DATA GOVERNANCE MATURITY MODEL: A CASE OF THE KENYA DEPARTMENT OF DEFENCE

GILLY GITAHI GATHOGO

A RESEARCH THESIS SUBMITTED TO THE DEPARTMENT OF DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY IN THE SCHOOL OF COMPUTING AND MATHEMATICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN CYBER SECURITY OF THE COOPERATIVE UNIVERSITY OF KENYA

November 2025

DECLARATION

Declaration by Candidate

This research thesis is my original work, and has not been presented to any other examination body of higher learning institution. Therefore, no part of this work should be reproduced without my consent or that of Cooperative University.

Signature …… …… Date 20 Nov 2025

Gilly Gitahi Gathogo

MCSC01/6012/2022

Recommendations by the supervisors

This research thesis has been submitted for examination with our approval as University Supervisors.

Signature …… …… Date …..22 Nov 2025.....

Prof. Simon Maina Karume

School of Mathematics & Computer Science

Cooperative University

Signature … …. Date…....23 Nov 2025.....

Dr. Josphat Karani,

School of Pure and Applied Sciences

Kirinyaga University of Kenya

DEDICATION

I dedicate this work to my beloved wife Josephine Wanjiru and my children Myles & Leo. I also dedicate it to my parents, Mr and Mrs Stephen Gathogo.

ACKNOWLEDGEMENTS

I want to begin by giving thanks to God for being in my life and for His unwavering guidance, without which I would not have had the strength and discernment to embark on this academic journey. Additionally, I want to sincerely thank my supervisors, Dr. Josphat and Prof. Simon Karume, for their unwavering support. Their guidance and assistance have had a significant impact on this study.

I want to thank my family for their unwavering support as I pursue my academic objective. I am fortunate to have a supportive and loving family. To my classmates and coworkers: your company and common experiences have made my academic journey even more enjoyable. I am grateful for the relationships and collaborative spirit that have developed during our time together.

Finally, I would like to express my gratitude to Cooperative University for providing the academic environment and resources needed for this study.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

COBIT ……Control Objectives for Information and Related Technologies

DOD……..Department of Defence

DGME ……Data Governance Maturity Evaluation Model

DGMI ……Data Governance Maturity Index

DPA ……..Data Protection Act

GDPR ……...General Data Protection Regulation

IBM ……International Business Machines

ISO …… International Standards Organization

KPI …… Key Performance Indicator

MCDA ……Ministry, Counties, Department, Agencies

MI ……Maturity Index

Definition of Terms

Data:	Any data, whether organized or unstructured, that an organization collects, processes, manages, stores, and uses to further its strategic goals and operational goals.
Data Governance:	A set of practices, processes, policies, guidelines, and standards that businesses implement to manage their data assets efficiently and responsibly.
Model:	A representation or abstraction of an idea, concept or process designed to capture essential aspects of the system while omitting unnecessary details in order to make it easy to understand.
Maturity Index:	A measuring or assessment scale that demonstrates the level of progress accomplished by an organization in relation to the problem or element it is meant to address.
Maturity Model:	A tool used by an organization to assess the success of a project and their capacity for ongoing improvement.
Web Application:	Any piece of software that resides outside of a device's memory and is accessed via a network connection using the HTTP (Hypertext Transfer Protocol).

ABSTRACT

Data is very important to military groups for planning, gathering intelligence, carrying out missions, and making decisions. This study created a Data Governance Maturity Model (DGMM) custom to DOD operations. The study entailed a comprehensive analysis of current maturity models and data governance frameworks, with an emphasis on modifications most appropriate for military institutions. A DGMM prototype for easy access and evaluation was created and put into use based on the results. In order to accomplish these goals, the model's data governance components were given equal weights in line with ISO/IEC 38500-1 principles that guided the model development. A mixed-method approach that combined qualitative (interviews) and quantitative (surveys) data collection techniques was used in data collection. Design Science Research (DSR) informed the methodology in derivation of the model as well as Principal Component Analysis (PCA) which was used in dimensionality reduction of the selected Key Performance Indicators (KPIs). 117 DOD employees was the sample size which got a 91.3% return rate. Required ethical considerations were followed throughout the proposed study. This study successfully identified essential KPIs required in developing a Data Governance Maturity Model, derived the model and validated it within DOD. In addition to enhancing data governance within the DOD, the proposed study aimed to provide a useful framework for other government ministries and agencies functioning in comparable settings.

Keywords: Data Governance, Model, Maturity Model, Maturity Index

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the Study

The amount of data generated globally has increased exponentially over the past decade, driven by the growth of social media, the Internet of Things (IoT), e-commerce, big data platforms, and the widespread availability of high-speed internet (Alhassan et al., 2016). As data volumes and complexity surge, organizations face a critical choice: adopt structured data governance to harness data's strategic value or succumb to information overload (Rivera et al., 2017). Data now plays a central role in informing policy decisions and influencing key domains such as politics, economics, sustainable development, and national security (MacFeely et al., 2022). Although data offers substantial benefits, it also poses risks when mismanaged, it can be weaponised by hostile actors, undermining organisational and national security. This escalating data landscape underscores the need for regulated, fair, and accountable data management practices.

In defence and military institutions, data is a mission-critical asset underpinning intelligence gathering, mission planning, operational readiness, and decision-making. However, defence environments continue to face structural and technological barriers that constrain effective data governance. Ceruti (2003) identifies persistent challenges within defence information systems, including legacy systems, limited data accessibility, weak data aggregation mechanisms, inadequate joint standards, interoperability difficulties, and limited reuse of information system components. Legacy systems often outdated, siloed, and costly to modernize exacerbate fragmentation in defence environments where units and operations are geographically dispersed, creating inconsistent data practices across the organization.

Despite the proliferation of data governance research, no single definition has been universally adopted within government institutions (Yebenes & Zorrilla, 2019). Weber et al. (2009) define data governance as a *company-wide framework for assigning decision rights and responsibilities to ensure that data is managed as an organisational asset*. This conceptualisation highlights the need for alignment between data governance practices and an organisation's mission, strategy, norms, and culture. It also underscores that data governance comprises multiple components roles, processes, policies, controls, and metrics that vary across organisations.

Maturity models provide a structured means for assessing and improving such governance systems. Permana and Suroso (2018) describe a maturity model as a tool used to develop, assess, and refine an expansive programme. In this study, the Data Governance Maturity Model (DGMM) is conceptualised as a systematic tool for evaluating and improving data governance practices across the organisation. Applying a DGMM enables objective assessment, identification of strengths and weaknesses, and prioritisation of interventions. For the Kenya Department of Defence, such a model provides a practical mechanism for enhancing data management capabilities, strengthening compliance, and aligning data practices with mission and strategic objectives.

In this study, the Data Governance Maturity Index (DGMI) is introduced as a quantifiable metric for assessing the extent to which an organisation's data governance practices comply with regulatory requirements and internal policies. Different formulations of a Maturity Index (MI) may be used depending on an organisation's data goals, governance objectives, and selected indicators. As Nadal et al. (2022) note, data governance practices vary significantly across organisations and are often implemented inconsistently. Each organisation develops its own data architecture and data flows based on its operational needs, constraints, and priorities. Therefore, a tailored DGMI provides a context-specific and evidence-based approach to assessing data governance maturity within the Kenya Department of Defence.

1.2 Statement of the Problem

The exponential growth in data volume and complexity presents a critical juncture for organizations worldwide: to adopt structured data governance and harness data's strategic value or succumb to information overload and inherent risks (Alhassan et al., 2016; MacFeely et al., 2022). This challenge is particularly acute within defence and military institutions, where data is a mission-critical asset underpinning intelligence, operational planning, and national security. While the Kenya Department of Defence (DOD) has initiated progress in establishing data governance procedures, its operational environment is fundamentally distinct from other government Ministries, Counties, Departments, and Agencies (MCDAs).

The DOD's structure is uniquely complex, comprising separate branches the Army, Air Force, and Navy each with distinct missions, technologies, and data priorities. This creates significant

challenges for establishing a unified, interoperable data governance framework. Furthermore, the department's operational environment is characterized by stringent security protocols, the handling of classified information, and the necessity for secure data sharing across geographically dispersed commands (Ceruti, 2003; Orfanus et al., 2016). Existing data governance maturity models, which are predominantly designed for corporate or general public sector contexts (Alhassan et al., 2019; Were & Moturi, 2017), fail to adequately account for these defence-specific imperatives, such as mission assurance and joint-force interoperability.

This gap between a recognized organizational need and the lack of a suitable solution defines a classic class of problems addressed by the Design Science Research (DSR) paradigm. DSR is a pragmatic research paradigm that seeks to create and evaluate innovative artifacts in this case, a tailored maturity model to solve identified organizational problems (Hevner et al., 2004; vom Brocke et al., 2020).

The core problem, therefore, is the absence of a defensible, context-specific Data Governance Maturity Model capable of empirically assessing and guiding the improvement of data governance practices within the unique, high-stakes environment of the Kenya Department of Defence. Consequently, this study was initiated to develop and validate a bespoke Data Governance Maturity Model (DGMM) using a Design Science Research approach. The model aims to provide the Kenya DOD with a precise mechanism to evaluate its current data governance maturity, identify critical gaps, and prioritize strategic interventions, thereby enhancing data-driven decision-making and reinforcing national security.

1.3 General objective

The main objective of this study was to develop a maturity model for measuring data governance maturity in the Kenya department of defence.

1.3.1 Specific Objectives

The specific objectives that guided the research study were:

(i) To investigate the Key Performance Indicators (KPIs) required for measuring Data Governance Maturity levels in the Kenya Department of Defence.

(ii) To derive a Data Governance Maturity Model (DGMM) model for measuring data governance maturity levels based on the KPIs identified.

(iii) To implement a prototype model as a Prototype.

(iv) To validate the model at the Kenya Department of Defence.

1.4 Research Questions

The study employed the following research questions:

i) What are the KPIs required to measure data governance implementation within the Kenya Department of Defence?

ii) How can the model be derived?

iii) How can the model be implemented?

iv) How can the model be validated?

1.5 Significance of Study

The study was significant in that the findings of this research informed the Department of Defence's strategic leadership about the levels of success in data governance. Using this information, the leadership was able to make informed decisions in military operations. It also identified areas where data governance practices required improvement to ensure better decision-making outcomes.

The Department of Defense's strategic leadership used the results of this study to help create data governance policies, strategies, and frameworks that were in line with the Department's goals. The study also added to the body of work on how to measure the maturity of data governance. This deeper understanding led to more new ideas, which will change the way militaries and defense-related agencies handle data in the future.

1.6 Justification of the Study

The Department of Defence operates within a highly specialised and security-sensitive operational environment that differs fundamentally from the business sector. Its branches, units, and formations handle mission-critical functions such as intelligence collection, operational planning, situational awareness, and defence readiness, each supported by distinct data types, system architectures, and classified information protocols. In this context, the term operational

environment refers to the complex, multi-layered ecosystem in which defence data is generated, processed, shared, and secured across geographically dispersed commands and mission domains, often under strict confidentiality and interoperability constraints.

Because of this environment, existing data governance frameworks and maturity models of which majority primarily corporate and business focused , do not sufficiently account for defence-specific factors such as classified data handling, joint-force interoperability, mission assurance, or operational security. The Kenya Department of Defence also follows a Data Sharing Pools (DSPs) model, in which data is distributed across specialized commands and must be shared securely and selectively based on clearance levels and mission relevance. These structural and security realities create a distinctive governance landscape that is not addressed by generic data governance maturity tools.

Therefore, a custom Data Governance Maturity Index (DGMI) is essential ,one aligned to the Department's unique Key Performance Indicators (KPIs), operational workflows, and regulatory obligations. Such a tailored DGMM fills the specific research gap identified in this study: the absence of a defence sector-appropriate, empirical measurement tool capable of evaluating the maturity, compliance, and effectiveness of data governance practices within the Department of Defence. Developing this model provided a precise mechanism for assessing governance performance, guiding improvement initiatives, and reinforcing mission assurance in a high-security environment.

1.7 Scope of the study

This study aimed to explore the Key Performance Indicators (KPIs) essential for assessing the maturity of the data governance within the Kenya Department of Defence. It entailed developing a Data Governance Maturity Model (DGMM) founded on the recognized KPIs. Furthermore, the research included the deployment of a prototype model as a proof of concept and its later validation within the Kenya Department of Defence.

