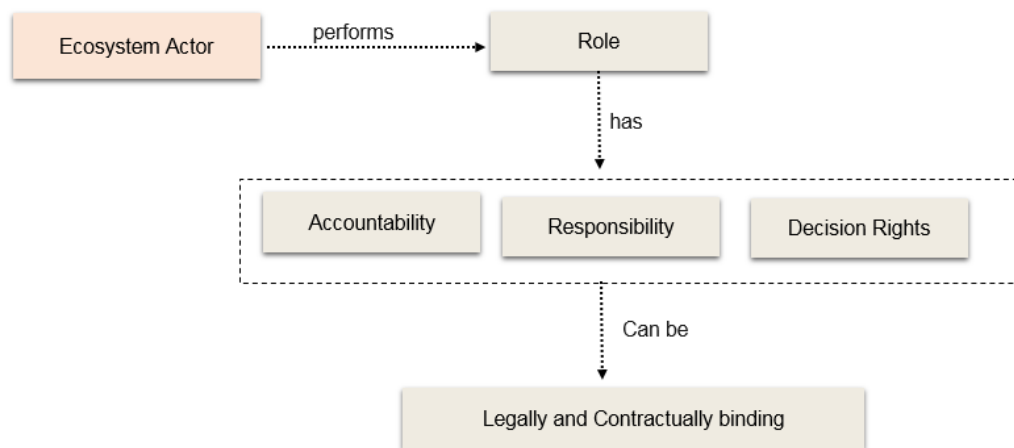


Figure 4-13: Ecosystem Actors' roles and responsibilities

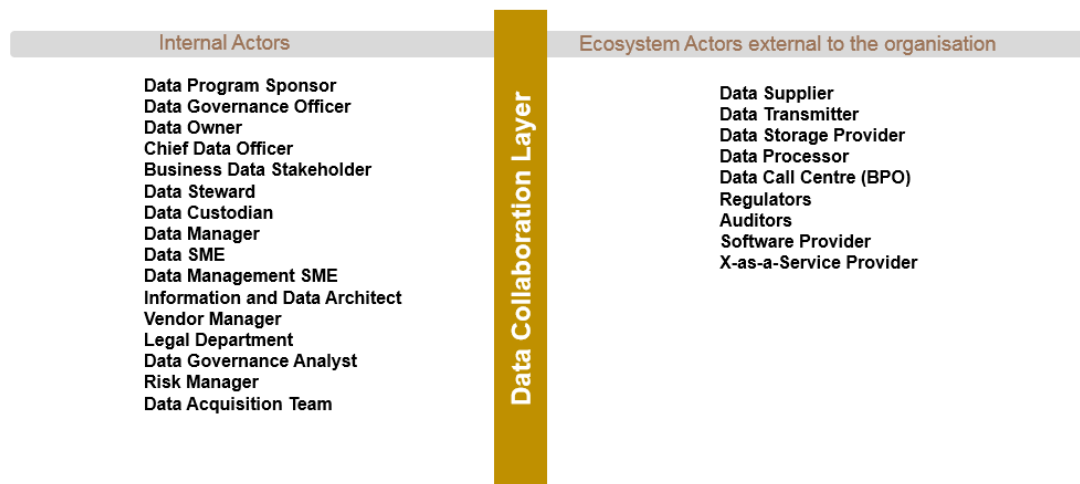


Figure 4-14: Actors in the 4I Framework

In the above figure 4-14 lists the internal and external actors considered in the 4I Framework. The internal players of the focal firm involved in DG and DM activities that are as follows:

Table 4-8: Key internal actors

Role	Description of the Role
Data Sponsors	A data sponsor is part of any firm's executive board and chairs the steering committee. A data sponsor drives the adherence to data guidelines, champions data quality for an identified critical data asset and leads the data as an asset culture. A data sponsor closely collaborates with the owner of the implementation of the data management of the data asset. The tasks involve assigning and communicating DG roles to relevant stakeholders, providing strategic and operational direction in the implementation of DGM processes, establishing the data-related objectives and execution strategy based on the priorities set by the CDO and working groups. Additionally, a data sponsor is accountable for and provides formal attestation concerning the management of the relevant data asset and guides the resolution of escalated issues for the data asset.
Data Owner	The data owner is accountable for the data asset. The data owner is usually the key data consumer or data user and plays a key role.
DG Officer (DGO)	The DG officer has the overall accountability when it comes to implementing enterprise-wide DG capabilities. DGO provides visibility of enterprise-level data initiatives and business priorities.
Chief Data Officer (CDO)	The CDO is accountable for the organisation's strategic and governance activities. The CDO develops and implements a company-wide data strategy to manage data as an asset and drive improvements in DG and DQ across delivery collaborators.
Business Data Stakeholder	The data users are influenced by data-related decisions. Depending on the situation, data stakeholders are involved in data management activities.
Data Steward	Data stewardship is the management of data on behalf of all stakeholders for their benefit. Data stewards act as SMEs for a data asset.
Data Custodian	Data custodians, sometimes referred to as data operators, perform business and technical onboarding, maintenance and end-of-life updates to data assets.
Data Referee	The data referee plays the role of coordinator when data is shared between organisations.
Data Manager	The data manager is responsible for helping business users understand and extract value from data by driving the adoption of enterprise information management capabilities. They are responsible for data content, definitions and implementation of policies, standards and business rules.
Data Subject Matter Expert	The data SME provides project or BAU team specialist knowledge of a data asset and the business rules.
Data Management	The data management SME is responsible for enterprise oversight of specific capabilities including data quality and standard business.

Subject Matter Expert	
Information and Data Architect	The information and data architect designs, creates, deploys and manages data architecture rules & definitions, business processes & data flows, master data and BI & reporting.
Vendor Manager	The vendor manager is responsible for contract and relationship management with the partners. The vendor manager is also responsible for ensuring third-party ecosystem players adhere to data management requirements. The vendor manager also develops contingency provisions to avoid third-party risk.
Legal Department	The legal department is responsible for setting up the contractual clauses in ecosystem partner engagement agreements.
Risk Manager	The risk manager role is to oversee the practice and development of RM throughout the organization.

The external players are as follows:

Table 4-9: Key external actors

Role	Description of the Role
Regulators	The government department that has the responsibility to administer and enforce compliance with a given sector. They can impact a variety of businesses through regulations on products, antitrust rulings, compliance laws, and so on. As an example, in Australia, APP fined VWFS out of concerns that the unlawful collection of data gives an unfair competitive advantage in the market.
Auditors	This is a trusted role that is responsible for auditing the data lifecycle-related activities. Auditors' duties include performing assessments, maintaining and archiving audit logs and conducting or overseeing internal audits.
Vendors (including suppliers and partners)	<p>Any legal entity with which the organisation has entered into a business relationship that provides a product or service to the organisation or its customers. Vendors can be:</p> <ul style="list-style-type: none"> • IT service providers (including API connector) • Third-party advisors (including legal services) • Software • Hardware • Licensed data supplier • Product • Marketing and media product provider • Telecommunication services • Non-IT service provider (including training or event partner) <p>Vendors can be involved as outsourcers for application development & maintenance, business process support and IT services.</p> <p>Vendors can be data producers or data consumers.</p>

In the 4I Framework, the intra-organisational operating **model** can vary depending on the nature of the organisation. For example, Figure 4-15 shows an example structure based on traditional DG practices.

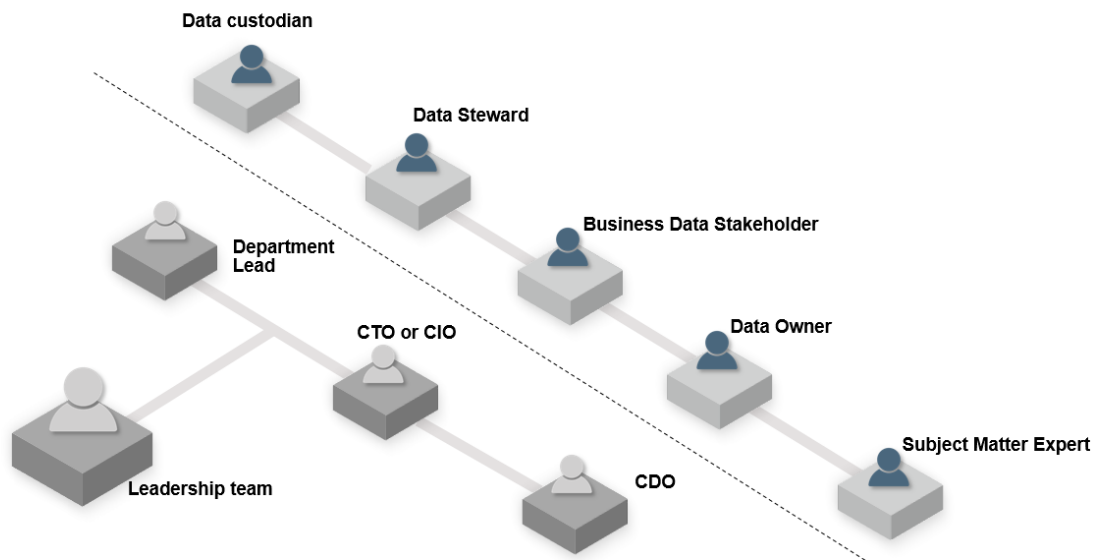


Figure 4-15: Sample governance structure of intra-organisational actors

However, when it comes to the inter-organisational context, different operating models such as shared governance or focal organisation-driven governance can be used.

Shared Decentralised Governance: Ecosystem actors in different organisations collectively govern, with each system owner taking the ownership of the respective system with the vendor manager of the organisation playing a key role. In this type of decoupled operating model, the practices are not consistent across the actors and continuous mass collaboration and co-ordination is required among the different actor. The shared decentralised governance is suitable for a dyadic relationship involving focal enterprises and another partner. It requires heavy reliance on formal contractual obligations such as non-disclosure agreements, security obligations, and service level agreements guiding data sharing and management arrangements.

Centralised Governance: In this model, one organisation is responsible for the proper administration of the data in the ecosystem. This ensures formal control power is with the focal enterprise throughout the collaboration lifecycle. For example, the owners of a platform, can play a dominant role and provide the requirements and instructions to the participating actors on how to handle data depending on the platform strategy. However, for co-responsibility to work, the external actors must provide certification of compliance to the rules periodically.

Hybrid or Federated Governance: In this model, the centralised and decentralised operating models are combined where DG activities are co-ordinated centrally with a central team facilitating the interactions. Each individual organisation operate independently and manages data autonomously, supported by a central function. In several cases, an external entity with no involvement in the data assets, such as a consulting firm or independent expert, can take the responsibility for the governance. In this operating model, the co-ordination between collaborating actors is performed by the data referee, who defines the rules of data exchange including data format, privacy, security and compliance requirements based on factors such as data asset and data risk. While the "data referee" provides high-level governance (see Figure 4-16), there is a need to have the active participation of different actors in the data decision-making process related to developing and approving policies, as well as ensuring the processes to support the implementation of the policy are adequate. The function of this role is to initiate, own, make decisions, act at specific points within a process or complete a process. Each responsibility can be assigned to a person, group or even machines, especially in the case of AI and IoT. Similar to the centralised governance

operating model, all actors in the federated operating model need to comply with the guidance provided.

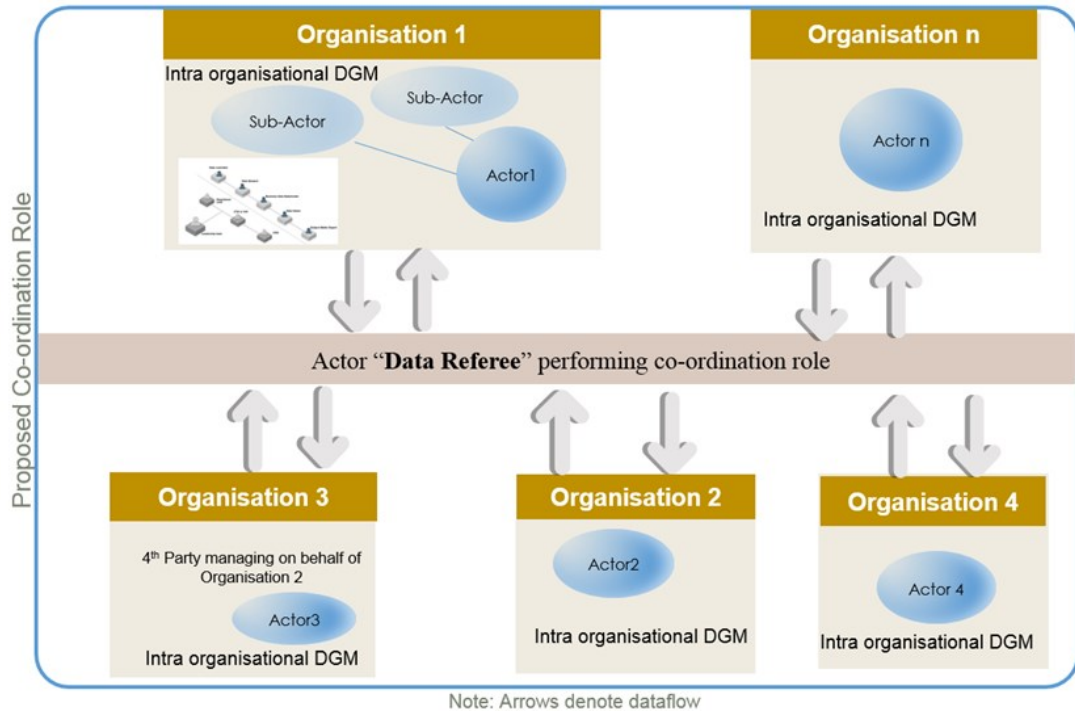


Figure 4-16: Inter-organisational governance operating model

The success of the above operating models are dependent on all actors in the DE. Therefore, it is crucial from the focal firm's perspective that appropriate controls are implemented to mitigate data risks, particularly those caused by external actors. Formal contract-based agreements can play a key role in ensuring the accountability around data processing and usage. This involves confirming that the external actor handling the data has DGM capability. In addition, it involves incorporating data retention, data deletion, data security and privacy requirements in the contractual clauses that the external actor are obligated to perform. Furthermore, including clauses to ensure staff, both external and internal, working with data are trained periodically to understand the guidelines to follow. Finally, a comprehensive assurance mechanism needs to be implemented to reduce risks from external ecosystem actor. This can

include monitoring of data related Service Level Agreements (SLA) through regular reporting and meetings held between dedicated roles such as “data referee” or vendor manager.

4.1.3 Stages of the Framework

The implementation of the DG and management framework in the DE is performed in stages or phases. The proposed 4I Framework is organised into four key stages (Identify, Insulate, Inspect Improve), with each stage performing specific steps. The stages comprises of the elements discussed in the previous section. The framework is flexible which enables it to be used in organisations with varying and unique requirements. The 4I Framework does not require businesses to use all elements, but only those components that are applicable to the firm's scope and context. The steps mentioned in each stage can vary depending on industry, type and size of the organisation and needs to be tailored depending on the context (see Section 5.2)

4.1.3.1 Identify

This stage is the starting point of the 4I Framework and specifies the key drivers, risks, requirements and context encountered from a DG and management standpoint. It includes several elements mentioned in the previous section such as risk, policy, processes, people, data asset and technology. This stage's intention is to lay the foundations for the DG and DM activities. This groundwork is essential to ensure that i) the firm knows the critical data assets and risks associated with it, ii) it engages the right actors and iii) it prepares the guidelines to properly handle data with the right processes.

Table 4-10: Key Steps of the Identify Stage

#	Steps	Output	Actors involved
1	Determine the driver to undertake the Data Management and Governance activities	Business Case Data asset Inventory Guidelines	Internal Actors mentioned in Table 4-8
2	Discover the scope of the data assets	Ownership of data asset	
3	Determine the relevant standards, rules and regulatory obligations and develop the guidelines to realise the objectives	Risk assessment report Security impact assessment Privacy impact assessment	
4	Determine an acceptable risk appetite for the assets		
5	Identify, assess, treat and report on risk including privacy impacts, data impacts as well as risks associated with the external actors		
6	Report on business unit risks and risk management performance		

4.1.3.2 Insulate

This stage covers the preventive control measures taken to alleviate the risks uncovered in the Identify stage. In this stage, the design, development and delivery of the controls are executed in alignment with the guidelines. This includes the deployment of multiple technical tools to help manage the data. This critical stage also includes implementing provisions to ensure and monitor the compliance of the external actors who handle the data. This can include several activities such as i) on-boarding new actors in the DE, based on the evaluation criteria of the firm, ii) negotiating contractual terms and conditions to introduce new clauses to mitigate risk related to

data access, data sharing, erasure of data at the end of the contract, etc. This stage can also look retrospectively into the existing arrangements with external actors and renegotiate contracts if required.

Table 4-11: Key Steps of the Insulate Stage

#	Steps	Output	Actors involved
1	Design a solution approach and the activities needing to be undertaken to realise the objectives	Controls to manage data-risk (at any point of the data lifecycle) and reduce its impact	Internal actors such as organisation's Technical team
2	Build and deploy solutions to realise the objectives	Contract clause changes (to ensure cross-organisation actors are legally bound to handle data.)	External actors such as Vendor manager
3	Conduct security assessment of external actors	Training program (making staff, both new and existing, to be aware of the data-related risks and obligations)	Data referee
4	Revisit and revise contractual obligations with external actors that benefit all actors involved		
5	Create awareness among actors handling the data-on-data management guidelines		

4.1.3.3 Inspect

The inspect stage is the controls-check phase of the framework. It establishes the monitoring process and reviews the progress of all the DG and DM activities. Assessing and reporting on the incorrect execution of processes and non-compliance to the established objectives, policies and standards are the key focus areas of this stage.