1.8 Limitations and Delimitations of the Study

This study recognized that the Department of Defense's (DOD) operations are sensitive and confidential. Therefore, it was considered essential to follow bureaucratic procedures during the time-consuming data collection process. Additionally, the unique and specialized nature of defence

operations limited the study's ability to be externally generalized, making direct applicability to other sectors difficult.

Employees of the Department of Defence, including data managers and other pertinent stakeholders, were involved in the study from the beginning to guarantee the collection of useful and pertinent data. Additionally, it developed systematic data collection methods, surveyed and interviewed experts who had in-depth knowledge of data governance procedures, and examined publicly available reports, policy documents, and guidelines pertaining to data governance in DOD departments.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Introduction

This chapter presents a comprehensive review of literature dedicated to data governance and maturity models, with a specific focus on their application within organizational contexts. It examines the theoretical foundations, existing frameworks, and empirical studies that guide the design and evaluation of data governance maturity. The analysis further explores the unique requirements and idiosyncrasies inherent in data governance within defence institutions. By synthesizing these insights, this review establishes the foundational framework for identifying appropriate Key Performance Indicators (KPIs) and developing a bespoke Data Governance Maturity Model (DGMM) tailored to the unique circumstances of the Kenya Department of Defence.

2.2 Theoretical Foundations of Data Governance

The conceptualization of data governance has undergone a significant evolution, transitioning from a technical discipline focused on data modeling and quality to a strategic, cross-organizational imperative essential for value creation, risk mitigation, and compliance (Abraham et al., 2019). Early definitions often centered on the specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, consumption, and control of data and analytics (Weber et al., 2009). This foundational perspective established data governance as a system of organizational controls, moving beyond mere technical management to encompass the people, processes, and policies required to treat data as a strategic asset.

In recent years, the scope of data governance has expanded dramatically, driven by the proliferation of big data, artificial intelligence, and heightened regulatory scrutiny. Contemporary scholarship posits data governance not just as a control function but as a critical enabler of digital transformation and data-driven innovation (Otto, 2019). As Alhassan et al. (2019) argue, effective data governance provides the "rules of the road" that allow organizations to leverage data safely and ethically, balancing the exploitation of data's value with the imperative to manage risks related to security, privacy, and reputational damage. This is particularly salient in the context of emerging

regulations like the European Union's General Data Protection Regulation (GDPR) and Kenya's own Data Protection Act (DPA), which have made robust governance a legal necessity, not just a strategic choice (Kevins & Brian, 2022).

The theoretical underpinnings of data governance are further refined by the distinction between its structural and behavioral components. Structurally, it involves the establishment of formal mechanisms such as data stewardship roles, data councils, standardized policies, and defined processes for data lifecycle management (Khatri & Brown, 2020). Behaviorally, it seeks to foster a data-centric culture where data ownership and quality are seen as shared responsibilities across the enterprise (Tallon et al., 2020). This dual nature highlights that successful implementation relies as much on influencing human behavior and organizational culture as it does on implementing formal structures and technologies.

Within the specific context of a defence institution, these theoretical foundations take on a heightened significance. The strategic asset in question is not merely commercial data but often classified information integral to national security. Therefore, the "decision rights and accountability" framework (Weber et al., 2009) must be mapped onto a strict military chain of command, and the "balance of value and risk" (Abraham et al., 2019) is weighted overwhelmingly towards mitigating risks of operational compromise. The defence data environment, characterized by legacy systems, interoperability challenges, and the need for secure data sharing across dispersed commands (Orfanus et al., 2019), presents a unique set of constraints that demand a tailored application of data governance theory. Consequently, a data governance framework for the defence sector must be theoretically grounded in these established principles while being pragmatically adapted to address the sector's unparalleled security requirements and mission-critical operational tempo.

2.3 Synthesis of Data Governance Maturity Models

A maturity model provides a structured means for assessing and improving governance systems. This study reviewed prominent models, including CMMI, IBM, Gartner, and Stanford DGMM. A synthesis of these models reveals common thematic domains critical for maturity assessment, yet also highlights a significant gap: their primary design and application reside in corporate and general public sector contexts (Alhassan et al., 2019; Were & Moturi, 2017).

Table 1: Synthesis of Maturity Model Domains and Defence Context Gap

Thematic Domain	Description (from Literature)	Relevance to Defence Context
Data Quality & Architecture	Ensuring data accuracy, consistency, and reliable structure (DAMA, 2014).	Mission-critical for intelligence accuracy and operational planning.
Data Security & Privacy	Safeguarding data against unauthorized access and ensuring compliance (GDPR, Kenya DPA).	Paramount for protecting classified information and national security.
Data Management & Lifecycle	Managing data from creation to disposal (Herselman et al., 2019).	Essential for audit trails, data sharing pools (DSPs), and records management in a dispersed command structure.
Organizational Framework	Establishing clear roles, responsibilities, and reporting structures (Alhassan et al., 2019).	Needs adaptation to fit military chain-of-command and joint-force interoperability.
Emerging Technologies	Leveraging AI, analytics, and automation for data-driven insights (Pankowska, 2020).	Critical for modern warfare, including Network Centric Warfare (NCW) and AI-driven intelligence.

As illustrated in Table 1, while the core domains are universally relevant, their implementation within a defence environment is fundamentally different. The Kenya DOD operates within a highly specialized ecosystem involving classified data handling, joint-force interoperability, and mission assurance under strict confidentiality constraints (Ceruti, 2003; Orfanus et al., 2016). Generic models do not sufficiently account for these defence-specific factors, creating a clear research gap for a tailored maturity model.

2.4 The Local and Regional Context

The literature on data governance maturity within the Kenyan and broader African public sector is notably sparse. A key regional study by Herselman et al. (2019) developed a Data Governance Maturity Evaluation Model (DGMEM) for South African government departments, focusing on data quality, lifecycle, and security. This study provides a valuable methodological reference but does not address the defence sector.

Locally, the enactment of the Kenya Data Protection Act (DPA), 2019, establishes a critical regulatory framework that any data governance initiative must comply with (Kevins & Brian, 2022). Were & Moturi (2017) explored data governance in Kenyan health regulatory authorities, highlighting challenges like unclear policies and limited resources. However, to the best of our knowledge, no study has developed or applied a data governance maturity model within the Kenya DOD or similar high-security Kenyan government entity. This absence underscores the necessity and novelty of the present research.

2.5 Derivation of Key Performance Indicators (KPIs)

The KPIs for this study were derived through a synthesis of the global literature and a critical analysis of the local context. The universal domains from established models provided the initial theoretical constructs. These were then critically evaluated and refined to ensure their relevance and measurability within the defence environment.

Table 2: Operationalization of Theoretical Constructs into Measurable KPIs

Theoretical Construct (from Literature)	**Defence-Specific Consideration**	**Operationalized KPI for this Study**
Data Quality (DAMA, 2014)	Accuracy of intelligence and operational data.	Data Accuracy (DAC): The correctness and precision of mission-critical data.
Data Security (Weber et al., 2009)	Protection of classified information from hostile actors.	Multi-Factor Authentication (MFA): Securing access with additional authentication layers.
Data Policies (Alhassan et al., 2019)	Formalizing rules for handling classified and sensitive data.	Data Policies & Standards (DPS): Enforcement of policies across the data lifecycle.
Organizational Structure (Herselman et al., 2019)	Aligning with military chain-of-command.	Reporting Structure (RS): Clarity of data governance reporting within the military hierarchy.
Compliance (Kenya DPA, 2019)	Adherence to national data protection law.	Data Protection (DP): Ensuring practices comply with the Kenya DPA.
Emerging Technology (Pankowska, 2020)	Leveraging AI for battlefield advantage.	AI & Analytics (AI): Use of artificial intelligence for data analysis and insights.

This process of operationalization ensured that the final list of KPIs like Data Encryption, Data Governance Team, Strict Enforcement) was not merely a replication of corporate lists but a curated

set of indicators grounded in both global best practice and the specific, high-stakes operational reality of the Kenya DOD.

2.6 Data Governance Maturity Model Key Performance Indicators

Alhassan et al., (2019) employing five paradigm models and through selective coding has defined breaks own Critical Success Factors (SCF) for data governance into seven fundamental components: a company's workforce at the appropriate skill levels, defined data processes and procedures, agile data tools and technologies, clear easy-to-follow formulated policies and standard operating procedures, designated roles and duties regarding data, inclusive data requirements, focused and actionable strategies.

As pointed out in (MoALFC, 2022), data governance serves as a foundational structure for decision-making and aims to achieve several objectives such as: “Providing clarity and consistency on roles and responsibilities, establishing rules for data use such as collection and sharing, Minimize the risks of collecting, storing and using data, Help regulatory expectations to be met, Enhanced decision making, Enhanced communication”.

DAMA framework (Cupoli, Earley, & Henderson, 2014) describes data governance through ten core knowledge areas, each with goals, processes, technology, inputs, suppliers, responsibilities, stakeholders, deliverables, and consumers. Literature varies with terminology, for instance, Ekundayo et al. (2023) refers to them as ‘Specializations’. This study found out that, according to the DAMA framework (Cupoli, Earley, & Henderson, 2014) , Data Governance can only be successful if the following key specialisations are also successful: Data architecture, data modelling, data quality, data integration, master and reference data management, data warehousing, data storage, data security, and document and content management are all important. The study also noted that (PANKOWSKA, 2020) supports that the goal of data governance in a business is to process data as quickly as possible while keeping the best quality possible.

The author says that data processing speed, data privacy, and data consistency are all common problems that need to be fixed in order for data governance to work. According to Cupoli et al. (2014), the following main areas work together to help govern data: data architecture, data quality, data modelling, data storage, data security, data warehousing, business intelligence, data

integration, interoperability, and meta-data management. Alhassan, Sammon, & Daly (2019) also highlights that there haven't been many studies that have looked specifically at what makes data governance mature and successful. Still, it's not easy to measure the success of data governance, and the metrics can be different for different businesses and industries.

A data governance checklist by Cupoli, Earley, and Henderson (2014) focuses on the following metrics: who has the power to make decisions, policies, procedures, and standards for protecting data privacy and security; data inventories; content and records management; data quality control; data access; data security and risk management; and data sharing and dissemination. Hence it's crucial to note that different models had different data governance metrics. This is due to the fact that various organizations have varying opinions about what constitutes effective data governance and what their own priorities are.

2.6.1 Data Governance Key Performance Indicators (KPIs)

According to the Cupoli, Earley, and Henderson (2014) study, some of the metrics were described as the key performance indicators of data-governance (KPIs). These were records management, data-quality control, data access, data security, risk management, data sharing, data dissemination and privacy protection. Other pillars of good governance were said to be business intelligence, data integration, interoperability, data warehousing, data architecture, data storage, data security and meta-data management. Later, KPIs given by the organization DAMA International included Data Architecture, Data Modelling, Data Quality, Master and Reference Data Management, Data Warehousing, Compliance, Document and Content Management and Data Privacy (Cupoli et al., 2014).

The additional data-governance KPIs brought forward by Alhassan et al., in 2019, include Employee Data Competencies, Clear Data Processes, Flexible Data Tools and Technologies (including AI and analytics), Standardised Easy-To-Follow Data Policies and Established Data Roles and Responsibilities. These novel measurements testify to a growing demand of technological as well as human capacity when in looking after present-day data systems.

Pankowska (2020) revealed the quality of data is one of the key factors of effective data governance. Before, data security and data integration in addition to the importance of data quality had been introduced by Cupoli et al. (2014). The Ministry of Agriculture, Livestock, Fisheries and

Cooperatives (MoALFC) released new set of key performance indicators (KPIs) in 2022 with respect to data governance in the areas of clear data rules, minimised data collection risks, data sharing, data storage and data automation. Such parameters indicate the existence of a larger dream concerning the efficiency of the processes accompanied by the minimization of risks.

When comparing the proposed KPIs in the context of different data governance maturity models, one will find that in spite of the fact that each of the frameworks has slightly different metrics to be adhered to, some indicators have been universal across the board. These are data warehousing, data security, data storage and data quality. Therefore, the paper has shown that there is much convergence in the core elements of data governance programmes in various fields. However, there are KPIs that have taken shape as a result of the advancement in technology and as a result, organisations are now able to respond to opportunities as well as challenges that come along with data. Some of them include data automation, analytics, privacy, and compliance.