Table 4-12: Key steps of the Inspect Stage

#	Steps	Output	Actors involved
1	Perform ongoing monitoring of controls after implementation	Control effectiveness assessment Compliance with agreements assessment Issue and risk logging	Ecosystem actors (both external and internal)
2	Detect any issues, incidents or lapses around data handling guidelines, including changes to consent, non-disclosure agreement, privacy policy associated with data		
3	Escalate and report gaps to stakeholders based on pre-determined metrics and measures		
4	Collaborate and communicate with stakeholders to ensure data processed externally is managed according to the guidelines		
5	Perform periodic audits as well as respond to requests from regulators		
6	Perform checks around critical questions from regulators including those around compliance attestations.		

The necessary toolkits such as audit reports, maturity models and software agents required to periodically monitor, assess and report on the DG and DM maturity are part of the inspect stage.

4.1.3.4 Improve

The final stage Improve is an iterative improvement approach intended to improve the efficacy of the key insulation mechanisms and remove bottlenecks.

Table 4-13: Key steps of the Improve Stage

#	Steps		Output		Actors involved
1	Ensuring appropriate remedial actions are taken, developed and monitored		Corrective and preventive actions to ensure activities meet the stakeholders' goals		Ecosystem actors

4.2 Stages and Elements

Table 4-14 shows an example of mapping the elements in the stages. However, organisations can decide the level of granularity when they operationalise the 4I Framework based on the discretion of the actors involved.

Table 4-14: Sample mapping of the proposed elements of the framework to the stages

Row: Element Column: Stage	STAGES			
	Identify	Insulate	Inspect	Improve
Data asset	Data asset value Data properties	Control mechanisms	Monitor control effectiveness based on metrics and measures Relationship Communication Collaboration	Guidelines Processes Controls Technology Interactions
Data risk	Risk threats Risk vulnerabilities			
Ecosystem Actors	Actor, Activities			
Guidelines	Standards Policies Rules/Regulations			
Processes and Procedures	Procedural mechanisms (derived from guidelines)			
Technology	Applications Data technology Infrastructure			

4.3 Implementation of the framework

This section outlines the steps that needs to be performed to assist stakeholders in implementing the 4I Framework. Few experts who participated in the survey emphasised the benefits of providing the operationalisation steps. The implementation of the 4I Framework comprises of five steps as summarised in Figure 4-17 below and elaborated in Table 4-15 below.



Figure 4-17: Implementation Activities

Sample operationalization of the 4I framework can be broken down as shown in Table 4-15.

Table 4-15: Sample Implementation Steps

Step	4I component / Element: Stage	Proposed Activities	Target State
1	Driver	Prepare the business case	A document that describes the key reasons to undertake DGM activities. The document aims to obtain funding from the organisation's management. This step is a prerequisite to commencing any DGM activities. It outlines the data process or requirements to be met in the design or output of a data process
2.1	Element: Identify	Undertake a review of the applicable legislative obligations	A high-level requirement document that captures and analyses the list of obligations that need to be specifically addressed in response to the legislative mandates such as GPDR, APP, CCPA
2.2	Element: Identify	Undertake a review of existing data assets	An inventory of crucial data assets that are produced, stored, used, shared and disposed of during the course of the data lifecycle journey
2.3	Element: Identify	Undertake a review of ecosystem actors involved in the data processing	A mapping of all the actors, both external and internal to the organisation and their responsibilities.
2.4	Element: Identify	Undertake a risk-based approach to determine the risks linked with the data asset.	A document listing the gaps identified in the existing processes ranked according to impact . This includes a review of the contracts with other contracts,
2.5	Element: Identify	Create guidelines and uplift the existing processes	This involves developing guidelines articulating the recommended approaches to manage the requirements identified in previous steps
3	Element: Insulate	Uplift technology and contract management	A process to implement appropriate controls to mitigate the risks raised in Step 2.4 using the guidelines developed in 2.5
4	Element: Inspect	Monitor controls and report any lapses	A report to track how the control mechanisms are working
5	Element: Improve	Improve the process	Iterative process to improve steps 2-4.

4.4 Chapter Summary

To sum up, this chapter presented the main components of the final version of the 4I Framework. The framework comprises three main components (Drivers, Elements and Stages). Each component has its own focus. The framework derives its name from the first letter of the four stages of the framework. The evaluation of the framework is described in the next chapter.

Chapter 5: Framework Evaluation

The 4I Framework described in Chapter 4 is the final version of the framework, developed through several iterations involving a series of actions. As stated in Chapter 3, the evaluation strategy of the 4I Framework follows the DSR approach adopted from the study by Prat, Comyn-Wattiau & Akoka (2014).

This chapter is divided in three main sections. The first section of this chapter presents the scenario-based testing to demonstrate the applicability of the 4I Framework. The second section of this chapter presents the results of the industry survey that was conducted with data and software professionals from various organisations. The survey data is analysed to arrive at the final evaluation of the 4I Framework. The final section discusses the novelty of the framework.

5.1 Framework evaluation overview

This section presents the high-level assessment of the framework to provide proof of concept of the proposed framework in a real-world scenario. To evaluate the framework, two types of **evaluation techniques** are chosen: scenario-based testing and survey. The results are used to determine the applicability of the framework in meeting the evaluation criteria (Table 3-2) for RSQ4 mentioned in Chapter 3 of this thesis. Following recommendations of Prat, Comyn-Wattiau & Akoka (2014), of the several characteristics available, only six were included in this thesis. The characteristics on which the 4I Framework was assessed (based on the scope of this thesis) are in Table 5-1.

Table 5-1: Evaluation Criteria

Characteristic	Description	Assessed using
Coverage	The aim of this characteristic is to evaluate if the framework included sufficient components required to implement DG and DM	Survey
Generality	The aim of characteristic is to assess how well the framework supported different businesses spread across different domains, as well as how easy it is to tailor the 4I Framework to different environments and circumstances	Scenario
Clarity	The aim of this characteristic is to appraise whether the 4I Framework is explained clearly and how easy it is to understand the framework	Survey
Efficacy	The aim of the efficacy characteristic is to determine the degree to which the artefact produced meets its desired goal of managing DG and DM activities in the context of DE	Scenario and survey
Importance and Relevance	The aim of this characteristic is to determine whether the elements included in the framework are relevant and important to address the DG and DM problems and challenges facing an organisation.	Survey
Novelty	This characteristic was aimed at demonstrating the Framework's novelty (new knowledge)	Literature gap review

The assessment was conducted using the following approaches:

a) **Descriptive Illustrative Scenario-based testing:**

Three scenarios based on recent data breach events that occurred in the retail, healthcare and supply chain sector were chosen to demonstrate the effectiveness of the framework by showing that the real-world problems can be solved. As shown in Figure 5-1, the characteristics evaluated through scenarios were generalisation and efficacy. The scenarios are discussed in detail in Section 5.3 of this chapter.

Evaluation Criteria	Description	Scenario
Generality	Framework is not tied to one domain or type of organisation and can be tailored to similar contexts	Scenario1 Scenario 2 Scenario 3
Efficacy	Framework is able to meet the goal of managing DG and DM activities	

Figure 5-1: Criteria evaluated using scenarios

b) **Question-Based Empirical Survey:**

An industry survey was given to a cohort of 30 professionals from local and international companies. 30 is considered a reasonable representative sample size for this kind of research (Anwar 2021). The survey was offered to participants in a Google form format. The actual wording of the survey can be found in Appendix A. The data was collected anonymously to help provide proof of concept for the 4I Framework and to determine its relevance from the experts' point of view. The criteria evaluated using the survey are listed in Figure 5-2 and elaborated in Section 5.3.

Evaluation Criteria	Description	Survey Question
Coverage	Framework includes sufficient components required to implement DG and DM	Question 3
Clarity	Framework is clearly explained and easy to understand	Question 6
Relevance and Important	Framework and the components included in it are relevant and important to address the DG and DM challenges	Question 1, 2, 3, 4, 5, 7
Efficacy	Framework is able to meet the goal of managing DG and DM activities	Question 9

Figure 5-2: Criteria evaluated using surveys

The novelty of the framework is justified through the literature review of the related work and the research gaps outlined in Chapter 1 and 2.

5.2 Scenario-based testing demonstration

To assess the suitability of the framework proposed in Chapter 4, three scenarios involving the mishandling of data that lead to data breaches are considered.

Table 5-2: Mapping of Scenarios to Evaluation Problems

#	Scenario	Real life DGM issue	Criteria Evaluated	4I Framework components involved
1	A Europe-based car company needs to on-board the details of a new car buyer. Driving license data is captured as part of this activity. Abiding by regulatory requirements related to data handling and sharing is of the utmost importance to the car company.	Non-compliance concerns (Anderson 2017)	Generality Efficacy	Elements: All
				Phases: All
2	A wearable device manufacturing company needs an approach to safeguard customers' health data and abide by HIPPA laws.	Mishandling of data (Banerjee, Hemphill & Longstreet 2018; tegan)	Generality Efficacy	Elements: Data-Asset Data-Risk Technology Guidelines Technology Ecosystem Actors
				Phases Used: All
3	A smart fridge collects information about the food habits of the user. To the service provider of the smart fridge, handling a customer's private data is of the utmost importance.	Mis-share of data (NBC 2015)	Generality Efficacy	Elements used: Data Asset Data Risk Guidelines Technology
				Phases Used: Identify Insulate

5.2.1 Use Case: Application of the 4I Framework to support regulations

5.2.1.1 Scenario Overview

Car vendor's business process involves on-boarding new customers who purchase a vehicle. The purchase process captures information pertaining to the driving license of the buyer. During the course of customer onboarding and after sales support, various systems (Ordering, Warehousing, Inventory and Transportation) and suppliers are involved. The activities performed include capturing, storing, processing and deleting data.

5.2.1.2 Need

The car vendor needs to ensure that they are compliant to the latest regulations when it comes to managing sensitive data.

5.2.1.3 Evaluation Objective

The objective is to demonstrate the applicability of the framework to guide the firm to meet the requirements of the regulators and minimize financial loss, if any.

5.2.1.4 Execution of the 4I Framework

This section describes how the 4I Framework can support the stakeholder's compliance readiness activities.

Activity 1: Identify the key driver to undertake DGM activities

The requirements of the car organisation is to ensure that they understand the requirements. Therefore, key driver to engage in DGM is managing compliance obligations.

Activity 2: Identify the key elements

Keeping in mind the fact that the fundamental driver is to meet the regulatory obligations in terms of data compliance, the next steps are to determine the key elements including what data is in scope, what are the risks associated and who are involved.

Activity 2.1: Data asset

The organisation captures customers details which includes:

- Name
- Contact Details
- Date of Birth (dob)
- Digital signatures
- Driving License
- Payment Card Information (PCI)
- Financial Invoices

To ensure the proper governance and management and end-to-end visibility across all stages of the data lifecycle, it is essential to create a data lineage diagram showing flow of through several systems. In addition, metadata such as creation timestamps, data retention duration, the source, author and, classification of the data based on its sensitivity needs to be captured.

Activity 2.2: Data risk

The organisation wants to avoid repeating the mistakes made by a firm operating in the airlines sectors (Section 2.2). Information Commissioner's Office fined British Airways 20 million pounds for failing to follow GDPR and manage data appropriately. The root cause of the incident was inadequate of governance and oversight when using third party developed module or plugin. In this scenario, the main risk that the car

vendor wants to eliminate is any regulatory compliance breach caused by the external ecosystem actor (KPMG 2021).

Activity2.3: Guidelines

A key step is to understand which legislations the car vendor needs to comply with. An important task in this regard is to ascertain what needs to be done to achieve compliance, which regulations needs to be studied. In this instance, since the car firm was from Europe, GDPR was found to be the relevant regulations.

The table below outlines the GDPR requirements and the corresponding activities which need to be performed to identify the requirements.

Table 5-3: Key obligations identified at the Identify stage

Obligation	Activities
Find the legal requirements	Capture the legal obligations through discussion with the compliance team and identify the cost benefit analysis if the requirements are not met.
GDPR Article 5.1 “Purpose limitation” to ensure only data necessary to perform business is collected	Capture data collected through various interfaces. If possible, prepare lineage diagrams related to the flow of personal data throughout its lifecycle across the supply chain systems (Banakar et al. 2019)
GDPR Article 3.7 on “Designation of the data protection officer”	Set up governance charter that includes representatives of Data Protection DG team (Bastos et al.)
GDPR Article 30 “Records of processing activities” of GDPR	Gain information on relevant stakeholders in the car supply chain who has permission to access data including the external actors who process personal data (Collibra 2018)
GDPR Article 5.2 “Principles relating to processing of personal data”	Co-ordinate with the legal team to consolidate and analyse all <ul style="list-style-type: none"> • contracts, • third-party actor selection criteria • Statement of work, • Standard operating procedure involved in data exchange
GDPR article 32 “Security of data” GDPR article 33 “Notification of a personal data breach to the supervisory authority”	Identify and document <ul style="list-style-type: none"> • inventory of all infrastructure devices • inventory of software used with version number (including Cloud hosted solutions) • Review and update • security vulnerability monitoring rules • breach notification guidelines
Enable compliance related to GDPR article 20 “Right to data portability” and article 17 “Right to be forgotten”	Review and update storage compliance policies for data Data retention rules (Vargas et al. 2018) data sharing principles (Security 2018))

Protect data as required in GDPR article 32 “Security of data”	Frame the access management principles to design the security of data
Article 32 and article 30 “records of processing activity”	Prepare a template to capture and store information on <ul style="list-style-type: none"> • consent collections • data cleansing, • data sharing • log data management

For this scenario, GDPR was considered as an example. However, other regulations like APP might be applicable in case the car vendor is located geographically in other continent.

Activity 2.4: Ecosystem Actors

A car supply chain comprises multiple stakeholders such as the manufacturer, parts supplier, sales representatives, brokers, agents, dealers, IT service providers, transport authorities, cloud provider, car insurance provider and the customer. The process of data collection to gain deep business insights using analytics is highly complex, with several players working jointly to achieve overall performance, and personal information of the customer being shared at multiple steps. Certain activities are carried out through outsourcing and offshoring arrangements that involves third-party service -providers. The firm’s governance structure has characteristics of federated and centralised data governance mode. The compliance to GDPR is the responsibility of the car manufacturer. Therefore, the centralised governance model is a natural choice in this scenario. Alternatively, federated mode can be used where the “data referee” plays a key role in ensuring right activities are executed to ensure compliance.

Activity 3: Appropriate Controls

After identifying the key elements, the focus shifts to insulating the risk identified in the Identify stage. For instance, to align the data initiatives, legal contracts such as contract clauses and statement of work (SoW) can be altered to impose financial

penalties for external suppliers for data privacy violations. Once a formal agreement is reached, there must be an agreement on what data to share and who can access the data. Additionally, SoW can incorporate certification needs such as the Payment Card Information Data Security Standard (PCI-DSS) (Yulianto, Lim & Soewito 2016). Each of vendor partners responsibility, accountability and boundaries can be laid out in the Service Level Agreement (SLA). Moreover, SLA can mention provision to regularly the assess ecosystem player's capability to support in terms of its financial health, subcontractor clauses, breach notification and data management practices.

Table 5-4: Key activities to monitor risk from external actors

Activity	Action
Assess risk associated with ecosystem actors	<ul style="list-style-type: none"> • Evaluate supplier reputation • Evaluate supplier technology capability • Evaluate sub-contractor management (such as legal agreements) • Evaluate external political environment • Evaluate supplier's employee background check process • Evaluate supplier's ability to track data incidents • Evaluate supplier's ability to follow the guidance and provide assurance or attestation to DM activities DM created as part of Activity 2 above

From a technical perspective, controls in the form of automated batch jobs can be used to delete stored data using data retention rules such as changes in the ownership details of the car's vehicle identification number (VIN). Furthermore, techniques such as pseudonymization, obfuscation, hashing and encryption (Romanou 2018) can be used to further insulate the data from data privacy breaches by limiting the sensitive data shared between organisations and conform to GDPR article 25 on "data protection by design and by default".

Activity 4: Monitor the existing systems and processes

Next, we move to the Inspect phase where quality control checks are performed to ensure that the data from the supplier, product, vehicle owner and suppliers are accurate. To safeguard the privacy of the customer, it can include i) software agents to ascertain if non-authorized personnel accessed the data classified as confidential, or ii) ensure communication from the sensors of autonomous vehicles to cloud servers abides by data sharing rules. In the event of potential unethical usage of data from systems, the audit trail can be shared to notify data breaches and comply with GDPR article 31 on “cooperation with the supervisory authority”.

As mentioned earlier, the framework does not enforce any specific tools or technologies to be used. The organisation using the framework can implement solutions that are deemed suitable.

Activity 5: Continual Improvement

Finally, we move to the Improve phase where opportunities for continuous improvements are identified, assessed and actioned. This will ensure that the 4I Framework supports any adaptation needs within the overall context of DG. For instance, the additional new data minimization policy outlined in the Global Automakers Consumer policy (Akalu 2018) that limits the collection of personal information to a bare minimum can be implemented. This can be supplemented by performing assessments on cloud service providers using the Consensus Assessments Initiative Questionnaire (CAIQ) and Cloud Controls Matrix (CCM) (Sen & Madria 2018) developed by Cloud Security Alliances (CSA), an organisation that promotes the use of best security practices in cloud computing.

This example provides a perspective on the role of the 4I Framework to support the regulatory obligations of a car vendor participating in the Digital Automobile Ecosystem.

5.2.2. Use Case: Application of the 4I Framework to support data protection

5.2.2.1 Scenario Overview

Wearables were one of the main fitness trends for 2019 (Thompson 2018) according to the American College of Sports Medicine (ACSM). Presently, there are greater than 28 million users of Fitbit (Fitbit 2019). Wearables collect a consumer's private and personal data such as heart rate, daily steps etc. The device manufacturer engages different software vendors for the development, support and monitoring of the mobile applications associated with the device.

5.2.2.2 Need

The device manufacturer recognises its obligation to keep the personal health data it collects, captures, stores and shares with other ecosystem actors secure. The key requirement is to safeguard the mishandling, misused, or mis-share.

5.2.2.3 Evaluation Objective

The objective is to demonstrate the applicability of the framework to facilitate the data protection activities that involves several external actors. The aim is to provide the Wearable service providers with guidance on how the framework can be leveraged to avoid the unethical usage of data.

5.2.2.4 Execution of the 4I Framework

This section describes how the 4I Framework can support the retail wearable sector.

Activity 1: Identify the key driver to undertake DGM activities

The requirements of the organisation dealing with wearables is to ensure the manufacturer of the wearable device and associated services safeguard the stored

customer data from misuse. Hence, **data protection** obligations are a key requirement for any player in the DE handling a wearable device.

Activity 2: Identify the key elements

Given that the main driving factor is ensuring data protection, the next step is to determine the key elements that need to be protected, including what data is in scope, the associated risks, and who is involved.

Activity 2.1: Data asset

The following personal identifiable information (PII) is exchanged between the IoT-enabled wearable device and mobile application and the DE players.

- Name
- Address
- Health Data
- Contact Details
- Date of Birth

Activity 2.2: Ecosystem Actors

Transmission of the wearable device data occurs using Bluetooth technology. As illustrated in Figure 5-3, data is initially sent to the user's desktop or mobile application, before it is transmitted to the cloud. Several players are involved in the data exchange such as the device manufacturer, the software application provider, hosting provider, health service provider (HSP) and numerous IT systems to deliver the end user with a health service.

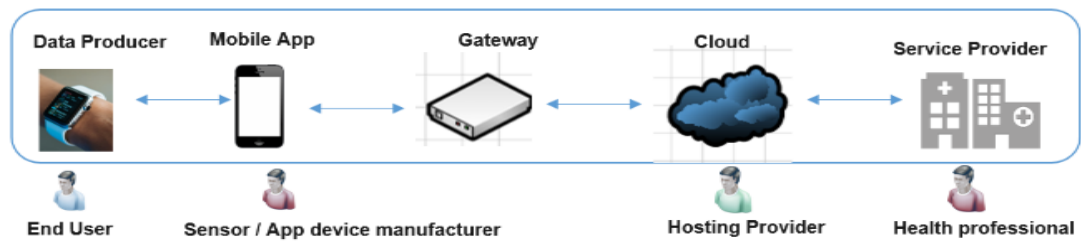


Figure 5-3: Wearable smart IoT-enabled ecosystem

Activity 2.3: Data risk

In 2019, nearly 41 million healthcare records were compromised with majority of them directly linked to the illegal disclosure of data by third parties. For instance, the athletic wear company *Under Armour* disclosed that data records tied to its fitness app was breached, affecting 150 million user accounts containing user passwords, account reset codes, smart camera recorded conversations (Kuhn 2018). In a few instances, the hackers accessed contact details and attempted to use the compromised data to break into other accounts. Therefore, it is critical for the focal wearable service provider to avoid similar incidents and ensure that data is protected.

Activity 2.4: Guidelines

To ensure the customer data is not compromised, the focal firm needs to review the laws (see Figure 5-4), particularly those related to healthcare such as HIPPA (Sharma, Chen & Sheth 2018) with regards to data protection. Furthermore, depending on the jurisdiction, the rights of the smart health device user need to be understood before formulating the rules of engagement with the external players.

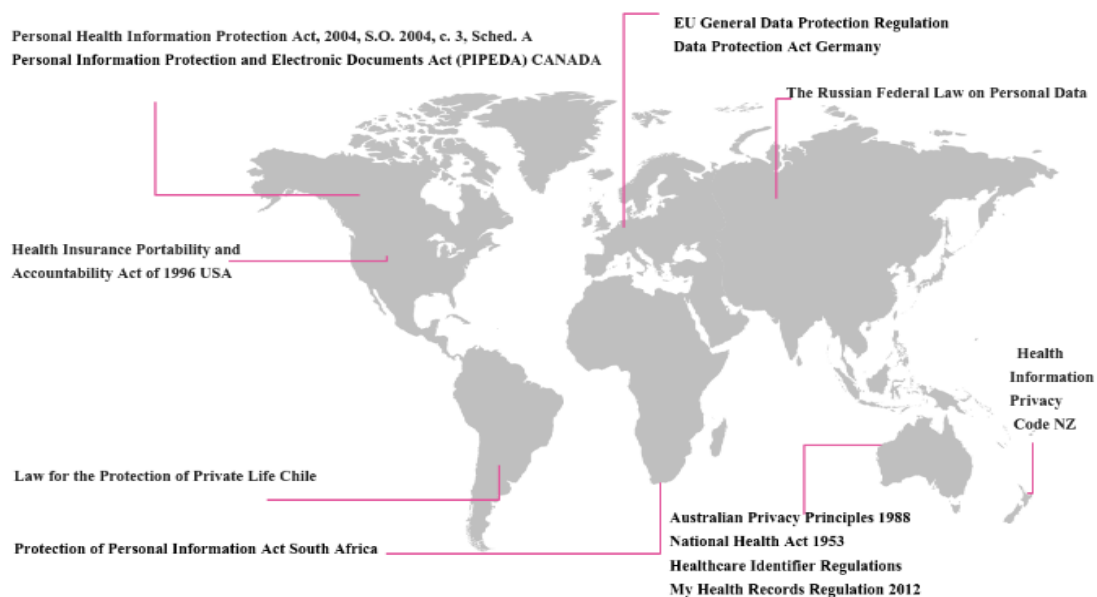


Figure 5-4: Laws related to health data

As mentioned above, the exploitation of security vulnerabilities are a phenomenon common to the wearable app industry where mobile applications running on Android are impacted by vulnerabilities from time to time (Linares-Vásquez, Bavota & Escobar-Velásquez 2017). Android Security Bulletin publishes details of security vulnerabilities affecting Android devices. This publicly available information can be included in the patch management guideline.

Activity 3: Appropriate Controls

The second stage Insulate includes the preventive measures taken to mitigate the risks identified in the previous stage. Technology and cyber controls can be used to implement the data protection policies related to healthcare equipments. For instance, this can include preventive measures such as ensuring software patching is carried out in accordance with the patch management policy or password is forcibly updated in a

periodic basis. As mentioned previously, the framework does not prescribe any fixed approach and suggest that firm's use mechanisms that address the requirement.

Activity 4: Monitor the existing systems and processes

The inspection phase is a made up of real-time monitoring, auditing and reporting performed manually or through the use of existing software agents available in the market that can ensure protection of the IoT devices. This can include initially mapping each wearable device and data it collects to the application and database server. Next, Security Information and Event Management (SIEM) or Data Loss Prevention (DLP) agents can be installed in the devices or network endpoints. The agents raise automated alerts to the relevant stakeholder whenever any non-encrypted record is detected.

Activity 5: Continual Improvement

In the improve phase, the effectiveness of the processes and procedures are investigated, analysed and enhanced. For example, activities such as performing an annual review of third-party SaaS providers cloud storage vulnerability or updating the ecosystem partner service level agreement policy can be conducted. Additionally, rewording the contracts to pass the liability of a data breach to the third party can be another outcome of the improvement phase.

In this example, the 4I Framework outlines the systematic approach to help in addressing the challenges arising due to inadequate governance and management of data within the retail wearable sector.

5.2.3 Use Case: Application of the 4I Framework to support ethical data usage

5.2.3.1 Scenario Overview

Home automation and smart homes are one of the emerging industry sectors in today's world. Television sets, thermostats, self-operating vacuum cleaners, refrigerators, home security, cleaning and maintenance devices are common examples of connected appliances. Motion and light sensors are also grouped into this category. The mode of operation of smart appliances involves collecting data that may contain private and sensitive information. For example, by tracking a person's food preference, searching and ordering history of food from online stores, user preference traits can be discovered. The type of data involved includes location and address, network access and connectivity-related information, photos and videos.

5.2.3.2 Need

Consumers are required to consent to the processing of data when enrolled to smart fridge services. The manufacture of the fridge needs to ensure that customer's personal data is not shared to other actors without consent from customer.

5.2.3.3 Evaluation Objective

The objective is to demonstrate the applicability of the framework to enforce data usage governance thorough consent mechanisms.

5.2.3.4 Execution of the 4I Framework

The following steps illustrate the use of the 4I Framework in IoT-enabled homes using a smart refrigerator.

Activity 1: Identify the key drivers to undertake the DGM activity

The requirements of the organisation dealing with smart fridges is similar to the case of the wearable device discussed in the previous section. The trust and confidence of

end user needs to be ensured. Hence, in an organisational context, data protection is the main driver that provides justification to conduct DGM activities.

Activity 2: Identifying the key elements

Activity 2.1: Data asset

A smart refrigerator functionality includes sending out information such as the personal food habits of the user. In some newer versions of smart fridges, credit card information is also stored to order food directly from the fridge. Thus, PI and financial records make up the data asset in this scenario.

Activity 2.2 Data risk

The fridge manufacturer wants to avoid incident similar to that data trust incident reported against the HealthEngine app (Table 2-4). In that instance, HealthEngine app provided contact information of 135000 patients to private health insurance brokers for a fee without adequately disclosing to consumers.

In this example, the refrigerator owners' eating habits can be inferred from the search queries. The records are distributed to third-party businesses for the purpose of sending targeted advertisements to consumers. In the absence of explicit consent obtained from the user, this may result in an ungoverned collection of data and legal issues.

Activity 2.3 Guidelines

One of the activities that the manufacturer of the smart fridge can take is the establishment of a guideline detailing the consent. The consent collection form can be created and included in the terms and conditions (T&C) or privacy policy (PP) related to the features of products or services (Banerjee, Hemphill & Longstreet 2018; Skierka 2018).

Activity 3: Appropriate Controls

Different cyber security and technical controls can be deployed to ensure inappropriate data exchange does not occur. For example, as part of initial setup or registration of the device, the users will have the option to accept the data sharing arrangements through a software agent included in the web interface of the product. This interface stores and saves the accepted data transmission rules. Furthermore, formal agreements can be arranged with the ecosystem partner to include “non-disclosure of customer data” clauses, or sharing of anonymized data.



Figure 5-5: Data Collection Consent

Based on the settings, a smart refrigerator can do several things such as i) transmit data to a cloud-based consent setting, ii) transmit only to a preconfigured address or iii) transfer data only if the firmware version of the device is up to date. The exact control may vary from one organisation to another.

Activity 4: Monitor the existing systems and processes

The aim of this activity in the Inspection stage is to ensure that the organisation can capture events and is able to detect and correct any issues effectively. For example, the agent can be equipped with logging and monitoring capabilities to trigger alerts in the case of abnormal patterns in the network traffic. The data can be correlated across other events and reported periodically. Additional monitoring tools like FireEye and Tanium can be deployed to assist in the inspection activities. As stated earlier, the 4I

Framework is technology agnostic and is not fixed on any specific technology stack or tool.

From a governance perspective, smart fridge manufacturer can seek evidence from external party handling the refrigerator owners' data confirming that data is not reused inappropriately.

Activity 5: Continual Improvement

Similar to the previous two cases, continuous improvement initiatives can involve keeping check on the changing data requirements which is critical.

In this example, the 4I Framework outlines the supporting controls to ensure that the company ensures data use and reuse is not permitted without consent. The controls rely on using a combination of formalised contractual clauses, guidelines and configuration management techniques.

5.3 Empirical Evaluation: Survey

5.3.1 Survey plan

The intention of the survey was to obtain experts' feedback and opinions on the 4I Framework. The survey gathered both qualitative and quantitative data from the participants. The survey data analysis is used to determine whether the 4I Framework meets the evaluation criteria (see Chapter 3, Table 3-4)

5.3.2 Design of the survey

The study focused on capturing collective feedback and recommendations about implementing DG and management practices from expert professionals from the ICT industry. Initially, a questionnaire using Microsoft Excel was piloted with three participants to test for clarity and applicability. This enabled us to validate the questionnaire by rewording and deleting ambiguous statements. Based on the feedback with regard to the ease of use, the Microsoft Excel based questionnaire was replaced by an online survey constructed using survey design principles based on well-established guidelines by Hyndman (2008).

5.3.3 Survey procedure

The main source of data was the anonymous online surveys and one-on-one interviews. The data collection period was from March 2019 to January 2021. The interviews tracked various objectives including determining the importance, usefulness and practicality of the framework as well as suggestions to modify the framework. It allowed the participants to provide ratings (1-5) to convey their agreement, disagreement in relation to the framework components. The survey had options for the respondents to provide feedback or suggestions. This allowed me to

understand and gauge the effectiveness of the 4I Framework and to obtain valuable input on the crucial attributes which could be added to the framework.

The survey was given to the participants and experts in the IT industry who specialise in DG and DM. The participants were recruited through relevant online LinkedIn groups as well as my UTS supervisor's industry contacts. All participants were contacted via email relaying information about the purpose of the study and its agenda and the consent form. The invitation letter was approved by the UTS Human Research Ethics Committee (HREC) before it was sent to the participants in accordance with the ethics approval outlined in Appendix A. A follow-up email was sent to those who agreed to participate in the study with the survey form link (see Appendix B). Few individuals preferred a face-to-face session or a telephonic conversation and the information was keyed onto the form by the author.

The survey did not collect any PI information from the participants. The original signed consent forms are stored in CloudStor (see Appendix E).

5.3.4 Survey respondent profile

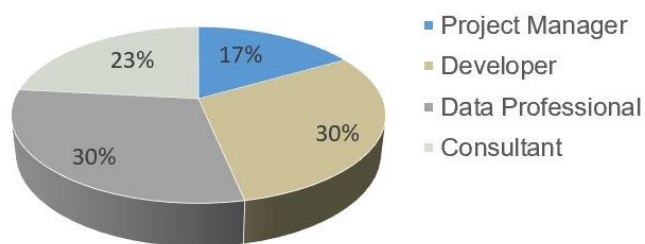


Figure 5-6: Respondent designation profile

Figure 5-6 reports the respondents' positions in their organisation. By design, most of the respondents had an average minimum 10 years of experience. This is an indicator of the expertise of the participants in the IT industry. The minority of the participants

worked as a project manager. Of the 30 respondents, five worked as a project manager while nine were employed as a data professional either in a role as a DG coach, DG strategist or as data consultant. Seven consultants worked in the role of a security professional, business analyst, solution architect or as a cloud consultant. The remaining nine respondents were solution or technical architects, software engineers or developers and are included in the category of developer in Figure 5-6.

As shown in the demographic representation pie chart of the participants, the majority of respondents' organisation (60 percent) are headquartered in Australia, 26 percent have headquarters in Europe (Germany, Denmark, Spain and Netherlands) and 14 percent are located in Asia.

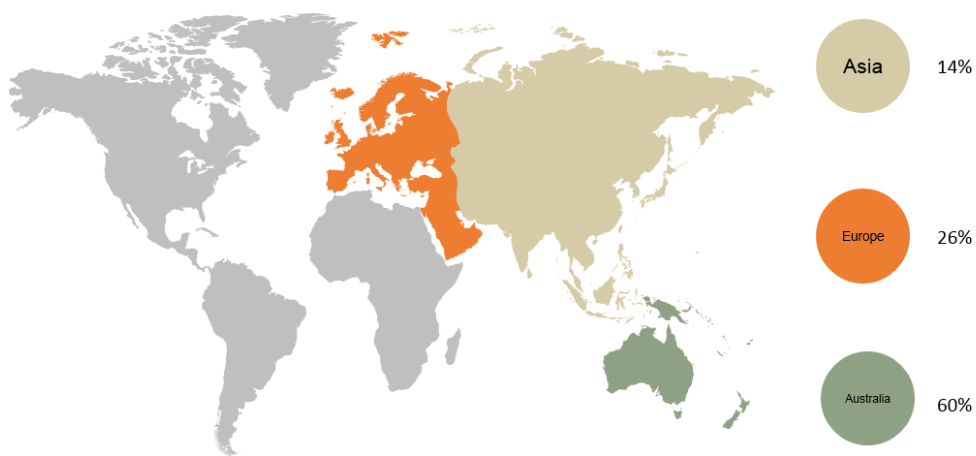


Figure 5-7: Respondents' demographic profile

The participants comprised practitioners employed in both the IT and business side of organisation. This diversity allowed the study to obtain a comprehensive synthesized finding from the different stakeholders in the DGM landscape.

5.3.5 Survey questions

To evaluate the 4I Framework, a pilot test was conducted using a spreadsheet-based questionnaire and survey with a representative sample of three respondents. The purpose of the pilot test was to assess if the instructions were adequate and understandable. The feedback of the pilot resulted in the revision of the format from the spreadsheet-based survey to a Google form-based survey.