2.7 Data Governance Maturity Model Design

With the predetermined key performance indicators, the second goal of this study is to come up with a unique maturity model to determine the degree of Data Governance maturity at the Department of Defence in Kenya. This discussion hence reflects on the major existing Data Governance maturity models. Since the Defence sector is exercised in the changing military environment, it is prudent to profile the present military environment and existing facts management trends first before subjecting themselves to the more complex aspects of DM maturity.

2.7.1 DOD Operating Environment

The various elements of the military operating environment coordinate to pursue one overall goal. The three Services are the Army, Air Force and the Navy that each of them falls under United States Department of Defence (Redmond et al., 2015). During peaceful as well as during conflict periods, the corresponding Service has active and reserve forces, and their workers are needed to provide delegated missions. Under non-combatant conditions as well all of the departments have particular abilities and responsibilities to complete certain tasks. A subculture, equipments, training and doctrine are further built in each Service to reflect its particular organisation and purpose, but all these are directed toward fulfilment of national security goals united as a nation.

Military command and control (C2) systems have undergone considerable improvements in the last decade, with the recent technologies having a major impact on the development of military systems that can give commanders thorough intelligence and situation awareness (SA) to aid informed decision-making (Orfanus, De Freitas, & Eliassen, 2016). Unmanned aerial vehicles (UAVs) have found their invaluable use in a contemporary battlefield with the UAVs being targeted at acquiring targets, surveillance, and reconnaissance (Orfanus et al., 2016). The practicability of remotely piloted UAVs has already been proven in the earlier and current instances of deployment of the Kenyan Defence Forces (KDF). Military units around the world are becoming more and more interested to learn how the Network Centric Warfare (NCW) doctrine, which takes advantage of new technologies to strengthen information- and technology-based warfare, can be incorporated in their operations.

Although the military has resorted to an extensive use of advanced technological tools, the peculiarities of the operation environment poses challenges to maintaining uninterrupted interconnection and integration of various systems of communications that do not relate to each other. During peaceful times, the military bases are normally located in remote areas but during wars, they have a higher chance of emerging in enemy territory offering daunting access issues. Orfanus et al. (2016) consider interference and jamming as the informative examples of disruptive phenomena that often interfere with data communications. Such disruptions can be short-term, but they would render useless the technological solutions based on the conventional communications infrastructure.

2.7.2 Maturity Models Design

DataFlux (Dataflux Company, 2019) and Oracle (Oracle, 2021) are technology-focused mostly and the major means by which data can be managed according to them is through the use of technological tools. DMM and IBM models opt to retain the definition of each maturity level in DMM as defined in CMMI with slight changes. Oracle and MD3M localize CMMI to support data-specific elements that supplement, though they do not replace, the current component. However, the authors ascertain that, even after making such refinements, the models have not captured the strategic aspect of data governance; that is, elements of organisational culture, decisional processes, oversight, and alignment with its strategic goals.

2.8 Case Studies

2.8.1 Data Governance Maturity Model for Microfinances

An assessment matrix centred on five domains—Data Principles, Data Lifecycle, Metadata, IT Usage, and Regulatory Compliance—was employed in this model by Rivera et al. (2017). As illustrated in figure 2 below, the assessment criteria were assigned using the Capability Maturity Integration Model (CCMI) (David Patón-Romero et al., 2019) maturity levels: Level 1- Initial, Level 2- Managed, Level 3-Defined, Level 4-Quantitatively Managed, and Level 5-Optimized as shown in figures , and 3 below.

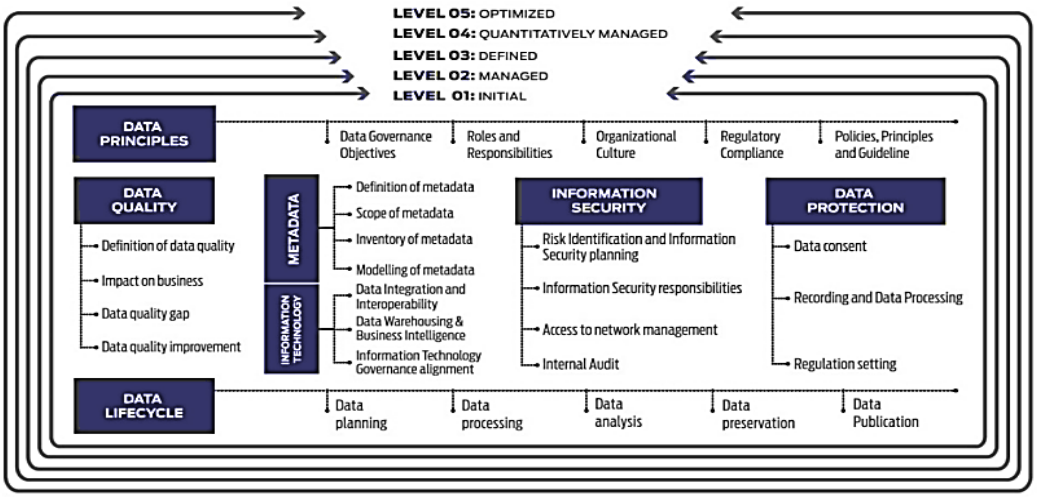


Figure 1: Data Governance Maturity

Source: (Rivera et al., 2017)

Maturity weightage is assigned based on compliance levels to the processes set at each of the five domains as shown in the figure 3 and 4 below:

		Level 1 Initial			Nivel 2 Managed			Nivel 3 Defined			Nivel 4 Quantitatively Managed			Nivel 5 Optimized		
Domain	Assessment Criteria	NI	PI	I	NI	PI	I	NI	PI	I	NI	PI	I	NI	PI	I
Data Principles	Data Principles		✓													
	Data Lifecycle					✓										
	Metadata						✓									
	Information Technology								✓							
	Regulatory Compliance						✓									

Indicator	Description	Weight (w)
NI	Not implemented	0.00
PI	Parcially implemented	0.50
I	Implemented	1.00

Figure 2: Maturity Assessment Matrix

Score	Maturity Level
0 > Score <= 1	Level 1: Inicial
1 > Score <= 2	Level 2: Managed
2 > Score <= 3	Level 3: Defined
3 > Score <= 4	Level 4: Quantitavely Managed
4 > Score <= 5	Level 5: Optimized

Figure 3: Score Equivalences

2.8.2 Stanford Data Governance Maturity Model (DGMM)

The Stanford Maturity Measurement Tool, according to Zúñiga et al. (2018) and Al-Ruithe & Benkhelifa (2017), uses both qualitative and quantitative metrics to track how well an organization's data governance procedures are working. While quantitative metrics track the quantity of activities carried out, program participants, and artifacts produced, qualitative aspects evaluate the organization's attributes at various stages of maturity. Each dimension of the tool has a qualitative scale that goes from level one, which is the beginning of a data governance program, to level five, which is the intended goal of data governance in that particular field of study. This model focuses on three foundational components: Awareness, Formalization, Metadata, and three

project components: Stewardship, Data Quality, and Master Data which measure how effectively data governance concepts are applied. People, policies, and capabilities are the three dimensions on which all of these are measured.

Figure 4 below records the scores for each Component-Dimension and computes the average for each row and column in order to evaluate the maturity of qualitative aspects in an organization's data governance program. The organization's level of maturity in each area is indicated by the average score for each Component and Dimension.

Foundational	People	Policies	Capabilities	Average
Awareness	2	2	2	2
Formalization	1	2	1	1.3
Metadata	2	1	1	1.3
Average	1.6	1.6	1.3	
Project	**People**	**Policies**	**Capabilities**	**Average**
Stewardship	2	1	1	1.3
Data Quality	2	2	1	1.6
Master Data	1	1	1	1
Average	1.6	1.3	1	

Figure 4: Qualitative Score Card

According to this model, the organization can create a roadmap for data governance success by assessing the program's baseline maturity levels, establishing short- and long-term goals during the initiation phase, and conducting regular evaluations of the Component-Dimensions." The data governance board sets both short- and long-term maturity goals that determine success. A data governance program may be hampered by the complexity and scope of these projects, according to Permana & Suroso (2018). For this reason, a parallel approach—or rather, two concurrent sets of activities are used to counteract obstacles as illustrated in Figure 6 below.

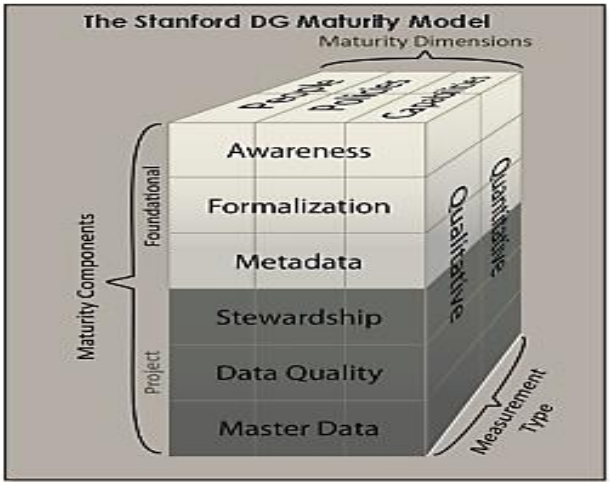


Figure 5: Stanford maturity assessment component

Source: (Permana & Suroso, 2018)

2.8.3 IBM Maturity Model

IBM Data Governance Maturity Model was created through incorporating techniques methods and gathered from the council members (Were & Moturi, 2017). According to Al-Dossari and Sumaili, the IBM model is centered on eleven domains that are categorized into four groups, as illustrated in the following figures 6, 7, 8, and 9:

• *Ooutcames* Domain: ✓ Data risk Management ✓ Creating value,	• Core Domain ✓ Data Quality Management ✓ Information Life Cycle Management ✓ Privacy & Security Management
• Function Domain ✓ Organization Structure & Awardness ☑ Data Stewardship ✓ Policy	• Support Domain ✓ Data Architecture ✓ Classification & Metadata Management ✓ Information Audit & Reporting

Figure 6: IBM Data Governance Domains

Source: (Prasetyo, 2016).

The IBM model maturity levels are borrowed from the scientific Capability Maturity Model (CMM) by Software Engineering Institute are as described in the figure below:

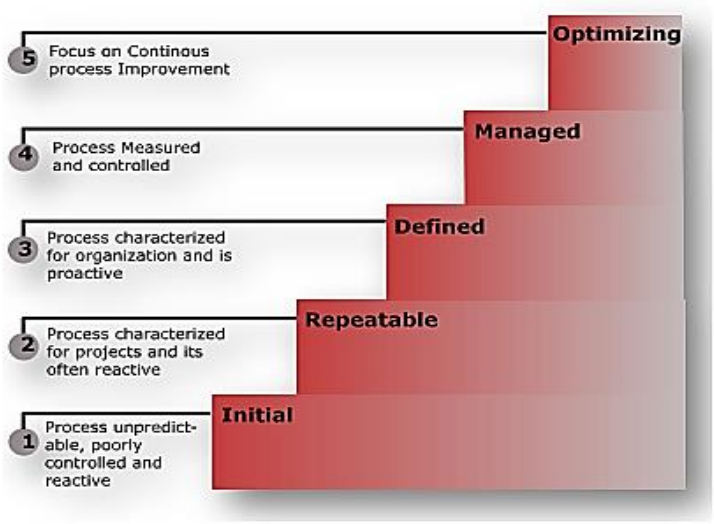


Figure 7: Five Levels of CMM Maturity

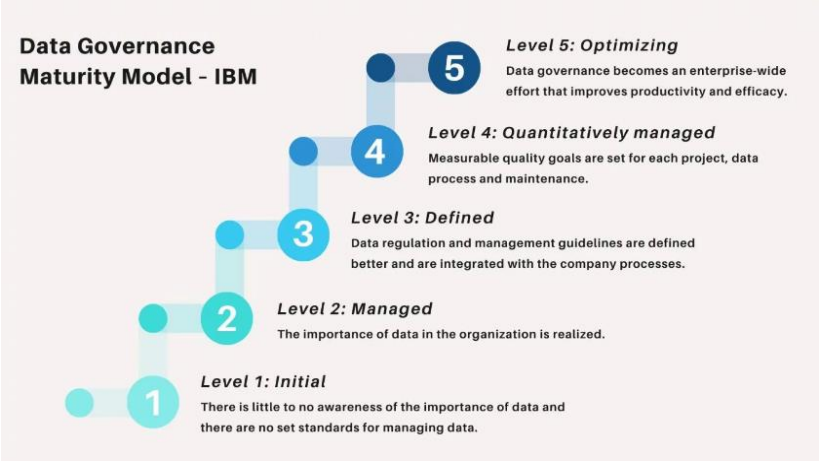


Figure 8: IBM Data Governance Maturity Model

2.8.4 Gartner Maturity Model

The Gartner model (Al-Dossari & Sumaili) encompasses six stages that represent varying levels of maturity.

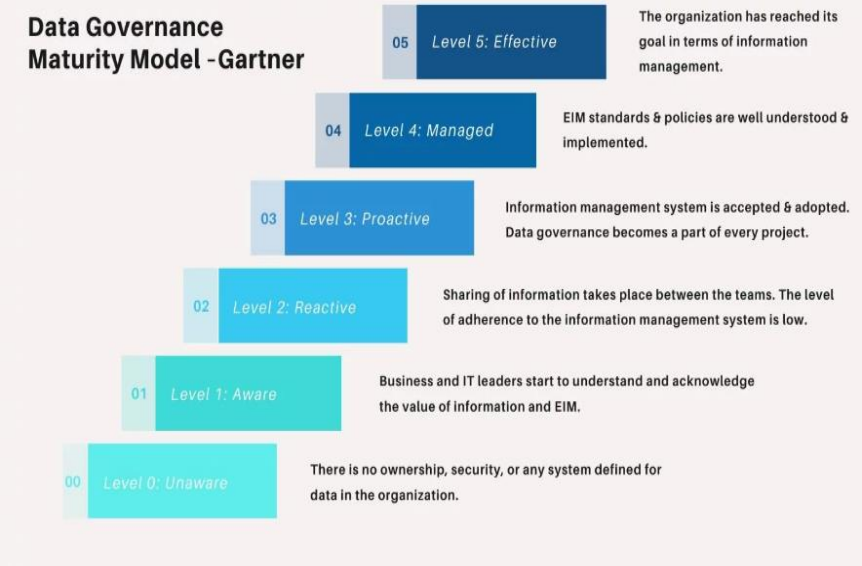


Figure 9: Gartner IEM Model

From the models reviewed, we summarize the maturity levels applied by different maturity models covered earlier as below:

Table 3: Comparative analysis of Maturity Levels

Level	Oracle	Data Flux	IBM	Gartner	DGMM	Stanford	CMMI
0	--	--	--	Unaware			Initial
1	Marginal	Initial	Initial	Aware	Initial	Initial	Repeatable
2		Reactive	Managed	Reactive	Managed	Managed	Defined
3	Stable	Managed	Defined	Proactive	Defined	Defined	Managed
4	Best Practice	Practice	Quantitatively Managed	Managed	Quantitatively Managed	Quantitatively Managed	Optimized
5	Transformational	Strategic Performance	Optimizing	Effective	Optimizing	Optimizing	

It was clear from the models examined that the maturity levels vary according to the model used. According to table 1 above, the scientific Capability Maturity Model (CMM) has been adopted by the majority of Data Governance Maturity Models, according to this study.The main reason why organizations may select different models is because their business objectives differ.Prasety

o(2016) asserts that while some organizations focus on customer satisfaction, others prioritize da ta quality, and still others prioritize customer data confidentiality.

Even though the CMM provides very clear, quantifiable, and easily adjustable steps that are com patible with many organizations, it misses a crucial component of assessing the human capacity of data governance skill sets. Tables 4, 5, and 6 below provide knowledge transfer summaries and the Unaware Level, which is a component of the Gartner Model that aids in measuring the degree of data implementation awareness. The importance of developing a data governance-focused maturity model that considers all facets of an organization's data governance is highlighted by this study. It carefully evaluates the overall maturity of a data governance program and emphasizes the significance of adding a sixth maturity level that complies with ISO/IEC 38505-1 guidelines.

Table 4: Summary of the proposed Six-Level Maturity Model

Levels	***CMMI Maturity Model***	***Gartner Maturity Model***	***Guiding Principles (ISO/IEC 38500-1)***	***Hybrid Maturity Model***	***Maturity Score***	***Description***
Level 1	Initial	Reactive	Conformance	Unaware/None	0	The organization is at a "Unaware" maturity level
Level 2	Managed	Aware	Responsibility	Aware	1	The organization is at a "Aware" maturity level
Level 3	Defined	Defined	Responsibility, Strategy, Acquisition	Defined	2	The organization is at a "Defined" maturity level
Level 4	Quantitatively Managed	Managed	Responsibility, Strategy, Performance, Conformance	Managed	3	The organization is at a "Managed" maturity level
Level 5	Optimizing	Optimized	All (Responsibility, Strategy, Acquisition, Performance, Conformance, Human Behavior)	Optimized	4	The organization is at a "Optimized" maturity level
Level 6			Performance	Mature	5	The organization is at a "Mature" maturity level

Table 5: Six-Level Maturity Model

Maturity Level	Description	Score (X)
Unaware/None	No formal Data Governance processes.	0
Aware	Basic recognition of Data Governance needs.	1
Defined	Formal processes are in place, but not standardized.	2
Managed	Processes are managed, measured, and controlled.	3
Optimized	Processes are continuously improved based on metrics.	4
Mature	Fully integrated Data Governance and Continuous improvement.	5

Table 6: Summary of Models Reviewed

S/No	Model	Year	Refence	Author
1	MD3M	2015	Spruit & Speizka,2015	Spruit & Speizka
2	DDG	2010	Vijay & Brown,2010	Vijay & Brown
3	Oracle	2011	Oracle,2011	Oracle
4	DataFlux	2010	Dataflux Company,2010	Data Flux Company
5	DGMM for Microfinance	2017	Rivera et al.	Rivera et al
6	Stanford Data Governance Maturity Model(DGMM)	2018	Zúñiga, Cruz, Ibañez, Dominguez, & Moguerza,2018	Stanford University's Data Governance Office
7	Capability Maturity Model (CMM)	2019	Herselman, Wayi, & Olaitan,2019	Capability Maturity
9	IBM Data Governance Council Maturity Model (DGMM)	2007	Al-Dossari & Sumaili. Were & Moturi, 2017	IBM
10	Gartner Model (EIM)	2016	Al-Dossari & Sumaili	Gartner

2.9 Developing a Data Governance Maturity Model

The third goal of this study was to implement the Data Governance Maturity Model and use it to measure data governance maturity at the Kenya DOD. In this section on the development of a Data

Governance Maturity Model, the study reviewed literature on a Data Governance Maturity Evaluation Model implemented by a South African government department, the ISO/IEC 38500, and COBIT recommendations on data governance model considerations, Compliance Laws and regulations.

2.9.1 Data Governance Maturity Evaluation Model (DGMEM)

Herselman, Wayi and Olaitan (2019) also define maturity model as a collection of elements organised in a form organising an index of the characteristics of effective processes at various levels of their development. Such models define the scope between successive stages and the means of passing to a certain stage. Maturity model aims at providing a concise, accurate guide on perfecting the desired improvements, and thus facilitates transformation of the current state to a desired level of maturity by a firm.

The present study evaluated the Data Governance focus areas through the interviews conducted across all organizational departments and through the comparison and evaluation of the currents practices of the organization and the recommended international practices frameworks. When developing and testing a conceptual Data Governance Maturity Evaluation Model (DGMEM) applicable to various government departments in South Africa, the review of a study by Herselman et al. (2019) established that the research work handles the three major areas of data quality management, data lifecycle management, and data security and privacy.

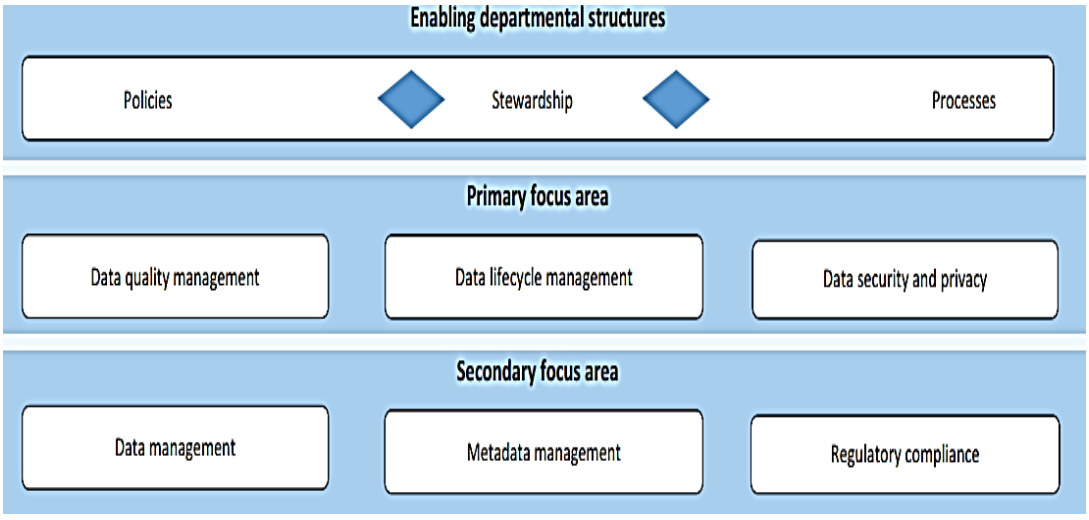


Figure 10: Data Governance Maturity Model

Source: (Herselman et al., 2019)

This study adopted the method by Herselman et al. (2019) that breaks down the Data Governance Maturity Evaluation Model into the three main parts discussed above. It also used the suggestions from COBIT and ISO/IEC 38500, which are shown in Figure 10 above.

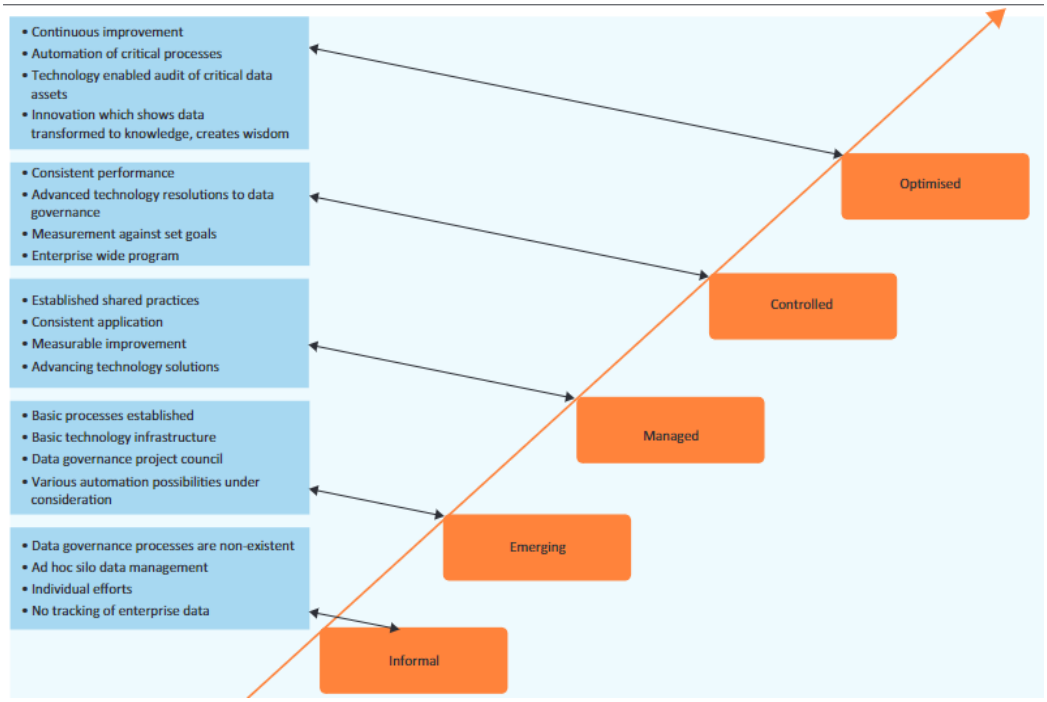


Figure 11: DGMEM maturity levels

Source: (Herselman et al., 2019)

2.9.2 Compliance

The governance of data involves protection of data, legal compliance and use of data protection and compliance in determination of the rights of data (Voss, 2019). The necessity to ensure that the data can be accessible and have a sound quality is one of the core goals of this field. According to Yebenes and Zorrilla (2019), data are an asset that should be handled with extreme care in order to develop their potential fully. Organizations, therefore, need to work out clear policies on data management and ensure these policies are effected through tracking the way data are used, ensuring their quality, maintaining privacy, and complying with legislative regulations like the General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), and the Sarbanes-Oxley Act (SOX).