The survey comprised nine questions and took approximately 20-30 minutes to complete. The respondents were asked to review and rate the different elements of the framework. The questions (refer to Appendix B) used for this study focussed on

- Stages of the framework
- Individual elements of the framework
- Efficacy of the framework
- Suggestions to improve the framework
- Overall feedback and rating of the framework

Qualitative feedback from the participants played a significant role in the iterative development and evaluation of the framework. The validation exercise not only resulted in suggested improvements to the framework, it also served to obtain empirical evidence supporting subsequent refinements to the framework design based on experts' views of DGM, regulatory policies and future strategies and policies.

5.3.6 Survey data collection

The data collected from the survey can be categorised in two groups:

- a) **Qualitative** (participants' feedback)
- b) **Quantitative** (rating data transformed into ordinal data)

Qualitative feedback from the participants played a crucial role to identify the concerns and provided the data required to iteratively develop and evaluate the framework.

5.3.7 Survey data analysis

Thirty experts from the industry reviewed the framework. The nature of this survey is hybrid, with both quantitative and qualitative data collected. Participants were encouraged to provide qualitative comments and suggestions that would guide the future enhancement of the framework. The open-ended questions provided the flexibility to apply a qualitative analytic method for analysis. The evaluation included:

- a) *Individual evaluation*: This included the questions related to the individual elements of the 4I Framework. It was based on the question sets 1-5.
- b) *Overall evaluation*: This included the questions related to the overall feedback on the framework. It was based on question sets 6-9.

By and large, the feedback returned from the survey indicated that the respondents had positive opinions about the usability of the approach with some of the comments providing input to support the subsequent refinements to the framework design. Most of the reviewers favoured the 4I Framework and agreed that 4I is acceptable, practical and provides guidance on how to implement DGM activities. Importantly, the respondents provided valuable comments on the drawbacks of the framework and offered several recommendations to enhance the framework. On numerous occasions,

the feedback resulted in modifications to the final 4I Framework artefact. The comments received were captured primarily from the answers given to questions 3, 8 and 9.

In the following sections, the analysis of the findings based on EC is discussed in detail.

EC Criteria 1: Clarity

The assessment of the EC “Clarity” was based on the responses to question 6, to what extent is the framework clear and easy to understand. The overall responses of the reviewers to this question was positive. 17 of the 30 respondents (56%) commented that they were able to comprehend the framework. Several comments such as “*The framework is easy to grasp for a person working on data*”, “*Understandable*”, “*Yes, I can make it out*” demonstrate this point. Three respondents working in the DG office suggested further simplification of the framework. One wrote, “*In my point of view, framework should be more simpler which could be implementable and workable for any organisation*”. The second wrote, “*The framework should be accompanied with instructions specifying how to use it, what are the outputs of each stage*”. A third one’s comment was “*Simplify the framework so that it could be easily implementable*”. These comments were fed back into the iterative development when designing the final version of the framework. Table 5-7 summarises how all the responses were incorporated in the final version of the framework.

Overall, based on the high (> 50%) affirmative responses, it can be concluded that the 4I Framework is reasonably clear.

EC Criteria 2: Relevance and Importance

Several questions were asked to ascertain whether the 4I Framework contained all the relevant and important components required to implement DG and DM. Question 1 asked about the overall four-staged approach whereas Questions 2-5 focussed on the details of the stages. Table 5-5 presents the results of the quantitative ratings received from the completed surveys which provides an indication of the reviewers' opinions on the stages of the 4I Framework.

Table 5-5: Statistical Results on the stages from the Empirical Survey

Survey Item	% Agree
Presenting the four-staged approach (Question 1)	80
Presenting the DG and DM related activities in the Identify stage (Question 2)	80
Presenting the DG and DM related technical control activities in the Insulate stage (Question 4)	70
Presenting the DG and DM related non-technical control activities in the Insulate stage (Question 5)	80

From Table 5-5, it is clear that most of the experts agreed on the stages. However, when it came to the elements and sub-elements of the Identify stage, the responses varied. Table 5-6 shows the statistical results of the survey.

Table 5-6: Statistical Results on the stages from the Empirical Survey

Element	Mean	Standard Deviation	Skewness	Kurtosis
Data	3.86	0.8	-0.46378	0.034431
Risk	3.93	0.9	-0.08861	-0.82253
Compliance (included as part of Guidelines in the final version of 4I)	3.82	1.0	-0.36548	-0.52578
Policy	3.71	0.8	-0.63959	0.502926
Process	3.61	1.0	-0.33076	0.64344
Technology	3.68	1.0	0.147937	1.05953
People (renamed "Ecosystem Actors" in the final version)	4.07	0.9	0.17175	-1.73695

It was observed that all the values of skewness are in the range (-1, 1). According to Hair (2009), this is considered normal and acceptable. The values of kurtosis are in the range (-2,2) which, according to Hair et al. (2014) lies well within the range.

Most of the reviewers made positive comments on the relevance and importance of the 4I Framework. It is interesting to note that one participant objected to the “IT focus” and inclusion of DM principles. The respondent stated that “*your scope is too far wide and includes other DM disciplines*”. Another participant’s opinion was that the “*relevance of an element will depend on the business and strategic objectives of the organisation*”.

The overall summative assessment of this category is considered satisfactory based on the positive feedback received.

EC Criteria 3: Coverage

The aim of the assessment category “Coverage” was to determine if the framework included all the relevant elements needed to manage and govern data in the DE. The survey did not include any quantitative questions on the coverage criteria and relied on the participant’s comments received in the qualitative section. The majority of the common responses were “*It seems to cover all necessary fields*”, “*Looks satisfactory*” or a variant of these two comments. Valuable input was provided on different DG and DM-related topics that assisted in the expansion of the number of sub-elements included in the final version of the framework. The comments were reflective of the participant’s in-depth experience on a particular topic. For example, a data consultant mentioned “*More visibility needs to be included in the framework around data deletion*

/erasure” whereas another data engineer recommended the inclusion of topics such as *“Data masking policy”*, *“Report Management”* topics. The list of all the suggestions received and the approaches taken to address them are discussed in Table 5-7.

Overall, given the high percentage of positive comments, it can be concluded that the 4I Framework’s coverage is sufficient.

EC Criteria 4: Efficacy

Participants rated the practicality of the 4I Framework positively. The feedback received on the overall rating (Question 9) is depicted in Figure 5-8.

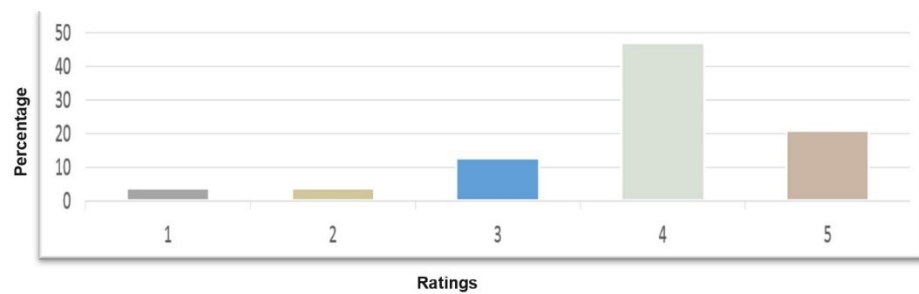


Figure 5-8: Overall rating of the framework

A total of 70% of the experts reported satisfactory feedback with only 30% giving a rating of 3 or below.

This gives an impression that in general, the 4I Framework is considered to be practical to implement DG and DM activities in the DE context. Therefore, the overall validation can be considered satisfactory.

5.3.8 4I Framework enhancement (based on feedback)

The responses received in the survey resulted in some changes to the initial version of the framework. Table 5-7 captures the key suggestions and the approach taken to address them, post examination of all the comments received.

Table 5-7: Suggested changes and gap mitigation approach

#	Suggestion Received	Mitigation Approach
1	<i>DG is not a one size fits all prescription, but it is a more workable framework/ solution which works for any organisation for effective data management. From my point of view, it is very detailed technically but it lacks the business side of things like how you start working on the DG initiative e.g., alignment with the corporate strategy. Every organisation has different data-related challenges, some have data quality issues, some have data but do not know what to do with that data etc. From my point of view, the framework should be simpler so that it is implementable and workable for any organisation.</i>	<p>The framework was updated to reflect the business view through the introduction of the Drivers component. The relationship among the different business objectives, stages, elements, sub-elements is elaborated in Chapter 4(Section 4.1.1)</p> <p>The user or consumer of the 4I Framework can now drill down based on the requisite granularity needed depending on the context.</p> <p>The applicability of the 4I has been demonstrated through the scenario-based evaluations to show it is workable for organisations of diverse backgrounds.</p>
2	<i>Roles differ from company to company and it is not possible to have an exhaustive list - you would be better to list the responsibilities than the role title</i>	Attention was given to incorporate the suggestion in the final version of the framework. The responsibilities associated with the roles are now included in detail. (Table 4-8)
3	<i>There is a merger of compliance requirements and controls in the framework. The implementation will require substantial funding that is not captured in the framework</i>	The Driver component was introduced to incorporate this feedback (Section 4.1.1)
4	<i>More visibility needs to be included in the framework around data deletion /erasure</i>	Data deletion related concept was already included in the 4I Framework as an integral part of the guidelines (Section 4.1.2). However, the details of this topic were beyond the scope of this thesis.
	<i>There are ecosystem players like translation services who handle 'data' on behalf of the organisation. Similarly, subscription services and loyalty programs also collect data. The people category should include them.</i>	The 4I Framework is adaptable. Organisations can adjust the people element of the 4I Framework to include new players, as appropriate (Section 4.1.2)
5	<p><i>Simplify the framework so that it is easily implementable.</i></p> <p><i>Framework is important but I feel that this one is too complex to be useful.</i></p>	The structure of the framework and the associated thesis documentation on the 4I Framework was subsequently modified to take into consideration the feedback provided by the evaluators.

6	<i>Add something on data protection in 3rd-party Clouds, software including infrastructure vulnerability</i>	The intent of this thesis was not to evaluate the details of data protection. However, implementing controls to protect data through existing established practices such as were briefly mentioned in this thesis and highlighted in Section 5.2 (scenario-based examples)
7	<i>Add a section on approved storage requirements</i>	The concept suggested by the reviewer is now included as a sub-element in the guidelines.
8	<i>Include other industry laws like Association for Retail Technology Standards</i> <i>These are primarily relating to Financial Services - what about other industries?</i>	The industry laws (GDPR, APP) cited in the version 1 of the 4I Framework were for illustrative purposes. As part of identify phase of the framework, organisations need to find the legislations applicable to them.
10	<i>I think it is too complicated to be implementable</i>	The final version of the framework was modified to make it more usable. Refer to figure 5-3 for the step –by –step execution activities as a guidance to implement this framework
11	<i>Covers a lot of other data management disciplines</i>	With regards to the DM disciplines included in the framework, the assessor may not have fully understood the stages from the evaluation documentation. Chapter 2 was updated to clarify the DM and DG.
12	<i>Very detailed and technical and it does not give an overall picture or the important components linked to the business. I would include a "DG Checklist" i.e. assessment of current state of the organisation before taking any steps in identity. Followed by a maturity model to show where the organisation is in terms of maturity and then devise the next steps.</i>	The high-level activity list has been added in Table 5-3. The maturity assessment is beyond the scope of this thesis.

To gain a better understanding and insight from the qualitative data for the individual components of the 4I Framework, a thematic network analysis approach was used to structure the outputs. The thematic analysis approach comprised of three steps, a) screening data, b) categorising recurring data in the answers into common patterns and c) interpreting a text (in this case the qualitative text). This study focussed on three

levels of themes: basic theme, organizing theme, and global theme. Initially, the basic or lower order theme was construed from the full text of the answers provided by the respondents by selecting appropriate codes. For example, there were two comments from two different stakeholders about access management and cyber security. Both these comments refer to security. Hence, these were combined and referred to as the concept of data protection. This helped in establishing the basic themes. In total, nine concepts were identified which are listed in the following table.

Table 5-8: Concepts and sub-concepts derived from the survey responses

Underlying Concept	Basic Theme
Customer access management Cyber security Physical safeguard	Data protection
Minimizing operational risk Considering inbound regulations Risk awareness Risk acceptance	Regulatory compliance
Suppliers Outsourcers Regulators Coordination Supplier exiting Onboarding new supplier	Partners
Funding	Return on investment (Business benefit)
Stakeholder collaboration Staff collaboration Change enablement Staff awareness Staff learning	Organisational change management
Infrastructure Platforms Reference architecture Tools	Technology
Data sharing Data lineage Data traceability Data erasure	Data lifecycle management
Roles and responsibility Operating model	Governance
Data quality audits Reporting Benchmark	Monitoring and measurement

Next, the basic themes were collated and the higher order aggregated themes were created. Finally, the global themes were generated by grouping the organising themes as a whole. The feedback provided by the participants was analysed and categorised into different concept categories.

Table 5-9: High-level concepts derived from the survey responses

Basic Theme	Aggregated Organizing Theme	Global Theme
Regulatory compliance	Risk management view	Integrated people-focused risk-driven DG and DM view
Return on investment / Business benefit		
Technology	Data governance and management view	
Data lifecycle management		
Data protection		
Monitoring and measurement		
Governance		
Organisational change management	People view	
Partners		

As is evident from Table 5-9, there was an overlap between the themes and subthemes that emerged from the survey with the initial version of the 4I Framework. However, few of the sub-themes that arose were unique. Overall, the feedback ensured that there was a suitable level of improvement from the initial version (refer to Appendix F) to the final version of the framework.

5.4 Novelty of the framework

The 4I Framework provides a solution to support DG and DM in DEs. It offers new knowledge that has not been previously discussed in the form of cross-organisational DG and DM. As previously mentioned in the research gaps (Chapter 2), the concepts of DM and DG have existed for decades. Several researchers have worked on this topic in the past and different frameworks already exist, but the focus has been on single organisation. Apart from the scarcity of DE-oriented literature, another obstacle is the absence of an appropriate level of granularity to support businesses in the existing frameworks that explored inter-organisational context. The 4I Framework proposed in this thesis provides systematic guidance to streamline data exchange and assist organisations in an intra-organisational and inter-organisational context.

The rapid pace with which new regulations are being created means that the mandates to manage data are highly volatile. This means organisations are required to be prepared to adopt them at short notice. DG and DM is a mature field but very few researchers have dealt with the growing requirements of data-related regulations and cross-organisational data handling. The increase in data-related privacy and security incidents shows that the existing practices of DG and DM are becoming increasingly difficult to sustain due to unpredictability arising from multi-organisation-driven DE. To the best of our knowledge, different framework exists but they do not take into consideration the recent regulatory requirements on data sharing or takes into consideration the data risks arising due to the misuse of data by the ecosystem partner. Though the framework draws on from existing studies, its usefulness lies in providing a structured step-by-step risk aware approach to govern and manage data.

The framework also introduced a new co-ordinating role called data referee which plays a critical role in ensuring proper collaboration occurs between organisations in the DE.

5.5 Chapter Summary

In this chapter, the 4I Framework was validated through scenario-based case studies and practitioner feedback from industry experts. The three scenario-based case studies indicated that the 4I Framework is domain agnostic and can be used in different industries and contexts. The scenario of the car industry showed how the 4I Framework is able to provide the auxiliary support needed for GDPR compliance. The example involving wearable IoT devices showed that the 4I Framework can be used to support HIPPA regulations and the smart home refrigerator example showed that control mechanisms related to consents can be implemented to support APP principles. The quantitative and qualitative data obtained through the online survey distributed to professionals working in the data space indicated that most of the participants appeared to agree on the applicability, relevance and usability of the framework. The feedback helped in confirming that the framework is fit for the purpose of guiding DGM initiatives in an organisation and meets the evaluation criteria discussed in Chapter 3 (DSR). The survey participants pointed out the strengths and flaws of the framework which immensely helped in enriching the quality of the final version of the thesis outputs, limitations and key contributions are discussed in Chapter 6.

Chapter 6: Discussion and Summary

This chapter outlines the research journey since its inception in 2017. This chapter also presents an overall summary of the research project and the main output of the research, the 4I Framework. The publications emanating from this study are listed in this chapter and the limitations are discussed.

6.1 Research Timeline

The research journey began in autumn 2017 at a higher degree research (HDR) level and was upgraded to a PhD in spring 2018. The study was pursued on a part-time basis from 2020. The research thesis was submitted for review in spring 2021. The research timelines with the key activity performed are illustrated in Figure 6-1.