2.9.3 General Data Protection Regulation (GDPR)

This study established that GDPR brings a lot of new rules for data management, and many of them are based on the fair information practice principles that a committee in the U.S. Department of Health, Education, and Welfare wrote about in a report in 1973 (Voss, 2019). According to Voss (2019), metadata can help with good data governance, but using new business technologies like cloud computing can make it harder to follow the rules. Cloud providers often move data around based on load and other factors, which makes it hard to know exactly where data is at any given time. Good data governance makes it easier for organizations to follow data privacy laws, so we can say that good data governance makes it easier to follow data protection laws. This study aimed at including important metrics from the Data Protection Act and the General Data Protection Regulation (GDPR) that are necessary for comprehensive data governance.

2.9.4 Health Insurance Portability and Accountability Act (HIPAA)

HIPAA indicated that an organization must follow the basic rules of the HIPAA rule, train its staff about them, and create more awareness on them. This is to be done by putting in place rules and procedures that protect a patient's private health information. Following the rules also means you have to do more with your data. According to Aiello et al. (2021), following the GDPR affects all personal information. In contrast, HIPAA only applies to "Protected Health Information" (PHI), which is any data that can directly or indirectly identify a person in connection with their past, present, or future health status.

2.9.5 Kenya Data Protection Act (DPA)

Kenya Data Protection Act Article 31 (c) (b) of the constitution of Kenya is the statutory authorization to the Kenya Data Protection Act. The Act enforced the adoption of Article 31 (c) (d) that operationalized the Article 31 (c) (b) and came into force on 25 November 2019 (Kevins & Brian, 2022). The gauge makes Kenya data-protection regime compliant with the provisions of the General Data Protection Regulation (GDPR) and therefore increases harmony, ease cross-border data transfer, and cooperation between nations in terms of data management.

Particularly, the law builds protection shields that seek to guarantee confidentiality and integrity of personal information. In the case of Kenya department of defence, an institution concerned with the national security and defence, there is a necessity to implement complete compliance with the Kenya Data Protection Act. To achieve these goals, the statute sets a row of the requirements, such

as the data encryption, access control, authentication, data minimization, retention plans, data-impact analyses, training on security-related matters, incident reporting, managing consent, privacy-by-design concepts, checks on vendors, data transfer procedures, and the nomination of a Data Protection Officer. The following table will present a brief description of all the models that were considered in the legislation.

Table 7: Summary of Models reviewed

S/No	Model/Standard	Year	Metrics /KPIs	Author
1	COBIT 5	2012	Data Security, Risk, Continuity, Transparency, Adherence to processes and polices, Regular Audit Assessments, Privacy by Design, Alignment to Business Strategy	COBIT
2	IS0/IEC 38500	2008	Training, Data quality, Data Acquisition, Accountability, Compliance to law, Data reporting, Data management initiatives, Vendor Management, Regular Audit Assessments, Incident Management.	ISO
3	GDPR	2016	Privacy Awareness, Data reporting, Training, Data processing, Data Security, Risk Mitigation, Transparency, Vendor Management, Consent Management, Data Impact Assessment, Data Breach Response Plan, Data Subject Rights, Data Minimization, Data Retention and Disposal, Privacy by Design, Privacy by Design	GDPR
4	HIPAA	1996	Employee awareness, Controls,Data Security,Data Privacy, Data Availalbity, Consent Management. Data Breach Response Plan, Data Subject Rights, Incident Management	HIPAA
5	Kenya DPA	2021	Transparency, Vendor Management, Consent Management. Data Impact Assessment, Data Breach Response Plan, Data Subject Rights, Data Minimization, Data Retention and Disposal, Privacy by Design	Government of Kenya

2.10 Validating the Data Governance Maturity Evaluation Model

The third purpose of the research is to test the validity of Data Governance Maturity Model to the Kenya Department of Defence. The Model was tested out in a Data Governance Maturity Assessment at the Defence department specifically in the Department of Defence. This critique was an overall examination of the extent to which the model was appropriate, its relevance and adjustment in the context of application in the department. Through experimenting with the model in the real world environment, the researchers managed to conclude about its general efficacy and

its ability to address the goals of the department, thus ensuring that the model indeed stays the right and efficient tool to achieve the data governance maturity within the organisation.

2.11 Research Gap

A careful review of global and regional literature on data governance reveals important gaps that motivated the development of the Data Governance Maturity Model (DGMM). First, many existing maturity models (international standards and vendor frameworks) emphasise governance principles at a high level but provide limited operationalisation for context-specific KPIs; this reduces their applicability to organizations operating in low-resource or mixed-regulation environments common in parts of Kenya and sub-Saharan Africa. Second, there is sparse empirical evidence from Kenyan/African organizations demonstrating how specific governance practices map to measurable outcomes like trust, data quality, compliance; most regional studies are descriptive or narrowly focused on privacy/legal compliance without integrating technical, organizational and process KPIs into a single index.

Third, weight assignment and KPI selection in many published indices rely on expert judgement alone, with limited statistical validation or sensitivity analysis increasing subjectivity and reducing reproducibility. Fourth, few studies report systematic validation (factor structure, internal consistency) of the chosen indicators on locally collected data, and even fewer make comparisons across departments or over time. Finally, practical toolkits (dashboards, exportable reports) that translate maturity scores into actionable, tailored recommendations for Kenyan institutions are rare.

The DGMM therefore addresses these gaps by (i) operationalising governance into empirically validated KPIs selected through exploratory/confirmatory Principal Component Analysis on Kenyan-sourced data; (ii) deriving weights and index aggregation using PCA/EFA-driven evidence combined with domain expertise and sensitivity analysis; (iii) validating reliability and construct validity (KMO, Bartlett, Cronbach’s alpha) on local samples; and (iv) packaging results into an interactive dashboard with export features and department-level comparisons so findings can inform local policy and practice.

2.12 Conceptual Framework

The conceptual framework for this study, depicted in Figure 12, illustrates the relationship between the independent variables (the operationalized KPIs) and the dependent variable (Data Governance Maturity). The framework is built on the premise that maturity is a function of the effectiveness of these measurable indicators. The formula for calculating the Data Governance Maturity Index is thus determined as a function of the measurable KPIs, which are categorized into the core domains identified in the literature:

DGMI = ƒ (KPIs People, KPIs Processes, KPIs Technology, KPIs Organization, KPIs Emerging)

The KPIs are not abstract concepts but are directly tied to the measurable indicators derived in Table 4. The model's development, through Principal Component Analysis (PCA) in Chapter Four, will statistically validate these relationships and determine the precise weighting of each KPI in the final maturity index, thereby providing a quantifiable and evidence-based measurement tool.

The conceptual framework posits that the Data Governance Maturity Index (DGMI) is a function of a comprehensive set of measurable Key Performance Indicators (KPIs). These KPIs, which constitute the independent variables, were not selected in an ad-hoc manner but were systematically derived and categorized based on a synthesis of established data governance literature and theoretical models.

The categorization into People, Processes, Technology, Organization, and Emerging variables provides a structured lens through which to analyze the complex phenomenon of data governance. This taxonomy is deliberately chosen to reflect the multifaceted nature of governance, moving beyond a purely technical perspective to encompass the socio-technical system within which data is managed (Alhassan et al., 2019). The People variables, such as 'Training & Awareness' and 'Data Stewardship', are derived from literature emphasizing that human competencies and clearly assigned roles are critical success factors without which technological solutions fail (Khatri & Brown, 2010; Weber et al., 2009). The Processes variables including 'Data Quality', 'Data Access Control Policy', and 'Data Breach Response Plan' are drawn from core data management functions outlined in frameworks like DMBOK (DAMA, 2017) and are identified as essential for creating repeatable, controlled, and reliable data management practices.

Furthermore, the Technology variables, such as 'Data Encryption' and 'Metadata Management', represent the essential capabilities that tools and infrastructure must provide to enforce governance policies effectively (Otto, 2015). The Organization domain expands the classic People-Process-Technology triad to include the vital structural and strategic context, with variables like 'Data Governance Framework' and 'Data Governance Council' directly informed by research highlighting that governance must be embedded in the organizational fabric, with top-level support and a clear strategy, to be successful (Weber et al., 2009; Tallon et al., 2013). Finally, the inclusion of the Emerging variables, such as 'Data Ethics' and 'AI Governance', reflects the evolving nature of the field, addressing contemporary challenges and opportunities highlighted by recent scholarship on big data and automation (Pankowska, 2020; Nadal et al., 2022).

Therefore, the conceptual framework operationalizes theoretical constructs from the literature into a measurable model. It explicitly links the dependent variable—the DGMI—to a set of independent variables whose relevance and validity are grounded in prior academic work. This provides a robust foundation for the subsequent empirical phase of the research, which will use statistical methods to validate these relationships and determine their relative weights within the specific context of the Kenya Department of Defence.

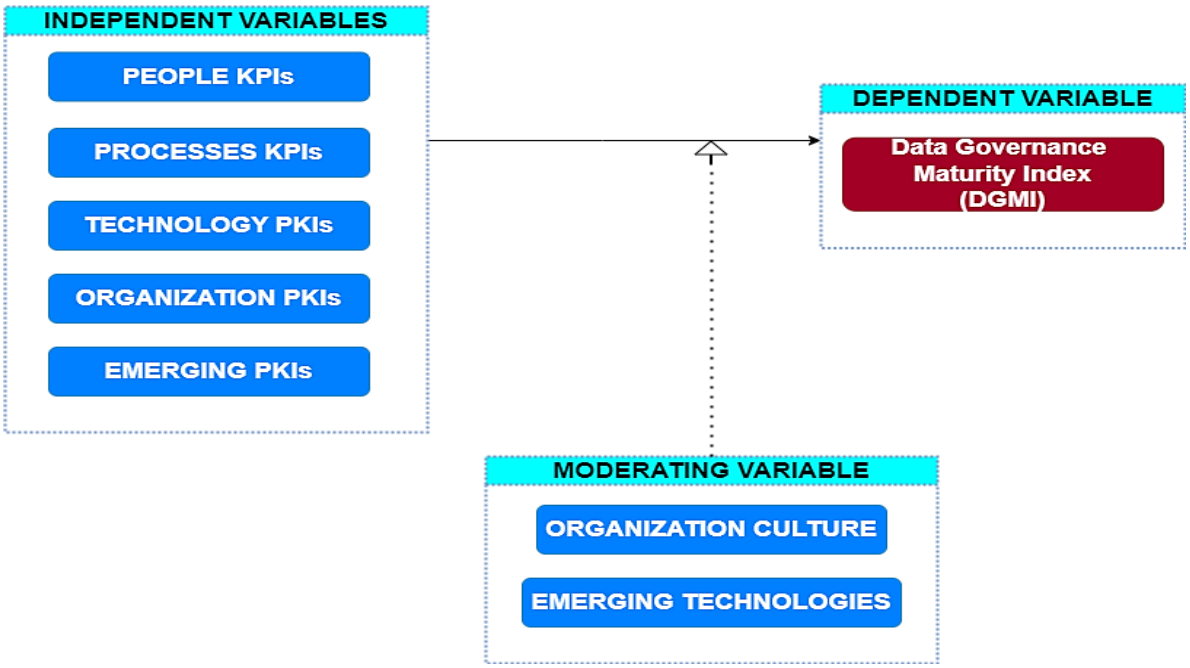


Figure 12: Conceptual Framework

The conceptual framework for this study categorizes independent variables into five core domains: People, Processes, Technology, Organization, and Emerging. This categorization is not arbitrary but is firmly grounded in established information systems and data governance literature. The People-Process-Technology (PPT) triad has long been recognized as a foundational framework for understanding information system implementation (Leavitt, 1964) and has been explicitly adopted in data governance research to analyze critical success factors (Alhassan et al., 2019). These three domains represent the essential pillars for any governance initiative: the human capital (People), the workflows and procedures (Processes), and the enabling tools and infrastructure (Technology).