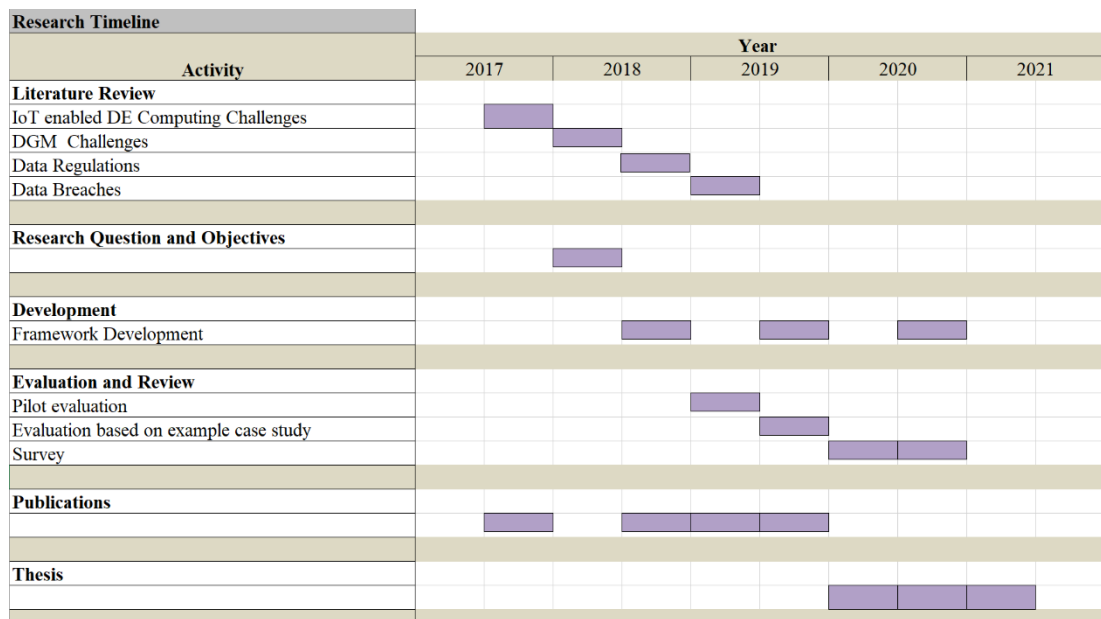


Figure 6-1: Research Timeline

6.2 Research Summary and Output

This research was undertaken to investigate the current challenges to govern and manage data in digital ecosystem. The study started by providing an overview of the key topics of the research such as the digital ecosystem (DE), data governance (DG) and data management (DM). The investigations pointed out the key gaps in governing and managing data beyond the organisational boundaries. To address the research gaps, the research question (RQ) and research objective (RO) were formulated (see Figure 6-2 below).

The main RQ was “*How to effectively assist in the governance and management of data in the distributed digital ecosystem*”. Given the complexity, RQ was subdivided into RSQs.

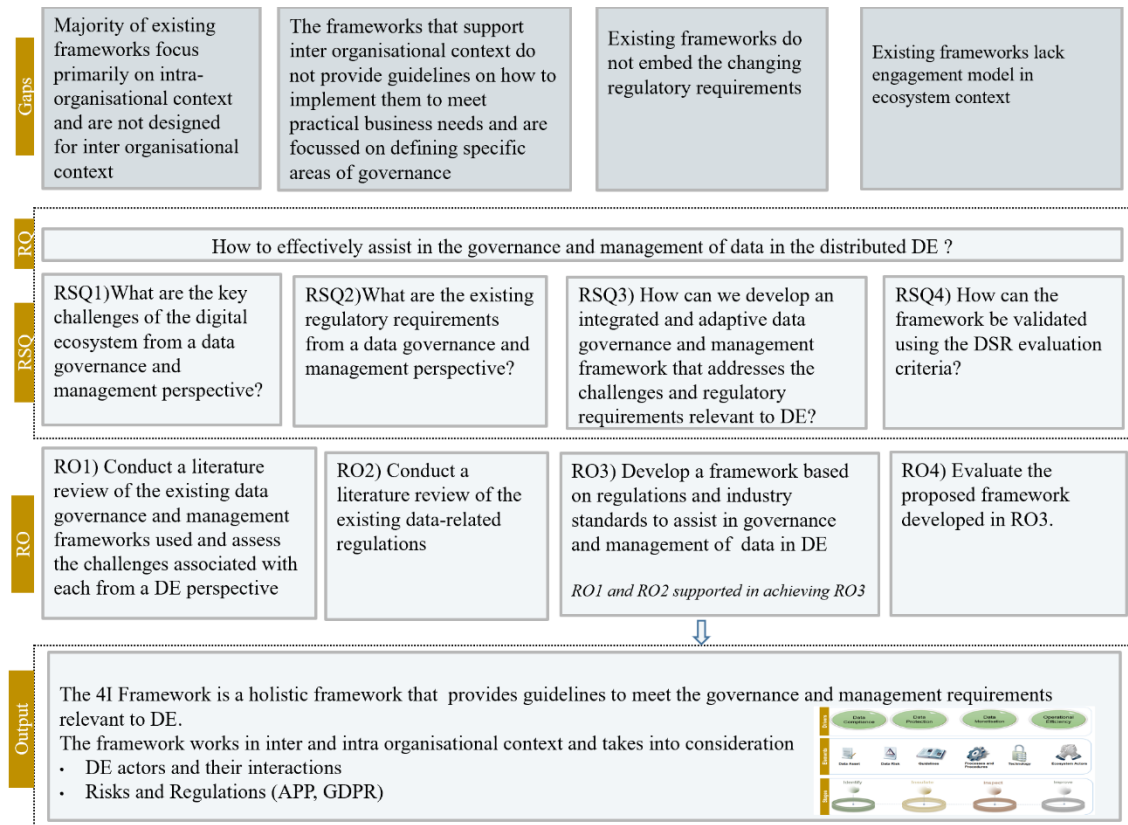


Figure 6-2: Research Questions, Objectives and Output

To address RSQ1 (*What are the key challenges of the digital ecosystem from a data governance and management perspective?*), this study conducted review of existing literature related to governing and managing data (see Chapter 2). The review of the literature highlighted that meeting legal requirements and ensuring protection of data in terms of privacy, security and usage were of paramount interest to organisations in a multi-party environment. From the analysis of the 43 papers examined, it emerged that though governing and managing data is a matured topic and works well in a single organisation context, it is not a widely researched topic in the context of DE. The study found that negligible number of frameworks discussed ecosystem scenarios. The frameworks that focussed on ecosystems were restricted to particular topics or technology and did not address stakeholder interaction related issues adequately.

Furthermore, they did not provide any practical step-by-step implementation guidance. The drawbacks in the existing frameworks provided the justification to create the 4I Framework that resolves the above gaps and meets the DE requirements. To get a better understanding of the real-world problems dominating DGM space in DE, a few recent data breaches were investigated.

The implications of regulations on DGM are undeniable. Different regulations related to data have been formulated and adjusted from time to time in the past few years (Dasgupta, Gill & Hussain 2019b) to protect the rights of the consumer by promoting risk driven practices to manage data. One of the challenges highlighted in the study was limited knowledge of legal requirements. For RSQ2 (*what are the existing regulatory requirements from a data governance and management perspective ?*), the study reviewed a few regulations to capture the key obligations of an organisation in the DE (see Chapter 2). General Data Protection Regulation (GDPR), and the Australian Privacy Act 1988 in Australia, which has stringent requirements in terms of the way data is processed, were investigated as part of the review work.

The outcomes of the review conducted to address RSQ1 and RSQ2 helped in identifying the research gaps and provided valuable information on the current practices that consequently helped in addressing RSQ3 (*How can we develop an integrated and adaptive data governance and management framework that addresses the challenges and regulatory requirements relevant to digital ecosystems?*).

Initially, a few research methods that are relevant to answering the research questions were discussed in Chapter 3. After discussing well known approaches available to IS research, the DSR method, based on the “build” and “evaluate” processes, was chosen as the appropriate strategy that most effectively addressed the RQ in hand. There are

several comprehensive models or processes for conducting the DSR. This research followed the approach based on awareness of problem, suggestion, development, evaluation and outcome (see Section 3.2).

This research was conducted to develop a new artefact, the 4I Framework. SLR data was the basis to construct the preliminary version of the 4I Framework. Chapter 4 presented the iteratively developed 4I Framework, which is the main research output of this thesis. The 4I Framework's systematic approach lists a range of activities that an organisation can perform, both in intra or inter organisation context. The activities, executed in a staged manner, can guide focal firms to perform DGM related business functions.

The 4I Framework comprises of three main components: framework driver, framework elements and is implemented in four stages as shown in Table 6-1:

Table 6-1: Key components of the 4I Framework

Component	Description	Why was this component introduced?
Driver	<p>This component aims to ensure the alignment of the fundamental guiding principles for the DG program with an organisation's strategy. It is subdivided into the following four fundamental reasons for an organisation to undertake DGM activities, namely i) data monetisation strategy, ii) demonstrating compliance, iii) efficient management and monitoring of data and associated risks and iv) increased data protection, resulting in better confidence among customers.</p> <p>Understanding the key driver is the pre-requisite prior to performing any other activities.</p>	<p>To ensure an appropriate justification and business case is in place to secure the funding required to perform the DGM activity.</p> <p>This was incorporated in the final version of the thesis based on feedback from the industry experts (see Section 5.3)</p>
Element	<p>The elements represent the fundamental elements of the framework. The first element is about the data, its benefits and the salient features. Data risk is the second element of the framework that considers the factors impacting the stakeholder. Guidelines, comprising of rules, regulations, policies and standards, is the third element. It establishes the expected activities taken for specific conditions that reflect the regulator's needs, industry best practices and the strategic direction of the organisation. Guidelines are implemented through Policies and procedures (fourth element) with the help of advanced systems, tools and technologies (fifth element). Technology emphasises the utility tools to manage, control and audit data.</p> <p>The final element is oriented towards human factors including roles, responsibilities, accountabilities, communication-collaboration plans and organisational structure.</p>	<p>To ensure appropriate information is available to manage data-related risks and undertake DGM activities to respond to the organisation's business goals.</p> <p>The elements suitable and required to govern and manage data were developed and evolved based on the analysis of existing literature (see Chapter 2) and feedback from the survey (see Chapter 5).</p>
Stages	<p>This component maps the DGM activities into different stages, namely Identify, Insulate, Inspect and Improve.</p>	<p>To ensure that the framework is implemented in a structured manner.</p> <p>The stages were based on the "planning and control" and "risk management" approach of DG (see Section 2.1)</p>

As for RSQ4 (*How can the framework be validated using the DSR evaluation criteria?*), the framework was evaluated in Chapter 5 against the EC outlined in Chapter 3. The validation process involved: 1) scenario-based examples and 2) a survey.

The three scenario-based examples involved different regulations and industry contexts such as Automobile, Wearable IoT device and Smart Home. The use cases demonstrated the applicability of 4I to address real-world DGM issues in DE using the systematic staged activities of the framework (see Chapter 5). This highlighted that the 4I Framework is generic, reusable and fit-for-purpose to support organisations to meet DGM goals, within and beyond organisational boundaries. Two of the three scenarios were published in peer-reviewed conferences (Dasgupta, Gill & Hussain 2019a; Dasgupta, Gill & Hussain 2020) while the third one was published as a book chapter (Dasgupta, Gill & Hussain 2019b).

The 4I framework was presented to thirty practitioners and experts from the industry through an anonymous survey. The participants spanned several continents and included experts from Asia, Australia and Europe. The objective of the survey was to determine the usefulness, relevance and importance of the framework and its components to meet business needs. The survey collected both quantitative and qualitative data. The quantitative data was analysed statistically (see Section 5.3.7). The results of the empirical survey indicated that respondents found the framework easy to understand (56%) and the framework was suitable to meet DG and DM objectives of a firm (70%). This aligned closely to the qualitative responses received such as *“in light of recent penalties issued to support, this is an important work to support management of data”* and *“having a set of guidance will help organisations in structuring data related projects”*.

Most of the participants agreed with the four staged approach (80%) and the elements included in the framework stayed within the acceptable range (mean range (3.6-4.7), skewness range (-0.6, 0.17), kurtosis range (-1.7,1.05)). The framework was further

refined based on suggestions from the participants (see Section 5.3.8). This included the addition of the new relevant component such as “Drivers”. Moreover, attributes related to few of the individual elements such as “training for external actors”, were updated in the final 4I version.

Overall, based on the assessment of the 4I Framework using scenario-based examples and empirical survey, it appears that the 4I Framework can be used to assist business to govern and manage data as a critical asset in DE in different contexts. The framework provides organisations with a step-by-step staged approach to meet DG and DM requirements.

6.3 Research Limitations and Future work

This thesis is based on an extensive study on documents, reports, research articles and industry practitioner feedback. While the proposed 4I Framework has been rigorously developed, few limitations remain.

- One of the limitations that needs to be considered when assessing and using this research is that the framework was applied to three sectors (Automobile, Wearable and Smart Home Retail). It would have been beneficial to apply it in other sectors such as the financial sector.
- The time constraint of the PhD work limited the ability to incorporate requirements from multiple regulations into the framework. The 4I incorporated requirements from two regulations: GPDR and APP and therefore needs to be tailored to satisfy compliance mandates from other jurisdictions. Future research could extend to encompass further regulatory contexts.
- Thirty participants who had the required expertise in the DGM space completed the survey. Another limitation is the lack of end-user participation in the survey, particularly those who have a good grasp of the data management spectrum. However, the results from the surveys are reliable since most of the users were involved with continuous feedback received from the stakeholders involved. Further research by involving additional participants can strengthen the current findings to further improve the 4I Framework and reduce the possibility of survey participant bias.

- Research output is influenced by individual creative mind and related body of knowledge. Thus, it is quite possible that the same research repeated by another researcher may result in different output which is not unusual.
- The final limitation of this research is that the framework takes a focal organisation perspective with the scope restricted to aspects such as risks and actors and interactions (see Section 1.5). Technical solutions and data management disciplines such as master-data-management or data catalogue were not discussed in detail in this dissertation.

6.4 Contribution to Research

Throughout this study, the gaps in existing literature related to data-sharing challenges in a distributed digital ecosystem was explored. The main contribution of the thesis is the 4I Framework that was developed to assist businesses in governing and managing data, both within and beyond organisational boundaries. The 4I Framework can be used in the form of guidelines to support DG and DM activities in the DE. Another contribution of this study is the synthesis of the comprehensive literature on the different DG and DM concepts and the best practices used in siloed organisations, which adds to the body of knowledge on DGM. The framework strives to improve integrated DG and DM and can be used as a basis for building data sharing theories in the area of DEs.

Table 6-2: Key contributions

Contribution	Source
4I Framework	Thesis
Book Chapter	Dasgupta, A., Gill, A.Q. & Hussain, F. 2019, 'Privacy of IoT-Enabled Smart Home Systems', IoT and Smart Home Automation, IntechOpen.
Conference	Dasgupta, A. & Gill, A.Q. 2017, 'Fog Computing Challenges: A Systematic Review', paper presented to the Australasian Conference on Information Systems,, Hobart.
Conference	Dasgupta, A., Gill, A. & Hussain, F.K. 2019a, 'A Conceptual Framework for Data Governance in IoT-enabled Digital IS Ecosystems', DATA, pp. 209-16.
Conference	Dasgupta, A., Gill, A.Q. & Hussain, F. 2020, 'A Review of General Data Protection Regulation for Supply Chain Ecosystem', Innovative Mobile and Internet Services in Ubiquitous Computing, eds L. Barolli, F. Xhafa & O.K. Hussain, Springer International Publishing, Cham, pp. 456-65.

6.5 Contribution to Practice

The 4I Framework has evolved through the study of existing practices, regulatory obligations and investigation of the root causes of data breach incidents (see Chapter 2). It was validated and refined through practitioner feedback from industry experts (see Chapter 5).

The examples involving cars, smart wearable devices and smart refrigerators demonstrated that the 4I Framework is suitable to address concerns about the use of data which is shared by external actors. Thus, the test examples demonstrate that the framework is fit-for-purpose and can be viewed as a practical use of the 4I Framework and can be replicated in other contexts by practitioners.

The framework can also be used to assist businesses identify the key requirements and activities which need to be performed when sharing data across organisational boundaries, but without data-related risks. In certain industries and jurisdictions, organisations are required to report their data governance and management practices to regulators. For example, Basel Committee of Banking Supervision's BCBS 239 requires banks to demonstrate the practical applications of their data governance and data (Grody & Hughes 2016). Thus, the framework is useful for industry practitioners to address specific of the reporting requirements.

Finally, it is anticipated that this research sets the basis to elaborate on the notion of collaborative DGM.

Conclusion

The rapidly growing DE presents an enormous opportunity for businesses. Both researchers and industry practitioners agree that the proper governance and management of data will be a key enabler for the growth of the DE. However, like other technological innovations, there are a range of data governance and management challenges faced by organisations. To become profitable in DE, businesses need to develop the right mechanisms to manage data and satisfy the evolving needs of the regulators.

In this thesis, the 4I Framework is presented which provides a practical framework for governing and managing data in the DE involving multiple actors. The 4I Framework was developed iteratively using a well-established DSR method and can be looked upon as a pragmatic guide that comprises of activities that could be undertaken in an organisation to address the DGM challenges. The intended users of this framework are the DGM teams (managers, data stewards, consultants, engineers, developers) in any organisational context. The 4I Framework is technology agnostic and is not fixed on any specific technology stack or tool. The framework's components can be traced to the literature sources or expert feedback. The next step of this research is to enhance the framework by developing a publicly available self-assessment tool to assess the maturity of an organisation. Another important area for consideration is to use the 4I Framework as a baseline to create a detailed inventory of risks, control taxonomies and regulatory requirements. This will help the ecosystem actors, particularly management, to accelerate DE adoption through collaborative DGM activities. Future research can further enhance the framework to include participants in addition to the already evaluated work.