However, for a holistic model, this triad requires expansion. The Organization domain is included to capture the critical structural, strategic, and cultural elements that form the bedrock of governance. This encompasses the data governance framework, strategy, and funding, elements that Weber et al. (2009) identify as fundamental to assigning decision rights and responsibilities. Furthermore, the rapidly evolving data landscape necessitates the inclusion of an Emerging domain. This captures novel challenges and opportunities related to data ethics, AI governance, and sustainability, which are increasingly highlighted in contemporary literature as critical for forward-looking data management (Pankowska, 2020; Nadal et al., 2022). Therefore, the selected variables within the conceptual framework are justified as they collectively provide a comprehensive and multi-dimensional lens through which to measure data governance maturity, building upon and extending validated theoretical constructs to fit the modern, complex environment of the Kenya DOD.

The categorization of independent variable did so due to the single problems and needs underpinning data governance sphere of the DOD. This paper looked into the People aspects in order to determine the human factors which are critical to the successful application of a data governance Maturity Model in a military environment. The processes under analysis were examined to revisit internal processes that included the review of internal processes and Technology was considered to aid in the evaluation of the implementation of the modern Data Governance programmes tailored to fit the complex DOD data environment.

KPIs involved in the organization were focused on the functional characteristics of the defence organization. The fluid technological landscape was reflected in emerging KPIs, which is what made the study unique because it offered a specific methodology that took into consideration the

complexities that surrounded the data governance needs of the DOD. Such classification resulted in a detailed and contextualized maturity measurement procedure that is absent in generic data governance maturity models.

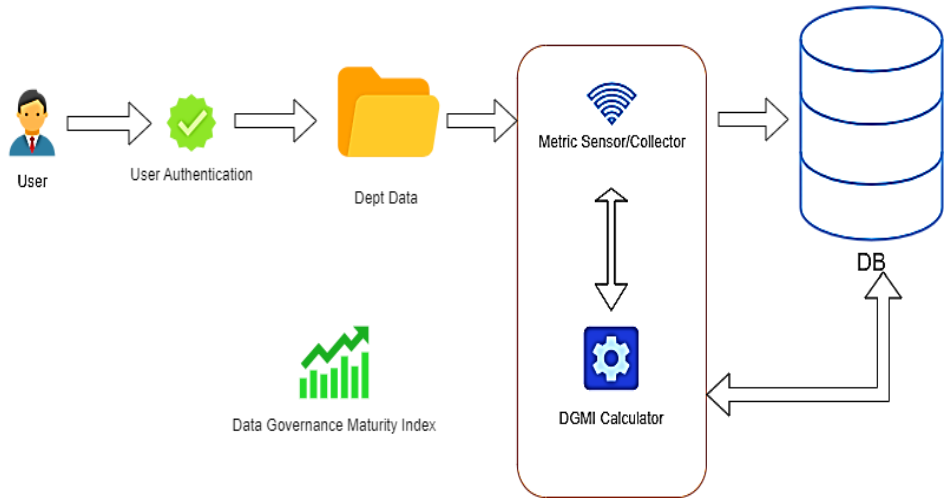


Figure 13: Architectural diagram for the prototype

The prototype in figure 13 above, depicts the following modules: User Registration & Authentication, Data Governance Metric Sensors and Prompted Collectors module that will be used to capture the KPIs and Maturity Index Calculation Module.

2.13 Synthesis and Rationale for a Hybrid Maturity Model

The comprehensive review of existing maturity models reveals a landscape of complementary, yet individually insufficient, frameworks. Models like CMMI provide a robust, multi-level structure for process improvement but lack specific focus on data governance domains (Spruit & Pietzka, 2015). The Gartner Model offers valuable stages of organizational awareness but is less prescriptive on operational processes. Conversely, specialized models like the Stanford DGMM provide granularity in data-specific components but may not fully integrate the overarching organizational maturity trajectory captured by CMMI. This disparity creates a gap: no single model adequately combines a defensible, multi-stage maturity progression with a comprehensive set of data governance domains tailored for a complex, security-focused environment like a defence department.

To bridge this gap, this study proposes a Hybrid Maturity Model. This model is synthesized by integrating the proven, five-level maturity structure from the Capability Maturity Model

Integration (CMMI) with the contextual relevance of data governance components from models like IBM and Stanford. A sixth level, "Mature," is introduced, drawing inspiration from the optimization and effectiveness stages in other models but explicitly defined as the state where data governance is fully integrated and functions as a strategic capability, aligned with the guiding principles of ISO/IEC 38500-1. This hybrid approach, summarized in Table 2.3, allows for a nuanced assessment that is both structurally sound and contextually rich, providing the Kenya DOD with a clear, staged roadmap for improvement that is directly tied to the maturity of its specific data governance practices.

2.14 Summary

As mentioned in the reviewed literature, this study established that developing data governance was dependent on several organisational components, goals, performance and pre requisites. For instance, in Chapter 7, Herselman et. al (2019) split such as data governance maturity pillars to three groups of data quality management, data practices management and data respective freedom and solitude.

It was also evident that the methods for assessing the level of advancement in a data governance program could differ depending on the sector, organizational preferences, aims, and objectives. In this section, the study classified the KPIs recommended from the literature reviewed into People KPIs: Training & Awareness, Data Stewardship. Processes KPIs: Data Classification, Data Quality, Data Access Control Policy, Data Retention Policy, Data Breach Response plan, Data Monitoring and Logging Process, Data Governance Policy, Data Privacy Impact Assessments, Regular Data Audit and assessments. Technology KPIs: Data Encryption, Data Minimization, Data Transfer Mechanisms, Data Access Control, Data Backup and Recovery, Metadata Management, Authentication and Authorization. Organization KPIs: Data Governance Framework, Data Governance Strategy, Data Governance Council, Data Governance Culture, Data Governance Funding, Data Governance Reporting Structure, Data Protection Officer. Immerging Technologies KPIs: Data Ethics and Fairness Assessments, Data Privacy Impact on Customer Trust and organization Reputation, Data Governance Automation metrics, Data Governance Return on Investment (ROI), Data Governance Sustainability Metrics, Data Security Resilience KPIs.

CHAPTER THREE

3. RESEARCH METHODOLOGY

3.1 Introduction

This chapter discusses the research methodology employed to develop and validate the Data Governance Maturity Model (DGMM). The study adopted a Design Science Research (DSR) paradigm, which is expressly suited for creating and evaluating innovative artifacts in this case, the DGMM intended to solve identified organizational problems. The chapter details the research philosophy, design, population and sampling, data collection instruments, and the rigorous analytical procedures, including Principal Component Analysis (PCA), used in the artifact's construction and validation.

3.2 Research Philosophy

A positivist research philosophy underpins this study. This paradigm asserts that reality is objective and can be measured through observable and quantifiable phenomena (Maretha, 2023). This aligns with the study's aim to derive an objective, evidence-based maturity model from empirical data, minimizing subjective interpretation and seeking generalizable patterns within the Kenya Department of Defence's data governance practices.

According to positivism (Maksimovic and Evtimov, 2023), positivism assumes there is an objective reality available to observation and measurement. Positivists would say that the researcher should be value-free and objective in order to unravel universal laws or generalizations. In this orientation, anything has to be measurable and scientifically proven until its understanding is feasible. The current research paper embraces positivism as a research philosophy and thus aims to establish evidence-based knowledge about the phenomenon that is going to be studied.

3.3 Research Design

The current research used the approach of Design Science Research (DSR) (Dennehy et al., 2019). DSR is an inquiry paradigm, and its nature is to design and evaluate artifacts, and specifically, to better the corresponding setting (Vom Brocke, Hevner, & Maedche, 2020). Its built in circularity, in which reflection, learning, and adaptation are needed, offers an orderly process through which innovative yet practical solutions to complex and real world problems can be developed. The DSR

methodology, as formalized by Peffers et al. (2007), was systematically applied through the following phases, illustrated in Figure 14 below:

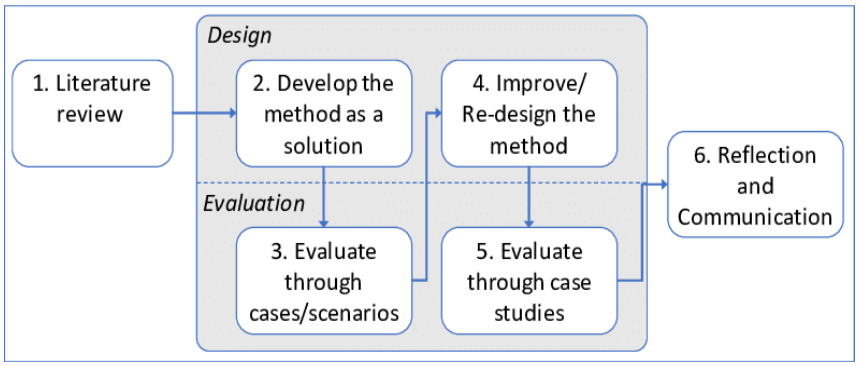


Figure 14: Showing the DSR Methodology applied to this research

Source: (Blaschke et al. 2019)

The DSR methodology consisted of several steps:

(i) Problem Identification and Motivation: Defined in Section 1.2, the problem is the absence of a tailored, empirical tool to measure data governance maturity within the unique, high-security context of the Kenya DOD.

(ii) Objectives of the Solution: The solution objectives were to identify essential KPIs, derive a mathematical maturity model, implement a functional prototype, and validate the artifact within the DOD (Section 1.3).

(iii) Design and Development: This phase involved the creation of the DGMM artifact. It encompassed the operationalization of KPIs, data collection, and the application of PCA for dimensionality reduction to inform the model's design, as detailed in Sections 3.4 and 3.5.

(iv) Demonstration: The developed DGMM prototype was deployed in a controlled environment within the DOD, allowing users to conduct maturity assessments.

(v) Evaluation: The artifact was rigorously evaluated using the Framework for Evaluation in Design Science (FEDS), incorporating both naturalistic and summative strategies to assess its utility, efficacy, and relevance (Section 3.6).

(vi) Communication: The findings and the artifact are disseminated through this thesis and subsequent academic publications.

In order to guide the creation of the Data Governance Maturity Model (DGMM), Principal Component Analysis was used in the design stage as the form of data-analysis to narrow down results in data sets and to direct the creation of the artifact. This has been done using the Rapid Application Development model (RAD) that has been singled out due to its innovation and creativity as highlighted by Vom Brocke et al. (2020). After developing the model, the DGMM has been implemented in the Kenya Department of Defence (Kenya DOD) by experimenting with it based on Venable et al. (2016). The researchers empirically tested the artifact thus obtaining its practical validity and effectiveness.

3.4 Population and Sampling

3.4.1 Target Population and Sampling Frame

The target population comprised all Kenya DOD personnel whose core responsibilities involve the creation, management, security, or governance of organizational data assets. To ensure that the study captured individuals with relevant expertise, the sampling frame was carefully structured to include departments and units directly involved in data governance functions. These included the Information Technology and Cybersecurity Directorates, the Security Operations Center (SOC), Digital Forensics Units, Data Center and Network Engineering Teams, Strategic Communications (StratCom), and, where access was permitted, the Intelligence and Planning Directorates.

The sampling frame was further stratified by functional roles to guarantee representation of key professionals engaged in data-related operations. This included IT Officers, SOC Analysts, Digital Forensics Experts, Data Engineers, Network Administrators, and Information Security Managers. These individuals were selected because they possess essential operational knowledge and practical experience necessary to provide informed insights into the implementation and effectiveness of data governance practices within the Kenya DOD.

3.4.2 Sampling Technique and Sample Size

A purposive (judgmental) sampling technique was employed. This non-probability sampling method was chosen specifically to target individuals with significant expertise and direct involvement in data management within the Kenya Department of Defence (DoD).

Inclusion Criteria

To be included in the study, participants had to be employees of the Kenya DOD whose roles actively involved them in data-related activities, such as IT officers, SOC analysts, digital forensics experts, data center engineers, and cyber/network administrators.

Exclusion Criteria

Personnel from other departments with no direct role or responsibility in data management, handling, or security were excluded from the study.