Bibliography

- Abraham, R., Schneider, J. & vom Brocke, J. 2019, 'Data governance: A conceptual framework, structured review, and research agenda', *International Journal of Information Management*, vol. 49, pp. 424-38.
- Adida, C., Crotty, P.L., McGrath, J., Berrebi, D., Diebold, J. & Altieri, D.C. 1998, 'Developmentally regulated expression of the novel cancer anti-apoptosis gene survivin in human and mouse differentiation', *The American journal of pathology*, vol. 152, no. 1, p. 43.
- AFR, A.F.R. 2019, 'LandmarkWhite faces regulator scrutiny over IT response, disclosure', <<https://www.afr.com/property/landmarkwhite-faces-regulator-scrutiny-over-it-response-disclosure-20190311-h1c8xr>>.
- Akalu, R. 2018, 'Privacy, consent and vehicular ad hoc networks (VANETs)', *Computer Law & Security Review*, vol. 34, no. 1, pp. 37-46.
- Al-Badi, A., Tahrini, A. & Khan, A.I. 2018, 'Exploring Big Data Governance Frameworks', *Procedia Computer Science*, vol. 141, pp. 271-7.
- Al-Ruithe, M. & Benkhelifa, E. 2017, 'A conceptual framework for cloud data governance-driven decision making', *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, pp. 1-6.
- Al-Ruithe, M., Benkhelifa, E. & Hameed, K. 2016a, 'A Conceptual Framework for Designing Data Governance for Cloud Computing', *Procedia Computer Science*, vol. 94, pp. 160-7.
- Al-Ruithe, M., Benkhelifa, E. & Hameed, K. 2016b, 'Key Dimensions for Cloud Data Governance', *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 379-86.
- Al-Ruithe, M., Benkhelifa, E. & Hameed, K. 2018, 'Data Governance Taxonomy: Cloud versus Non-Cloud', *Sustainability*, vol. 10, no. 1, p. 95.
- Alhassan, I., Sammon, D. & Daly, M. 2018, 'Data governance activities: a comparison between scientific and practice-oriented literature', *Journal of Enterprise Information Management*, vol. 31, no. 2, pp. 300-16.
- Aljeraisy, A., Barati, M., Rana, O. & Perera, C. 2021, 'Privacy Laws and Privacy by Design Schemes for the Internet of Things: A Developer's Perspective', *ACM Comput. Surv.*, vol. 54, no. 5, p. Article 102.
- Anderson, C. 2017, 'Swedish Government Scrambles to Contain Damage From Data Breach'.
- Anwar, M.J. 2021, 'Adaptive Digital Identity Verification Reference Architecture (ADIVRA) Framework'.
- Anwar, M.J., Gill, A.Q., Hussain, F.K. & Imran, M. 2021, 'Secure big data ecosystem architecture: challenges and solutions', *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, pp. 1-30.
- APRA 2013, *CPG 235 – Managing Data Risk*, <<https://www.apra.gov.au/sites/default/files/CPG-235-Managing-Data-Risk.pdf>>.
- B. Kitchenham, S.C. 2007, *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Keele University and Durham University, UK, (2007), vol. ver 2.3, EBSE Technical Report.
- Banakar, V., Shah, A., Shastri, S., Wasserman, M. & Chidambaram, V. 2019, 'Analyzing the Impact of GDPR on Storage Systems', *arXiv preprint arXiv:1903.04880*.

- Banerjee, S., Hemphill, T. & Longstreet, P. 2018, 'Wearable devices and healthcare: Data sharing and privacy', *The Information Society*, vol. 34, no. 1, pp. 49-57.
- Bastos, D., Giubilo, F., Shackleton, M. & El-Moussa, F., 'GDPR Privacy Implications for the Internet of Things'.
- Begg, C. & Caira, T.J.T.E.J.I.S.E. 2012, 'Exploring the SME quandary: Data governance in practise in the small to medium-sized enterprise sector', vol. 15, no. 1.
- Bennett, M. 2017, 'What is a digital ecosystem, and how can your business benefit from one?', *The Telegraph*, <<https://www.telegraph.co.uk/business/ready-and-enabled/what-is-a-digital-ecosystem/?>>.
- Brickson, P. 2016, 'Implementing Data Governance with Agile Project Management Methodologies', The College of St. Scholastica.
- Brous, P., Janssen, M. & Vilminko-Heikkinen, R. 2016a, 'Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles', Springer International Publishing, pp. 115-25.
- Brous, P., Janssen, M. & Vilminko-Heikkinen, R. 2016b, 'Coordinating decision-making in data management activities: a systematic review of data governance principles', *International Conference on Electronic Government*, Springer, pp. 115-25.
- Bughin, J., Catlin, T. & Dietz, M. 2019, 'The right digital-platform strategy', *McKinsey Quarterly*, vol. 2, pp. 1-4.
- California, S.o. 2018, *SB-327 Information privacy: connected devices.*, <https://leginfo.ca.gov/faces/billCompareClient.xhtml?bill_id=2017201805B327&showamends=false>.
- Cannon, C.H., Summers, M., Harting, J.R. & Kessler, P.J.J.B. 2007, 'Developing conservation priorities based on forest type, condition, and threats in a poorly known ecoregion: Sulawesi, Indonesia', vol. 39, no. 6, pp. 747-59.
- Cave, A.E. 2017, 'Exploring Strategies for Implementing Data Governance Practices', D.I.T. thesis, Walden University, Ann Arbor.
- Chatterjee, C. & Sokol, D.D.J.C.H.o.C. 2019, 'Data Security, Data Breaches, and Compliance'.
- Chaudhuri, A. 2016, 'Internet of things data protection and privacy in the era of the General Data Protection Regulation', *Journal of Data Protection & Privacy*, vol. 1, no. 1, pp. 64-75.
- Clarke, R. 1999, 'Internet privacy concerns confirm the case for intervention', *Commun. ACM*, vol. 42, no. 2, pp. 60-7.
- Collibra 2018, *6 TYPICAL GDPR & DATA GOVERNANCE QUESTIONS, EXPLAINED*, viewed 20 April 2018 2018, <<https://www.collibra.com/blog/8-gdpr-questions-answered/>>.
- Commission, E. 2016, 'European Commission. Eu data protection reform what benefits for businesses in europe'.
- Commissioner, A.I.C.a.P. 2020, *Flight Centre found to have interfered with privacy*, Australia, <<https://www.oaic.gov.au/updates/news-and-media/flight-centre-found-to-have-interfered-with-privacy/>>.
- Commissioner for Privacy and Data Protection, V., Australia 2016, *Guidelines for sharing personal information*, <https://www.cpdp.vic.gov.au/images/content/pdf/privacy_guidelines/CPDP_Information_sharing_guidelines.pdf>.
- Council, E. 2015, *EDMC DCAM version 1.0*, SI: sn.
- Cranefield, S. & Purvis, M.K. 1999, *UML as an ontology modelling language*, Department of Information Science, University of Otago New Zealand.
- Curry, E. & Sheth, A. 2018, 'Next-Generation Smart Environments: From System of Systems to Data Ecosystems', *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 69-76.

- Darking, M., Nachira, F., Nicolai, A., Dini, P., Le Louran, M. & Leon, L. 2007, 'Understanding the role of governance in the context of digital ecosystems', *Digital Business Ecosystems*, pp. 78-82.
- Dasgupta, A., Gill, A. & Hussain, F.K. 2019a, 'A Conceptual Framework for Data Governance in IoT-enabled Digital IS Ecosystems', *DATA*, pp. 209-16.
- Dasgupta, A. & Gill, A.Q. 2017, 'Fog Computing Challenges: A Systematic Review', paper presented to the *Australasian Conference on Information Systems*, Hobart.
- Dasgupta, A., Gill, A.Q. & Hussain, F. 2019b, 'Privacy of IoT-Enabled Smart Home Systems', *IoT and Smart Home Automation*, IntechOpen.
- Dasgupta, A., Gill, A.Q. & Hussain, F. 2020, 'A Review of General Data Protection Regulation for Supply Chain Ecosystem', *Innovative Mobile and Internet Services in Ubiquitous Computing*, eds L. Barolli, F. Xhafa & O.K. Hussain, Springer International Publishing, pp. 456-65.
- DataGuidance 2018, *Comparing Privacy Laws*, <[https://fpf.org/wp-content/uploads/2018/11/GDPR CCPA Comparison-Guide.pdf](https://fpf.org/wp-content/uploads/2018/11/GDPR_CCPA_Comparison-Guide.pdf)>.
- Deloitte 2019, *Data governance for next-generation platforms*.
- Development, U.N.C.o.T.a. 2021, *Data Protection and Privacy Legislation Worldwide*, Geneva.
- Dimick, C. 2013, 'Governance Apples and Oranges', *Journal of AHIMA*, vol. 84, no. 11, pp. 60-2.
- Dyché, J. & Levy, E. 2011, *Customer data integration: Reaching a single version of the truth*, vol. 7, John Wiley & Sons.
- Egelstaff, R. & Wells, M. 2013, 'Data governance frameworks and change management', *Studies In Health Technology And Informatics*, vol. 193, pp. 108-19.
- Ender, L. 2021, 'Data Governance in Digital Platforms: A case analysis in the building sector'.
- Engels, B. 2019, 'Data Governance as the Enabler of the Data Economy', *Intereconomics*, vol. 54, no. 4, pp. 216-22.
- Esposito, C., Castiglione, A., Pop, F. & Choo, K.K.R. 2017, 'Challenges of Connecting Edge and Cloud Computing: A Security and Forensic Perspective', *IEEE Cloud Computing*, vol. 4, no. 2, pp. 13-7.
- Fitbit 2019, *Fitbit to Be Acquired by Google*, Published, <<https://investor.fitbit.com/press/press-releases/press-release-details/2019/Fitbit-to-Be-Acquired-by-Google/default.aspx>>.
- Franke, U. 2017, 'The cyber insurance market in Sweden', *Computers & Security*, vol. 68, pp. 130-44.
- Fu, H. 2006, 'Formal concept analysis for digital ecosystem', *Machine Learning and Applications, 2006. ICMLA'06. 5th International Conference on*, IEEE, pp. 143-8.
- Gantait, A., Patra, J. & Mukherjee, A. 2018, *Defining your IoT governance practices*, IBM, 2018, <<https://www.ibm.com/developerworks/library/iot-governance-01>>.
- Gartner 2016a, *Data Risks in the Internet of Things Demand Extensive Information Governance*.
- Gartner 2016b, *Organizing for Big Data Through Better Process and Governance*.
- Gartner 2017, 'Industry Data Governance Is Key to Developing a Smart City Platform', no. G00325170.
- Gartner 2018, 'Get Ready for the Impact of GDPR on Content and Collaboration'.
- Geerts, G.L. 2011, 'A design science research methodology and its application to accounting information systems research', *International Journal of Accounting Information Systems*, vol. 12, no. 2, pp. 142-51.

- Geisler, S., Vidal, M.-E., Capiello, C., Loscio, B.F., Gal, A., Jarke, M., Lenzerini, M., Missier, P., Otto, B. & Paja, E. 2021, 'Knowledge-driven Data Ecosystems Towards Data Transparency', *arXiv preprint arXiv:2105.09312*.
- Gill, A.Q. 2015, *Adaptive Cloud Enterprise Architecture*, Book, World Scientific, Hackensack, New Jersey.
- Gill, A.Q. 2021, 'A Theory of Information Trilogy: Digital Ecosystem Information Exchange Architecture', *Information*, vol. 12, no. 7, p. 283.
- Government, A. 2019, *Seven Steps to Securing Your Smart Health Devices*, <<https://www.digitalhealth.gov.au/about-the-agency/digital-health-space/seven-steps-to-securing-your-smart-health-devices>>.
- Grody, A.D. & Hughes, P.J. 2016, 'Risk Accounting-Part 1: The risk data aggregation and risk reporting (BCBS 239) foundation of enterprise risk management (ERM) and risk governance', *Journal of Risk Management in Financial Institutions*, vol. 9, no. 2, pp. 130-46.
- Gromenko, B.D., 'Corporate Security in the EU and GDPR: Data breaches in British Airways and The Marriott International'.
- Guardian 2020, 'Travelers 'being held to ransom' by hackers said to be demanding \$3m', *Guardian*, <<https://www.theguardian.com/technology/2020/jan/07/travelers-being-held-ransom-hackers-said-demanding-3m>>.
- Hair, J., Black, W.C., Babin, B., Anderson, R. & Tatham, R. 2014, 'Pearson new international edition', *Multivariate data analysis, Seventh Edition*, Pearson Education Limited Harlow, Essex.
- Hair, J.F. 2009, 'Multivariate data analysis'.
- Handel, P., Skog, I., Wahlstrom, J., Bonawiede, F., Welch, R., Ohlsson, J. & Ohlsson, M. 2014, 'Insurance Telematics: Opportunities and Challenges with the Smartphone Solution', *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 57-70.
- Herald, S.M. 2020, 'Trusted inside access', <<https://www.smh.com.au/national/nsw/trusted-inside-access-sydney-it-contractor-arrested-over-landmark-white-data-breach-20191002-p52wup.html>>.
- Heredia-Vizcaíno, D. & Nieto, W. 2019, 'A Governing Framework for Data-Driven Small Organizations in Colombia', Springer International Publishing, pp. 622-9.
- Hevner, A. & Chatterjee, S. 2010, 'Design science research in information systems', *Design research in information systems*, Springer, pp. 9-22.
- Hitachi 2019, *Build a Data Governance Strategy for the New Digital Era*, <<https://www.hitachivantara.com/en-us/pdf/ebook/build-data-governance-strategy-for-new-digital-era-ebook.pdf>>.
- Horne, N.W. 1995, 'Information as an asset—The board agenda', *Computer Audit Update*, vol. 1995, no. 9, pp. 5-11.
- Hyndman, R. 2008, 'Quantitative business research methods', *Department of econometrics and Business Statistics. Monash University (Clayton campus)*.
- Ibrahim Alhassan, D.S.M.D. 2016, 'Data governance activities: an analysis of the literature', *Journal of Decision Systems*.
- Infotech, I. 2016, *Assessing and implementing a Data Governance program*, viewed August 2018, <<http://docplayer.net/11775196-Assessing-and-implementing-a-data-governance-program-in-an-organization.html>?>.
- Isaak, J. & Hanna, M.J. 2018, 'User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection', *Computer*, vol. 51, no. 8, pp. 56-9.
- Jagals, M. & Karger, E. 2021, 'INTER-ORGANIZATIONAL DATA GOVERNANCE: A LITERATURE REVIEW'.

- Janne J. Korhonen, Ilkka Melleri, Kari Hiekkanen & Helenius, M. 2013, 'Designing Data Governance Structure: An Organizational Perspective', *GSTF Journal on Computing (JoC)*, vol. 2.
- Janssen, M., Brous, P., Estevez, E., Barbosa, L.S. & Janowski, T. 2020, 'Data governance: Organizing data for trustworthy Artificial Intelligence', *Government Information Quarterly*, vol. 37, no. 3, p. 101493.
- Jifa, G. & Lingling, Z. 2014, 'Data, DIKW, Big Data and Data Science', *Procedia Computer Science*, vol. 31, pp. 814-21.
- Jonkers, H., Proper, E. & Turner, M. 2009, 'TOGAF 9 and ArchiMate 1.0', *White Paper. The Open Group*.
- Kalibatiene, D. & Vasilecas, O. 2011, 'Survey on Ontology Languages', Springer Berlin Heidelberg, pp. 124-41.
- Kees, A., Oberländer, A.M., Röglinger, M. & Rosemann, M. 2015, 'Understanding the Internet of Things: A Conceptualisation of Business-to-Thing (B2T) Interactions', *ECIS*.
- Kerber, W. 2019, 'Data Governance in Connected Cars: The Problem of Access to In-Vehicle Data', *JIPITEC*, vol. 9, no. 3, pp. 310-31.
- Kerber, W.J.J.I.P.I.T. & L., E.C. 2018, 'Data governance in connected cars: The problem of access to in-vehicle data', vol. 9, p. 310.
- Khatri, V. & Brown, C.V.J.C.o.t.A. 2010, 'Designing data governance', vol. 53, no. 1, pp. 148-52.
- Kim, H.Y. & Cho, J. 2017, 'Data Governance Framework for Big Data Implementation with a Case of Korea', *2017 IEEE International Congress on Big Data (BigData Congress)*, pp. 384-91.
- Kimmel, A. 1988, *Ethics and values in applied social research*, vol. 12, Sage.
- Kontzer, T. 2006, 'An End To Data Anarchy', *InformationWeek*, no. 1084, pp. 73-5.
- KPMG 2016, *data governance*, <<https://home.kpmg/sg/en/home/insights/2016/03/data-governance.html>>.
- KPMG 2021, *It's not what you know but who you share it with!*, viewed Jan 1 2021, <<https://home.kpmg/qm/en/home/insights/2019/05/It%E2%80%99s%20not%20what%20you%20know,%20but%20who%20you%20share%20it%20with!%20.html>>.
- Kuechler, B. & Vaishnavi, V. 2008, 'On theory development in design science research: anatomy of a research project', *European Journal of Information Systems*, vol. 17, no. 5, pp. 489-504.
- Kuhn, M.L. 2018, '147 Million Social Security Numbers for Sale: Developing Data Protection Legislation After Mass Cybersecurity Breaches', *Iowa Law Review*, vol. 104, no. 1, pp. 417-45.
- Ladley, J. 2019, *Data governance: How to design, deploy, and sustain an effective data governance program*, Academic Press.
- Lee, S.U., Zhu, L. & Jeffery, R. 2019, 'Data Governance Decisions for Platform Ecosystems', *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Lending, C., Minnick, K. & Schorno, P.J. 2018, 'Corporate Governance, Social Responsibility, and Data Breaches', *Financial Review*, vol. 53, no. 2, pp. 413-55.
- Li, Q., Lan, L., Zeng, N., You, L., Yin, J., Zhou, X. & Meng, Q. 2019, 'A Framework for Big Data Governance to Advance RHINs: A Case Study of China', *IEEE Access*, vol. 7, pp. 50330-8.
- Li, W., Badr, Y. & Biennier, F. 2012, 'Digital ecosystems: challenges and prospects', *proceedings of the international conference on management of Emergent Digital EcoSystems*, ACM, pp. 117-22.

- Linares-Vásquez, M., Bavota, G. & Escobar-Velásquez, C. 2017, 'An empirical study on android-related vulnerabilities', *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*, IEEE, pp. 2-13.
- Lis, D. & Otto, B. 2020, 'Data Governance in Data Ecosystems—Insights from Organizations'.
- Lis, D. & Otto, B. 2021, 'Towards a Taxonomy of Ecosystem Data Governance', *Proceedings of the 54th Hawaii International Conference on System Sciences*, p. 6067.
- Lütjen, H., Schultz, C., Tietze, F. & Urmetzer, F. 2019, 'Managing ecosystems for service innovation: A dynamic capability view', *Journal of Business Research*.
- Mahanti, R. 2018, 'Data Governance Implementation: Critical Success Factors', *Software Quality Professional*, vol. 20, no. 4, pp. 4-21.
- Mansfield-Devine, S. 2016, 'Data protection: prepare now or risk disaster', *Computer Fraud & Security*, vol. 2016, no. 12, pp. 5-12.
- March, S. & Smith, G. 1995, *Design and Natural Science Research on Information Technology*, vol. 15.
- McGrath, P., Blumer, C. & Carter, J.J.A.N. 2018, 'Medical appointment booking app HealthEngine sharing clients' personal information with lawyers'.
- Mell, P. & Grance, T. 2009, 'The NIST definition of cloud computing', *National Institute of Standards and Technology*, vol. 53, no. 6, p. 50.
- Merkus, J. 2015a, 'Data Governance Maturity Model'.
- Merkus, J.R. 2015b, 'Data Governance Maturity Model', Open Universiteit Nederland.
- Michael S. Smith 2015, 'Protecting Privacy in an IoT-Connected World', *Tech trend*
- Miles, M.B., Huberman, A.M., Huberman, M.A. & Huberman, M. 1994, *Qualitative data analysis: An expanded sourcebook*, sage.
- Moore, J.F.J.T.a.b. 2006, 'Business ecosystems and the view from the firm', vol. 51, no. 1, pp. 31-75.
- Mosley, M., Brackett, M.H., Earley, S. & Henderson, D. 2010, *DAMA guide to the data management body of knowledge*, Technics Publications.
- Nachira, F., Dini, P. & Nicolai, A. 2007, 'A network of digital business ecosystems for Europe: roots, processes and perspectives', *European Commission, Bruxelles, Introductory Paper*, vol. 106.
- NBC 2015, 'Smart Refrigerators Hacked to Send out Spam: Report', <https://www.nbcnews.com/tech/internet/smart-refrigerators-hacked-send-out-spam-report-n11946>.
- Nielsen, O.B. 2017, 'A comprehensive review of data governance literature', *Selected Papers IRIS*, vol. 8, pp. 120-33.
- Niemi, E. 2011, 'Designing a Data Governance Framework', *Proceedings of the IRIS Conference, At Oslo, Norway*, vol. 14.
- Niglas, K. 2001, 'Paradigms and methodology in educational research'.
- Nokkala, T., Salmela, H. & Toivonen, J. 2019, *Data Governance in Digital Platforms*.
- Norbib, K. & Bakar, N.A.A.J.O.I.J.o.I. 2021, 'Data Governance Model For The Ministry Of Education Malaysia Using Enterprise Architecture Approach', vol. 9, no. Special Issue 1, pp. 1-15.
- Offermann, P., Levina, O., Schönherr, M. & Bub, U. 2009, 'Outline of a design science research process', paper presented to the *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, Philadelphia, Pennsylvania, <https://doi.org/10.1145/1555619.1555629>.

- Ostrowski, L. & Helfert, M. 2012, 'Design science evaluation—example of experimental design', *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, no. 9, pp. 253-62.
- Otto, B. 2011, 'A morphology of the organisation of data governance', *ECIS*, vol. 20, p. 1.
- Otto, B. & Jarke, M. 2019, 'Designing a multi-sided data platform: findings from the International Data Spaces case', *Electronic Markets*, vol. 29, no. 4, pp. 561-80.
- Paananen, J.-P. 2020, 'A New Data Governance Model for the Bank of Finland'.
- Pappas, I.O., Mikalef, P., Giannakos, M.N., Krogstie, J. & Lekakos, G. 2018, 'Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies', *Information Systems and e-Business Management*, vol. 16, no. 3, pp. 479-91.
- Parycek, P. & Pereira, G.V. 2017, 'Drivers of Smart Governance: towards to evidence-based policy-making', paper presented to the *Proceedings of the 18th Annual International Conference on Digital Government Research*, Staten Island, NY, USA, <<https://doi.org/10.1145/3085228.3085255>>.
- Paskaleva, K., Evans, J., Martin, C., Linjordet, T., Yang, D. & Karvonen, A. 2017, 'Data governance in the sustainable smart city', *Informatics*, vol. 4, Multidisciplinary Digital Publishing Institute, p. 41.
- Pearce, G. 2017a, *Boosting-Cyber-Security-With-Data-Governance_joa*.
- Pearce, G. 2017b, 'Boosting Cyber Security With Data Governance and Enterprise Data Management', *ISACA Journal*, vol. 3.
- Peffer, K., Tuunanen, T. & Niehaves, B. 2018, 'Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research', *European Journal of Information Systems*, vol. 27, no. 2, pp. 129-39.
- Peffer, K., Tuunanen, T., Rothenberger, M. & Chatterjee, S. 2007, 'A Design Science Research Methodology for Information Systems Research', *J. Manage. Inf. Syst.*, vol. 24, no. 3, pp. 45-77.
- Perscheid, G., Ostern, N.K. & Moormann, J. 2020, 'Determining Platform Governance: Framework for Classifying Governance Types', *ICIS*.
- Peterson, R. 2004, 'Crafting information technology governance', *Information systems management*, vol. 21, no. 4, pp. 7-22.
- Pike, E.R.J.E.L.J., Forthcoming 2019, 'Defending Data: Toward Ethical Protections and Comprehensive Data Governance'.
- Prat, N., Comyn-Wattiau, I. & Akoka, J. 2014, 'Artifact Evaluation in Information Systems Design-Science Research-a Holistic View', *PACIS*, Citeseer.
- Prieëlle, F.D., Reuver, M.D. & Rezaei, J. 2020, 'The Role of Ecosystem Data Governance in Adoption of Data Platforms by Internet-of-Things Data Providers: Case of Dutch Horticulture Industry', *IEEE Transactions on Engineering Management*, pp. 1-11.
- PWC 2019, *The data opportunity*, <<https://www.pwc.com.au/assurance/protecting-and-realising-the-value-of-digital-assets/data-governance.html>>.
- R. Minerva, A.B., D. Rotondi, 2015, *Towards a definition of the Internet of Things (IoT)*, <<http://iot.ieee.org/definition.html>>.
- Rapoport, R.N. 1970, 'Three dilemmas in action research: with special reference to the Tavistock experience', *Human relations*, vol. 23, no. 6, pp. 499-513.
- Rashid, T. & Ahmed, M.J.S.A.f.t.l.o.E. 2020, 'Munir A. Saeed', p. 93.
- Redgate 2018, *Data Governance Implementation Report 2018*, <<https://www.red-gate.com/solutions/entrypage/database-governance-report>>.
- Reis, J.R., Viterbo, J. & Bernardini, F. 2018, 'A rationale for data governance as an approach to tackle recurrent drawbacks in open data portals', paper presented to the *Proceedings of the 19th Annual International Conference on Digital Government*

- Research: Governance in the Data Age, Delft, The Netherlands, <<https://doi.org/10.1145/3209281.3209354>>.
- Rob, P. & Coronel, C. 1997, 'Database systems', *Design, Implementation and Management, Third Edition, Course Technologies*.
- Robert C. Rickards 2012, 'DATA GOVERNANCE CHALLENGES FACING CONTROLLERS', *International Journal Of Business, Accounting, & Finance*, pp. 25-42.
- Romanou, A. 2018, 'The necessity of the implementation of Privacy by Design in sectors where data protection concerns arise', *Computer Law & Security Review*, vol. 34, no. 1, pp. 99-110.
- Rose, K., Eldridge, S. & Chapin, L. 2015, 'The internet of things: An overview', *The Internet Society (ISOC)*, pp. 1-50.
- Security, E.U.A.F.N.a.I. 2018, *Information Sharing and Analysis Center (ISACs) - Cooperative models*.
- Sen, A. & Madria, S. 2018, 'Data Analysis of Cloud Security Alliance's Security, Trust & Assurance Registry', *Proceedings of the 19th International Conference on Distributed Computing and Networking*, ACM, p. 42.
- Senyo, P.K., Liu, K. & Effah, J. 2019, 'Digital business ecosystem: Literature review and a framework for future research', *International Journal of Information Management*, vol. 47, pp. 52-64.
- Sharma, S., Chen, K. & Sheth, A. 2018, 'Toward Practical Privacy-Preserving Analytics for IoT and Cloud-Based Healthcare Systems', *IEEE Internet Computing*, vol. 22, no. 2, pp. 42-51.
- Sicari, A.R., Cinzia Capiello, Daniele Miorandi and Alberto Coen-Porisini 2018, 'Toward Data Governance in the Internet of Things'.
- Skierka, I. 2018, 'The governance of safety and security risks in connected healthcare', *Living in the Internet of Things: Cybersecurity of the IoT - 2018*, Institution of Engineering and Technology (IET), pp. 1-12.
- Society, R. 2017, *Data management and use : governance in the 21st century*.
- Stanley, J. & Briscoe, G. 2010, 'The ABC of digital business ecosystems', *arXiv preprint arXiv:1005.1899*.
- STEENBEEK, I. 2019, 'THE "ORANGE" MODEL OF DATA MANAGEMENT'.
- STEFANO, R. 2016, 'Improving data governance thought Ontology'.
- Sullivan, F. 2015, 'Game Changers—Artificial Intelligence: What You Need to Know', p. 89, Nov 2015.
- Tansley, A.G.J.E. 1935, 'The use and abuse of vegetational concepts and terms', vol. 16, no. 3, pp. 284-307.
- TDAN.com, R.S.S.K.C. 2015, 'Data Governance and the Internet of Things', *Monthly Webinar Series Hosted by DATAVERSITY*
- tegan, 'Self-Regulation within the Wearable Device Industry and The Alignment to Device Users' Perceptions of Health Data Privacy
- .
- Thammaboosadee, S. & Dumthanasarn, N. 2018, 'Proposed Amendments of Public Information Act Towards Data Governance Framework for Open Government Data: Context of Thailand', *2018 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, pp. 1-5.
- Thompson, W.R. 2018, 'Worldwide survey of fitness trends for 2019', *ACSM's Health & Fitness Journal*, vol. 22, no. 6, pp. 10-7.

- Tonkin, C. 2019, 'Comm Bank slapped over failed security',
<<https://ia.acs.org.au/article/2019/comm-bank-slapped-over-failed-security.html>>.
- Tsujimoto, M., Kajikawa, Y., Tomita, J. & Matsumoto, Y. 2018, 'A review of the ecosystem concept—Towards coherent ecosystem design', *Technological Forecasting and Social Change*, vol. 136, pp. 49-58.
- Valdez-De-Leon, O. 2019, 'How to develop a digital ecosystem: a practical framework', *Technology Innovation Management Review*, vol. 9, no. 8.
- van den Broek, T. & van Veenstra, A.F. 2015, 'Modes of governance in inter-organizational data collaborations'.
- Vargas, L., Hazarika, G., Culpepper, R., Butler, K.R.B., Shrimpton, T., Szajda, D. & Traynor, P. 2018, 'Mitigating Risk while Complying with Data Retention Laws', paper presented to the *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, Toronto, Canada.
- Viinikkala, M. 2004, 'Ontology in information systems'.
- Virgilio A.F. Almeida, D.D., Monteiro 2015, 'Governance Challenges for the Internet of Things'.
- Wand, Y. & Weber, R. 2004, 'Reflection: Ontology in information systems', *Journal of database management*, vol. 15, no. 2, pp. 3-6.
- Weber, K., Otto, B. & Österle, H. 2009, 'One size does not fit all---a contingency approach to data governance', *Journal of Data and Information Quality (JDIQ)*, vol. 1, no. 1, p. 4.
- Weber, R. 2016, 'Governance of the Internet of Things—From Infancy to First Attempts of Implementation?', *Laws*, vol. 5, no. 3, p. 28.
- Weill, P. & Ross, J.W. 2004, *IT governance: How top performers manage IT decision rights for superior results*, Harvard Business Press.
- Weinman, J. 2015, *Digital disciplines: Attaining market leadership via the cloud, big data, social, mobile, and the Internet of things*, John Wiley & Sons.
- Welfare, A.I.o.H.a. 2019, *Data Governance Framework*
<<https://www.aihw.gov.au/getmedia/a10b8148-ef65-4c37-945a-bb3effaa96e3/AIHW-Data-Governance-Framework.pdf.aspx>>.
- Wimmer, M.A., Boneva, R. & Giacomo, D.d. 2018, 'Interoperability governance: a definition and insights from case studies in Europe', paper presented to the *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, Delft, The Netherlands,
<<https://doi.org/10.1145/3209281.3209306>>.
- Wright, J.M. & Jones, G.B. 2018, 'Harnessing the Digital Exhaust: incorporating wellness into the pharma model', *Digital biomarkers*, vol. 2, no. 1, pp. 31-46.
- Wróbel, A., Komnata, K. & Rudek, K. 2017, 'IBM data governance solutions', *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC)*, pp. 1-3.
- Yang, L., Li, J., Elisa, N., Prickett, T. & Chao, F. 2019, 'Towards Big data Governance in Cybersecurity', *Data-Enabled Discovery and Applications*, vol. 3, no. 1, p. 10.
- Yannuzzi, M., Lingen, F.v., Jain, A., Parellada, O.L., Flores, M.M., Carrera, D., Pérez, J.L., Montero, D., Chacin, P., Corsaro, A. & Olive, A. 2017, 'A New Era for Cities with Fog Computing', *IEEE Internet Computing*, vol. 21, no. 2, pp. 54-67.
- Yebeles, J. & Zorrilla, M. 2019, 'Towards a Data Governance Framework for Third Generation Platforms', *Procedia Computer Science*, vol. 151, pp. 614-21.
- Young, M., Rodriguez, L., Keller, E., Sun, F., Sa, B., Whittington, J. & Howe, B. 2019, 'Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing', paper presented to the *Proceedings of the Conference on Fairness*,

Accountability, and Transparency, Atlanta, GA, USA,
<<https://doi.org/10.1145/3287560.3287577>>.

Yulianto, S., Lim, C. & Soewito, B. 2016, 'Information security maturity model: A best practice driven approach to PCI DSS compliance', *2016 IEEE Region 10 Symposium (TENSYP)*, IEEE, pp. 65-70.

Zachman, J.A. 1987, 'A framework for information systems architecture', *IBM systems journal*, vol. 26, no. 3, pp. 276-92.

Appendix A: Ethics Approval

Ethics is an integral part of IS Research (Creswell 2009) where data is collected from people. It governs the moral responsibilities of the researcher and his/her obligation as to what is morally acceptable behaviour (Kimmel 1988) during research to safeguard the rights of the respondents.

As part of the University's obligation to the Australian Code for the Responsible Conduct of Research and the National Statement on Ethical Conduct in Human Research, it is UTS policy that any research conducted by UTS staff or students that involves humans must receive ethical approval from UTS Human Research Ethics Committee (HREC) before proceeding. As part of this research's data collection requirement, an application for Ethics approval was placed before the UTS FEIT HREC panel. It comprised all the survey or interview questions.

Below is the Ethics approval email.

research.ethics@uts.edu.au
Wed 11/14/2018 11:57 PM

Asif Gill;
Avirup Dasgupta;
Laura McLean;
Renee Estrella;
Sofia Haidar

□

Dear

Applicant

Your local research office has reviewed your application titled, "DG in IoT", and agreed that this application 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 (<http://www.gsu.uts.edu.au/policies/research-management-policy.html>).

Your approval number is UTS HREC REF NO. ETH18-3165.

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 can be found at <https://staff.uts.edu.au/topic/sub/Pages/Researching/Research%20Ethics%20and%20Integrity/Human%20research%20ethics/Post-approval/post-approval.aspx#tab2>.

- The Principal Investigator will promptly report adverse events to the Ethics Secretariat (Research.Ethics@uts.edu.au). 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.

We also refer you to the AVCC guidelines relating to the storage of data, which require that data be kept for a minimum of 5 years after publication of research. However, in NSW, longer retention requirements are required for research on human subjects with potential long-term effects, research with long-term environmental effects, or research considered of national or international significance, importance, or controversy. If the data from this research project falls into one of these categories, contact University Records for advice on long-term retention.

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 Research.Ethics@uts.edu.au.

REF: 12a

Appendix B: Survey Questionnaire

The following is the sample of the survey questionnaire that will be used to record the response of the research participants.