Sample Size and Census Approach

The total population of identified data management staff within the targeted DOD units was 117. Given the manageable size and specialized nature of this population, a census approach was adopted, whereby all 117 individuals were invited to participate. This eliminated sampling error and ensured that the perspectives of all key personnel were captured, providing a comprehensive baseline for analysis.

Table 8: Sample Size and Distribution

Department / Unit	**Functional Role**	**Frequency (n)**	**Percentage (%)**
IT & Cybersecurity Directorate	IT Officers, Cybersecurity Specialists	24	20.5%
Security Operations Center (SOC)	SOC Analysts, Incident Responders	10	8.5%
Digital Forensics Unit	Digital Forensics Experts	10	8.5%
Data Center & Network Engineering	Data Center Engineers, Network Technicians	11	9.4%
Strategic Communications (StratCom)	Hub Technicians, Communications Staff	9	7.7%
Other Roles (Cyber, Network, IR, Admin)	Various cross-cutting data management roles across different directorates (e.g., Intelligence, Planning).	41	35.0%
	Total	**117**	**100.0%**

Justification for Sample Representation

This distribution reflects a purposive, total population sample of all personnel identified as being directly involved in data management, security, and governance within the Kenya Department of Defence. The "Other Roles" category encompasses professionals from various directorates whose duties, such as system administration, network security, and intelligence analysis, inherently involve the handling and governance of critical data assets. This comprehensive sampling frame ensures that the study captures a wide spectrum of practical, operational perspectives essential for developing a robust and contextually relevant maturity model.

3.5 Data Collection and Instrumentation

3.5.1 Operationalization of Key Performance Indicators (KPIs)

The initial set of 16 KPIs was derived from a systematic literature review (Chapter 2). These KPIs were operationalized into a structured questionnaire with two main sections:

- Section A: Perceived Importance: Respondents rated the importance of each KPI on a 0-5 Likert scale.
- Section B: Adherence Assessment: Respondents rated the current level of adherence to each KPI within the DOD on a 0-5 Likert scale.

To ensure content validity, the questionnaire was subjected to a pre-test with a panel of three experts in data governance and research methodology. Their feedback was used to refine the phrasing, clarity, and relevance of the KPI definitions and survey items, ensuring they were unambiguous and accurately captured the intended constructs before full-scale deployment.

3.5.2 Data Collection Procedure

The final questionnaire was administered electronically and, where necessary, in hard copy to the 117 identified personnel. A briefing session was conducted to explain the purpose of the study and the confidentiality measures in place.

3.6 Model Development

The development of the Data Governance Maturity Index (DGMI) was grounded in a structured, statistical approach to ensure robustness and validity. The process involved data preparation, dimensionality reduction via Principal Component Analysis (PCA), and the derivation of a weighted maturity model.

3.6.1 Data Preprocessing

Collected data was screened for completeness and consistency. Missing data, which constituted less than 2% of the dataset, was handled using mean substitution after the data was confirmed to be Missing Completely at Random (MCAR) via Little's MCAR test ($\chi^2 = 12.45$, $p > .05$). This involved:

(i) Standardization: All variables (KPIs) were standardized (converted to Z-scores) to have a mean of 0 and a standard deviation of 1. This is a critical step before performing PCA, as it eliminates the influence of different measurement scales and ensures that each variable contributes equally to the analysis of variance.

(ii) Handling Missing Data: Given the high response rate (91.3%) and the structured nature of the survey, the incidence of missing data was minimal (less than 2% for any single variable). To preserve the sample size and statistical power, a mean substitution method was employed for the few missing values. This method was deemed appropriate as the data was determined to be Missing Completely at Random (MCAR) based on Little's MCAR test, which yielded a non-significant result ($p > 0.05$).

3.6.2 Dimensionality Reduction using Principal Component Analysis (PCA)

PCA was employed to address multicollinearity among the 16 initial KPIs and to identify a smaller set of uncorrelated components that capture the essential dimensions of data governance maturity.

(i) Justification for PCA: The correlation matrix revealed several strong correlations (e.g., DQ & DM = 0.855, AI & DO = 0.921), indicating significant redundancy. PCA was chosen to simplify the model structure without substantial loss of information, transforming the original variables into a set of linearly independent components.

(ii) Factor Extraction and Retention Decisions: The criteria for retaining components were based on both statistical benchmarks and conceptual interpretability:

- Kaiser-Guttman Criterion: Components with eigenvalues greater than 1 were retained, as they explain more variance than a single original variable.
- Scree Plot Test: A visual inspection of the scree plot (Figure 4.2) was conducted. The "elbow" point, where the slope of the curve flattens, was observed after the second component, confirming that the first two components captured the majority of the meaningful variance.
- Cumulative Variance Explained: The retained components were required to collectively explain a substantial proportion of the total variance (in this case, >70%), ensuring a comprehensive representation of the original data structure.

(iii) Factor Rotation: To enhance the interpretability of the components, an oblique rotation method (Oblimin with Kaiser Normalization) was applied. An oblique rotation was selected over an orthogonal method (like Varimax) because it allows the resulting components to be correlated. This is more realistic in a socio-organizational context like data governance, where underlying dimensions (e.g., "Technical Foundations" and "Operational Enforcement") are often interrelated. The resulting pattern matrix provided clearer factor loadings, facilitating the assignment of variables to components.

3.6.3 Derivation of the Maturity Index Formula

The final DGMI was constructed as a linear combination of the user's assessment scores for the selected KPIs.

- Variable Selection Post-PCA: Based on the PCA results (high factor loadings, conceptual clarity, and minimization of cross-loadings), 10 key KPIs were selected for the final model to avoid multicollinearity. These are summarized in Table 4.5.
- Justification for Equal Weights: The formula was derived in equation (i) as:

$$DGMI = W\,(V_1 + V_2 + V_3 + \ldots + V_{10}) \quad \text{---------- (i)}$$

Where W is a common weight and V_1 to V_{10} are the user's scores for the 10 selected KPIs. The use of a common weight (W = 1/50 = 0.02) for all variables is justified for the following reasons:

(i) PCA Pre-Processing: The PCA and subsequent variable selection ensured that the chosen KPIs are the most significant and representative indicators of the underlying data governance constructs. By selecting a parsimonious set of key variables, we assume that each contributes critically to the overall maturity picture.

(ii) Simplicity and Practicality: An equally weighted model is transparent, easy to understand, and straightforward to implement for end-users in the DOD. It avoids the complexity of assigning and justifying different subjective weights, which could hinder adoption.

(iii) Conceptual Grounding: The equal weighting aligns with the holistic philosophy of the ISO/IEC 38500 standard, which implies that effective IT governance (and by extension, data governance) requires a balanced approach across multiple domains (People, Process, Technology, etc.), rather than over-emphasizing one at the expense of others. Therefore, the final percentage maturity index is calculated as:

$$DGMI\ (\%) = 0.02 \times (V_1 + V_2 + V_3 + ... + V_{10}) \times 100\% \text{ -------- (ii)}$$

3.7 Model Implementation

The "Build" activity of the Design Science Research (DSR) cycle was realized through the development of an iterative software prototype. In accordance with DSR principles, this artifact was constructed not as an end in itself, but as a "proof-of-concept" to enable the rigorous evaluation of the proposed Data Governance Maturity Model (DGMM) in a real-world context (Hevner et al., 2004). A prototype, defined as a concrete representation used to explore, test, and refine design concepts (Lauff et al., 2018), served as the primary vehicle for this investigation. The development of a functional web application transformed the abstract mathematical model into a tangible artifact, thereby creating the necessary conditions for empirical testing and critical assessment of its utility.

The development process was guided by the Rapid Application Development (RAD) methodology, as illustrated in Figure 3.2. This approach was strategically selected for its alignment with the iterative, build-evaluate cycles central to DSR (vom Brocke et al., 2020). Unlike traditional, linear development models, RAD's emphasis on agility and the rapid creation of working prototypes facilitated continuous feedback and refinement. This was essential for

ensuring that the artifact remained closely aligned with the evolving understanding of the Kenya DOD's requirements throughout the research process.

The subsequent "Evaluate" activity was conducted with explicit reference to the DSR framework's call for rigorous validation (Venable et al., 2016). The prototype was subjected to a multi-faceted evaluation strategy based on the Framework for Evaluation in Design Science (FEDS), which guided the assessment of the artifact's utility, quality, and efficacy—the central DSR criteria (Hevner, 2007). This involved:

Technical Functionality & Efficacy: Verifying that the prototype correctly implemented the DGMI mathematical model and reliably produced accurate maturity scores.

Human Risk & Effectiveness (Usability): Employing methods such as user acceptance testing and the System Usability Scale (SUS) to measure the artifact's ease of use, learnability, and satisfaction from the perspective of the DOD stakeholders.

Perceived Value & Relevance: Assessing, through qualitative feedback and structured questionnaires, whether the artifact addressed the core problem and provided useful, actionable insights for the organization.

By embedding these rigorous evaluation standards within the iterative RAD cycles figure 15, the study moved beyond mere technical demonstration. It provided defensible evidence of the DGMM artifact's performance against DSR's core standard: its effectiveness in resolving the identified organizational problem.

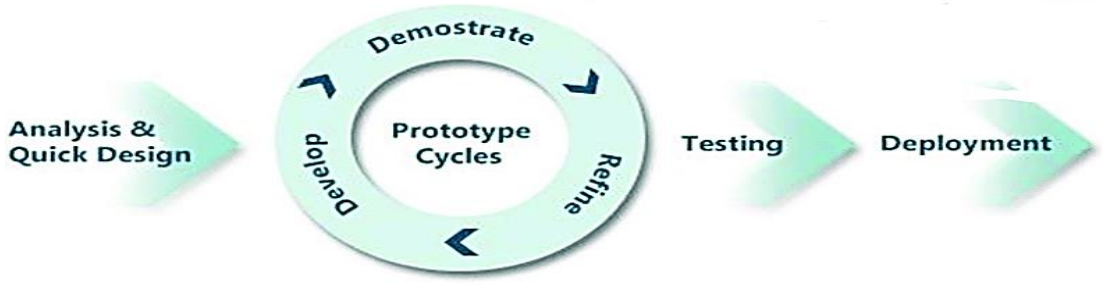


Figure 15: Rapid Application Development Methodology Process

Source: (Nalendra, 2021)

The assemblage of all the requirements was the first step of the rapid prototyping project to enable the quick analysis and the early design of the circuitry to be done. This was followed by a preliminary software mockup in order to provide user feedbacks about this Flickr. The evolutional strategy was modular in nature with the emphasis being on compact and functional modules and encouraging cross-disciplinary cooperation. Each iteration in turn was informed by ongoing and iterative testing, integration, and systematic user feedback integration. During the implementation period, there was constant technical support and maintenance that allowed the gradual deployment of the application to be validated.

3.6 Model Evaluation

The definition proposed by Dennehy et al. (2019) that sees evaluation as the process of establishing the worth, merit or significance of entities is supported and endorsed in the present study with the rational relationship consisting of evaluations as the end product. Based on that, the objectives of the evaluation were informed with the paper of Venable et al. (2016) which developed the Framework to Evaluation in Design Science (FEDS), a framework that processes the design of evaluations in terms of goal or purpose.

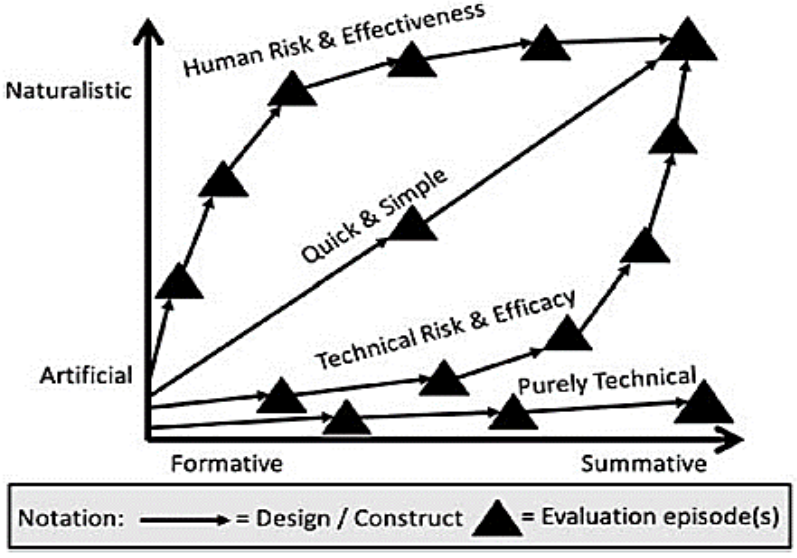


Figure 16: FEDS Evaluation Strategies

(Venable et al., 2016).