Survey: IoT Data Governance

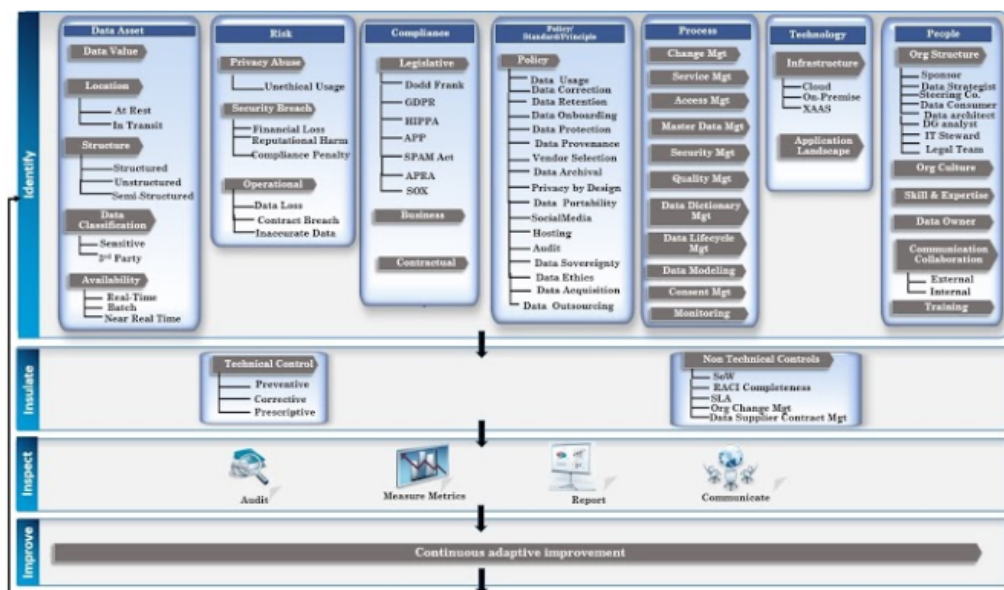
No.	Please provide your response to the following questions	Project Manager
PEOPLE: Data Values and Ownership		30%
1	The team knows the key business questions and goals that are driving the data collected from sensors	30%
2	The value of the data collected is ascertained periodically using established methods	50%
3	The team knows what percentage of the data captured is used	30%
4	The team knows exactly what "data" is captured and how it is used by 3rd party	10%
5	There is formal procedure in place to handle requests for access to personal data, including their purpose	10%
6	Are the roles of Data Steward, Data Owner, Data architect, Business Architect documented and appointed	30%
7	Consent of using sensitive data is acquired explicitly.	50%
RISK: Procurement Management		50%
1	Risk analysis was carried out prior to outsourcing to ensure suitability of the chosen outsourcing partner	10%
2	Procurement processes were clearly communicated to the delivery team	90%
Process: Data Operationalisation Process		53%
1	There is a system of records maintained for various type of data	30%
2	Data is anonymized when distributed to vendors	30%
3	Project team knows how data moves between systems including what ETL is applied in the process	70%
4	Data is destroyed at the end of its lifecycle	70%
5	There was a good documentation describing the tools, process and techniques used for data storage	90%
6	Proper access management is in place which determines who has access to data and how data is accessed	50%
7	Teams accountable for maintaining and operating the data stores (stand-alone, linked to a production application) are captured in RACI matrix	30%
REGULATIONS: Governance Risk and Compliance Management		50%
1	Clear defined rules exist to systematically capture, monitoring, and measure data assets	30%
2	Service levels agreement cover timeliness of data delivery, synchronization between copies of the data between parties in the data/service value chain, destruction of data, usage of data as per intended purpose, data breach notification process	30%
3	Verification and validation of source data is done frequently using automated process	50%
4	Data Privacy Policies exist and enforced	90%
5	Sophisticated solutions are used to ensure that data is kept secure regardless of whether it is controlled by an application system including mobilephones, copied to a test or training	90%
6	Is there a documented User Authorization process/User Management in place that ensures that authorizations for third parties are time-limited and ensures supervisor is traceably informed on any assignment, blocking or change of user rights	50%
7	Yearly identification, analysis and documentation of regulatory requirements applicable to IoT data services are documented	50%
8	Impact to existing and historical data and data processes if a given system change is implemented is well understood by project team	10%
TECHNOLOGY: Device Management		40%
1	Devices are not accessible by outsiders in physical space	70%
2	Documented process in place that ensure access IoT is only granted based on a unique user ID and password (principles of identity)?	10%

Assessment of the Framework

Please submit your feedback regarding the 4I Framework on Data Governance for Digital Ecosystems. This is created as part of a research project being conducted by Mr. Avirup Dasgupta (Email: Avirup.Dasgupta@student.uts.edu.au, researcher at the School of Software, University of Technology Sydney, Ultimo NSW 2007, Australia.)

High Resolution Photo can be found <https://drive.google.com/open?id=1DzmhBighAt9MxOlscD1okY51o1qVXxlV>

The 4I Framework



The 4I Framework

Questions

1. Do you agree with the statement that the overall four phases (Identify, Insulate, Inspect and Improve) in the framework are pertinent? If not, please make suggestions for improvements, deletions, additions etc. in Question 10 later *

Identify	Insulate	Inspect	Improve
Key risks, requirements , people , process, policy , technology and context	Precautionary measures taken to prevent lapses using technologies and non-technical risk remediation techniques	Toolkits such as maturity models, audit mechanisms, software agents required to continuously monitor, report and assess the Data Maturity	Focusses on continuous improvement

- ☐ Strongly Agree
- ☐ Agree
- ☐ Disagree
- ☐ Strongly disagree

2. Alright , the next few questions are about the Identify Phase. How well do you think the Identify phase of this framework covers all the dimensions which are critical to a successful digital ecosystem ? *

Dimension	Attributes
Data Asset	Benefits and salient features of the data
Risk	Factors that influence the end-user, Service Provider and Device manufacturer and impact the Organization financially and reputational
Compliance	Legislative requirements (e.g. User Consent), Business guidelines and Contractual obligations
Policy	Standards, principle and prescribed guidelines to manage data throughout its lifecycle.
Process	Various interfaces and functionalities which deliver a solution.
Technology	Infrastructure, Platforms, Applications, Software and Analytics which act as enabler for digital ecosystem.
People	Stakeholders such Data Consumer, Data producer, Service provider, Network provider and their responsibilities in digital ecosystem. Additionally the people dimension includes Communication, Organization and leadership structure.

1

2

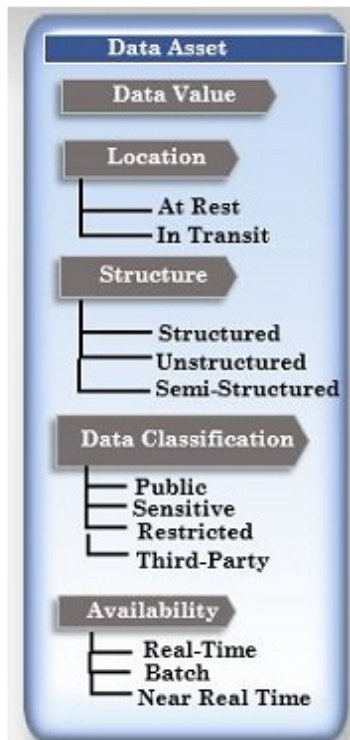
3

4

5



3 a. To what extent would you agree that the following Data asset components of the framework are relevant? *



1

2

3

4

5

☐☐☐☐☐

Are there any elements which you want to add here?

3b. To what extent would you agree that the following Risk components of the framework are relevant? *



1

2

3

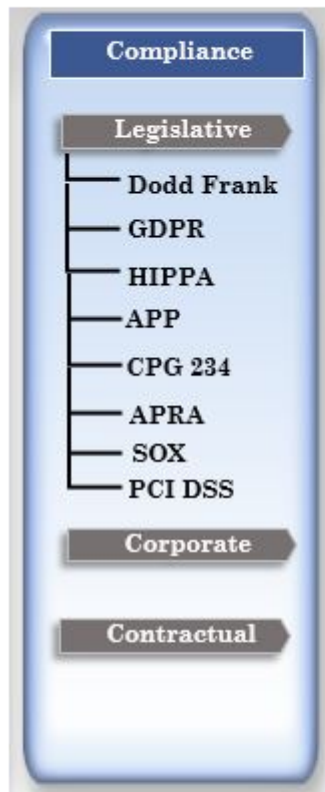
4

5

☐☐☐☐☐

Are there any elements which you want to add here?

3c. To what extent would you agree that the following Compliance components of the framework are relevant? *



1

2

3

4

5



Are there any elements which you want to add here?

Your answer

3d. To what extent would you agree that the following Policy component of the framework are relevant? *

Policy/ Standard/Principle	
Policy	
—	Data Usage
—	Data Correction
—	Data Retention
—	Data Onboarding
—	Data Protection
—	Data Provenance
—	Vendor Selection
—	Data Archival
—	Privacy by Design
—	Data Portability
—	SocialMedia
—	Hosting
—	Audit
—	Data Sovereignty
—	Data Ethics
—	Data Acquisition
—	Data Outsourcing

1

2

3

4

5

☐☐☐☐☐

Are there any elements which you want to add here?

3e. To what extent would you agree that the following Process component of the framework are relevant? *

Process
Change Mgt
Service Mgt
Access Mgt
Master Data Mgt
Security Mgt
Quality Mgt
Data Dictionary Mgt
Data Lifecycle Mgt
Data Modeling
Consent Mgt
Monitoring

1

2

3

4

5

☐☐☐☐☐

Are there any elements which you want to add here?

3f. To what extent would you agree that the following Technology component of the framework are relevant? *



1

2

3

4

5

☐☐☐☐☐

Are there any elements which you want to add here?

3g. To what extent would you agree that the following People component of the framework are relevant? *



1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Are there any elements which you want to add here?

Next couple of Questions are on Insulate Phase.



4. The non technical controls like Statement of Work(SoW), RACI Matrix, completeness. Service Level agreement, Organisation Change Management ,Contract management are relevant *

☐ Yes

☐ No

What else do you think can be added?

Your answer _____

5. "The Technical controls are relevant ". Do you agree? *

Choose ▼

6. Do you agree that the framework is clear, well thought out and easy to understand?

Long answer text _____

7. Did you think this framework is important and provide a roadmap to implement data governance initiatives in your organisation ?

Short answer text _____

8. Are there any modifications or improvements which you can suggest ? It will be fantastic, if you can also give the reasons also

9.. All things considered, how would you rate the practicality of this framework as a blueprint for your Data Governance initiatives?

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thanks!!

We'll use your feedback to improve the framework, and get in touch if we need to check anything with you.

Your Name and your IT role ^{*}

It's just to keep your details for audit trail purposes, not for spam.

Appendix C: Recruitment Email

Dear Sir,

I am a PhD student at the University of Technology, Sydney.

I am conducting research in the area of Information Systems. I developed a framework for governance around data in the digital ecosystem and would welcome your assistance. In order to evaluate the design and applicability of the framework, I would like to request you to participate in my research, review the framework and provide your feedback via a form. The research review and evaluation will involve filling up an online form and will not take more than 30 minutes of your time. I kindly request you to participate in this research because of your expertise in the field of software engineering.

This research has been funded by the Commonwealth Government as student funding help for higher education and research students.

I am looking forward to hear from you. I would be glad to provide more information, if required. Further, you may also contact UTS Graduate Research School and/or my supervisor Dr. Asif Q Gill (School of Software, University of Technology Sydney; Ultimo NSW 2007; Australia Asif.Gill@uts.edu.au).

You are under no obligation to participate in this research. In case you chose to kindly participate and help us evaluate the research output artefact, your contribution will be anonymous. No personal information about yourself will be collected or retained.

If you accept to participate, please sign and return the Consent form . I will then send you invitation to the survey form by email separately.

Yours sincerely,

Mr. Avirup Dasgupta

School of Software

University of Technology Sydney

Ultimo NSW 2007, Australia

avirup.dasgupta@student.uts.edu.au

+61 

NOTE:

This study has been approved by the University of Technology, Sydney Human Research Ethics Committee. If you have any complaints or reservations about any aspect of your participation in this research which you cannot resolve with the researcher, you may contact the Ethics Committee through the Research Ethics Officer (ph: +61 2 9514 2478 Research.Ethics@uts.edu.au), and quote the UTS HREC

reference number. Any complaint you make will be treated in confidence and investigated fully and you will be informed of the outcome.

Appendix D: Consent Form

DG driven framework for the digital ecosystems

I agree to participate in the research project on DG Driven framework for the digital ecosystem (UTS HREC REF NO. ETH18-3165) being conducted by Mr. Avirup Dasgupta (Email: Avirup.Dasgupta@student.uts.edu.au, researcher at the School of Software, University of Technology Sydney, Ultimo NSW 2007, Australia.).

I understand that funding for this research has been provided by Commonwealth Government.

I freely agree to participate in this research project and understand that I am free to withdraw at any time without affecting my relationship with the researchers or the University of Technology Sydney.

I understand that I will be given a signed copy of this document to keep.

I agree to:

- ☐ Receive the online survey form by email
- ☐ Participate in the questionnaire survey
- ☐ The collection of anonymous data from my response

I agree that the research data gathered from this project may be published in a form that:

- ☐ Does not identify me in any way
- ☐ May be used for future research purposes

I am aware that I can contact Mr. Avirup Dasgupta if I have any concerns about the research.

_____	_____
Name and Signature (participant)	Date

Avirup Dasgupta

_____	_____
Name and Signature (researcher)	Date

Appendix E: Versions of the 4I Framework

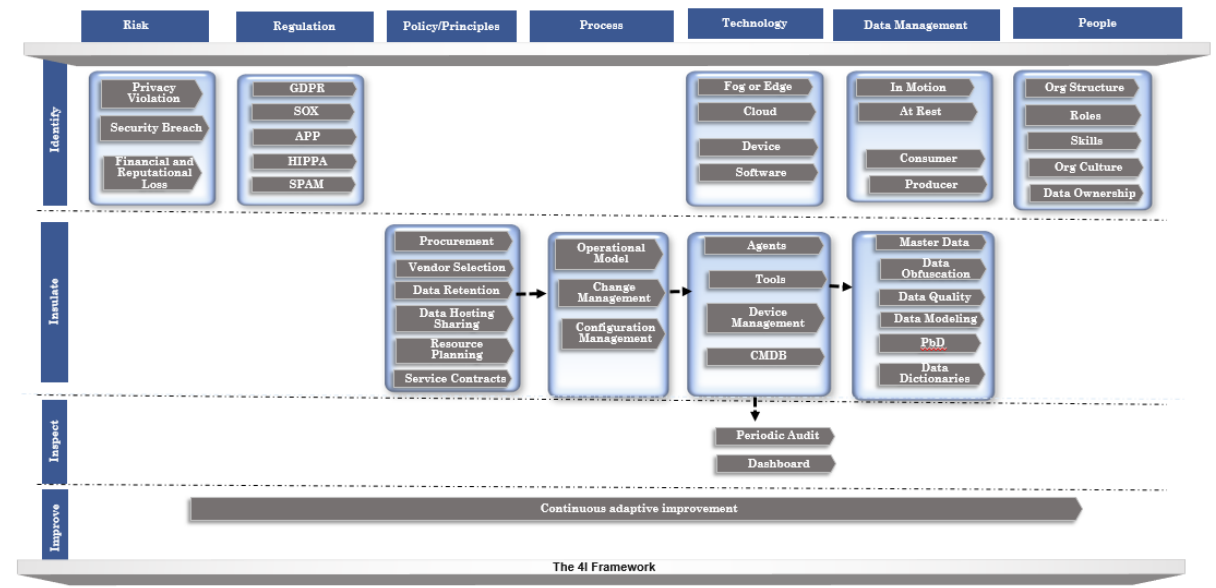


Figure F-1: Version 1.0 of the 4I Framework

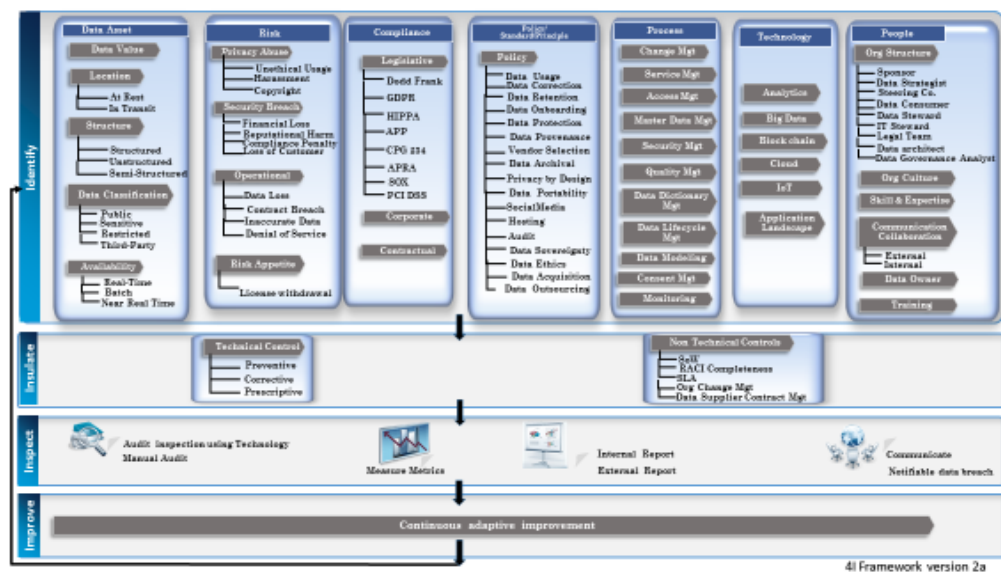


Figure F-2: Version 2.0 of the 4I Framework