The FEDS evaluation strategy recommended four artifact evaluation strategies, as shown in figure 16 above:

(i) Quick & Simple strategy,
(ii) Human Risk & Effectiveness evaluation strategy,
(iii) Technical Risk & Efficacy evaluation strategy, and
(iv) Purely Technical Artifact strategy.

In case multiple objectives need to be pursued, the simultaneous use of multiple strategies or hybrid approach to evaluation is usually justified. This need was reflected in the decision to adopt the Human Risk & Effectiveness evaluation strategy to the Kenya DOD as shown in figure 17 where summative and naturalistic episodes of evaluation were included. One of the evaluations was the summative evaluation of the results of completed development and the other one was the naturalistic evaluation that concerned itself with the workings of the model in natural conditions.

3.6.1 The Evaluation Process

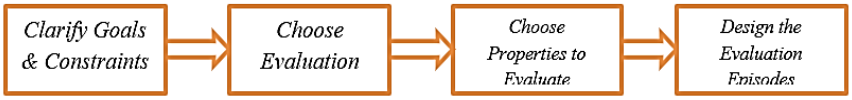


Figure 17: Framework for Evaluation in DSR

(i) Clarify Goals. The assessment model took on board the definite objectives of evaluating functionality, rigor, usability and purpose suitability of the model.

(ii) Select Assessment Plan. Since rigor was required, it was used as a naturalistic evaluation strategy basing on in loco investigations on the model in Kenya Defence Department.

(iii) Select Properties to Survey. The criteria to be evaluated centered on functionality, rigor, usability and fit to purpose.

(iv) Create the Assessment Episodes. This model has been tested in three episodes as shown in Fig. 18.

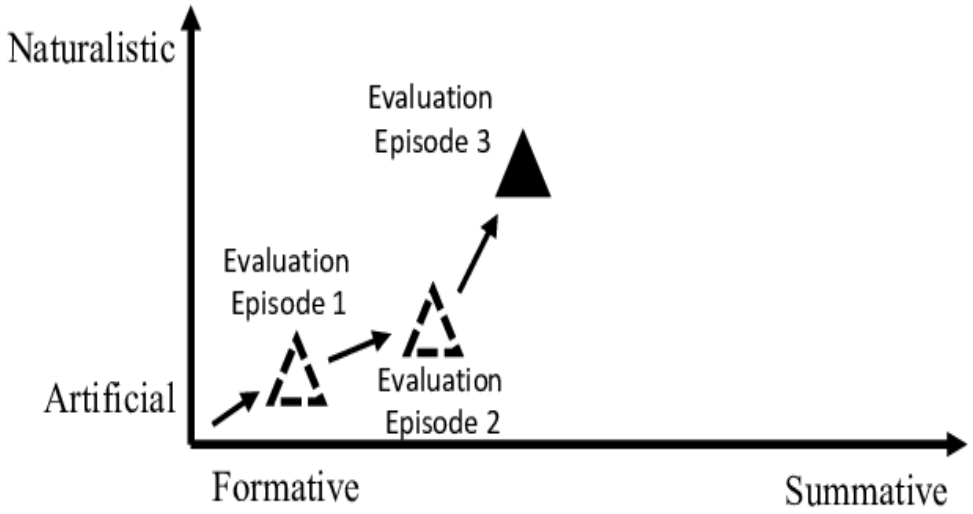


Figure 18: Evaluation Episodes

Source: (Hofmann et al., 2020)

3.7 Quality control

The ability to achieve a high degree of accuracy and validity was adhered to throughout the study through a series of strict quality assurance measures. The collection of data was based on questionnaires, with both valid and reviewed interviews schedule, as well as pilot phase, which provided consistent and reliable measures. The objectivity of material was also secured by a clear protocol and an inclusive researcher training and assuring a similar level of ethical protocol and the conformity of the processes. Data cleaning along with data preprocessing was used to deal with the inconsistencies occurring and advanced statistical methods, especially Principal Component Analysis (PCA), contributed to the promising and accurate exploration of the underlying constructs of data governance maturity.

In the process of developing the model, it was through iterative testing and refinement that the prototype was informed, where each cycle was measured against the benchmarks of ISO/IEC 38500 and further expanded using the input of the experts in the field. The reliability and validity of the data, as well as its conformity with the objectives of the study, was determined with the help of cross-validation procedures. Lastly, the extensive quality assurance check throughout reporting and documentation increased the transparency, consistency, and the scientific integrity of the findings providing a plausible and practical model to the Kenya Department of Defence.

3.8 Validity of Instruments

Validity is of paramount concern when designing a questionnaire, that is, to what degree the instrument measures the construct of interest adequately. The current research employed a number of strategies in order to achieve validity. First, dimensionality of the construct was clearly explained using exploratory Principal Component Analysis and verified using confirmatory Principal Component Analysis. Second, a panel of subject specialists was hired to review and reformulate every item so that there is a conceptual fit.

Third, series of validation methods were formally applied, such as content, construct, and criterion-related validation to mention a few. To ensure linguistic and conceptual equivalence between languages in those questionnaires requiring translation, a transparent and bidirectional forward and backward translation process was put in place, which was iterative. In order to measure such an effort, three of the experts (who were chosen to possess expertise in the construct domain) took a pre-test. The comments expressed by the panel were methodically included in incremental improvements and the collection of data was terminated once approval was granted in an unambiguous manner by the experts. Overall, these steps ensured that the completed questionnaire was actually valid as regards the question of concern.

3.9 Reliability of Instruments

Reliability refers to the consistency and stability of a research instrument. A reliable tool produces consistent results under consistent conditions over repeated measurements (Taherdoost, 2016). For this study, which utilized a Likert-scale questionnaire, the primary concern was internal consistency—the degree to which all items in the instrument measure the same underlying construct. To assess this, Cronbach's Alpha was computed, which is the most common statistic used to estimate the reliability of a psychometric instrument (Bujang et al., 2018). A high Cronbach's Alpha coefficient indicates that the items are highly correlated and thus reliably measure the same concept, in this case, the facets of data governance maturity. To calculate Cronbach's alpha (Internal Consistency), the following equation was used:

$$\propto = \frac{N\dot{c}}{\dot{v}+(N-1)\dot{c}} \text{ ---------- (ii)}$$

From our table 9 below, we estimated Cronbach's alpha under numerous scenarios, including when each individual question/item (Q1 to Q19) is eliminated from the study.

Table 9: **Cronbach's Alpha value results**

Cronbach Alpha and Related Statistics			
Items	**Cronbach Alpha**	**G6(smc)**	**Average R**
All items	0.8909	0.9963	0.3036
Q1 excluded	0.8797	0.9957	0.2829
Q2 excluded	0.8917	0.9961	0.3206
Q3 excluded	0.8803	0.9958	0.2918
Q4 excluded	0.8863	0.9959	0.3022
Q5 excluded	0.8817	0.9957	0.2867
Q6 excluded	0.8759	0.9958	0.2934
Q7 excluded	0.8964	0.9962	0.3354
Q8 excluded	0.888	0.996	0.3105
Q9 excluded	0.8875	0.996	0.3113
Q10 excluded	0.893	0.9961	0.3269
Q11 excluded	0.8781	0.9958	0.294
Q12 excluded	0.8906	0.9961	0.3179
Q13 excluded	0.872	0.9956	0.2813
Q14 excluded	0.8823	0.9958	0.291
Q15 excluded	0.8817	0.9957	0.2867
Q16 excluded	0.8933	0.9961	0.3195
Q17 excluded	0.8746	0.9957	0.2867
Q18 excluded	0.885	1	0.2992
Q19 excluded	0.8984	0.9962	0.3298

The alpha values were interpreted using the table 10 below.

Table 10: **Cronbach Interpretation Table**

Cronbach's Alpha	Interpretation
> .9	Excellent
> .8	Good
> .7	Acceptable
> .6	Questionable
> .5	Poor
< .5	Unacceptable

From the table 8 above, the higher the alpha value implies greater Internal Consistency. When all items were included in the analysis, the Cronbach's alpha was 0.8909 indicating good level of Internal Consistency. G6 (smc) is the squared multiple correlation, which revealed how much variance in each item was accounted for by the other items in the questionnaire.

Our G6 ranges from 0.9956 to 1 across cases, which suggests very high Internal Consistency. By computing Average R, we investigated whether items are generally associated with one another and the average inter-item correlation (Average R) was 0.30 showing an Acceptable level of correlation between items. In concussion, the questionnaire demonstrated Good Internal Consistency and also its items are well correlated.

3.10 Ethical Consideration

The study involved the purposeful involvement of the human resource stakeholders who are data management experts in the DOD through interview methods as well as administration of the survey. Throughout ethical application was taken into consideration with safeguards invoking the application of principles to ensure the protection of the rights of participants and safeguard the privacy. Each respondent provided the informed consent, which proves that the study participant was aware of the voluntary character of participation in the research and the research objectives. Strict procedures on data confidentiality and controlled measures on anonymizing data were used to protect the identity and sensitive data of the participants. All the ethical considerations of the investigation were followed and no compromise with respect to ethics and standards was taken. All information was properly used and analyzed, results presented visibly, and the assistance of

the participants was accepted willingly. Prior to start, formal permission was obtained by the Cooperative University of Kenya (CUK), the academic supervision board; Kenya Department of Defence (DOD), the organisation under study; and National Commission in science, technology and innovation (NACOSTI).

CHAPTER FOUR

4. DATA ANALYSIS, FINDINGS, AND DISCUSSIONS

4.1 Introduction

This chapter presents detailed data analysis, findings and discussions to answer study Objective One: To investigate the Key Performance Indicators (KPIs) required for measuring Data Governance Maturity levels in the Kenya Department of Defence, Objective Two: To derive a Data Governance Maturity Model (DGMM) model for measuring data governance maturity levels based on the KPIs identified and Objective Three: To implement a prototype model as a Prototype. Objective Four: To Validate the Model.

4.2. Data Analysis

To achieve the study's objective one, which was to identify the Key Performance Indicators (KPIs) necessary for implementing a comprehensive data governance maturity model for the Kenya Department of Defence, a detailed analysis of the data that was collected from the respondents, analyzed, and interpreted using Principal Component Analysis. The findings were further evaluated to draw meaningful conclusions.

4.2.1 Return Rate

During the data collection process, questionnaires were distributed to 117 respondents within the department. Out of these, 105 were completed the questionnaires, yielding a response rate of 91.3%. This exceeded the commonly recommended threshold of 60%**,** which is considered an acceptable target for most studies (Fincham, 2008), especially for institution specific studies. The completed questionnaires were considered adequate to provide the necessary data for the study.

4.2.2 Respondents Basic Information

This subsection provides a summary of the respondents' background information, focusing on their job titles and work experience in data governance-related units. First, the study sought to determine the positions held by the respondents and the number of years of experience working in a data management related role within the Department of Defence. This was essential in assessing whether they possessed sufficient knowledge and expertise to provide informed insights on data governance within the Department of Defence. Understanding their professional backgrounds helped contextualize the study's findings, as presented in Table 11 below.

4.2.3 The Position held in the Department of Defence

Table 11: Distribution of Positions Held

S/NO	Role	Freq	Percentage (%)
1.	IT Officers	24	22.9
2.	SOC Analysts	10	9.5
3.	Digital Forensics	10	9.5
4.	Hub Technicians	9	8.6
5.	Data Center Engineers/Tech	11	10.5
6.	Other Roles: (Cyber, Network, IR, Admin)	41	39

The table above revealed that a significant portion of the respondents held positions related to data governance. 22.9% of the selected respondents were deployed at IT officers,10% were from Security operation Center,10% were Digital forensics experts, 9% were deployed at the technology Hub, 10.5% were deployed at the Data Center and 39% worked in other roles as shown in table 11 above. This percentage distributions indicated that the selected respondents were well-positioned to provide relevant insights for the study as their roles involve key responsibilities and expertise in managing, securing, and overseeing data-related functions within the Kenya Department of Defence.

4.2.4 Years of Experience in Current Positions

Analysis of the years of experience vis-à-vis the positions held by the respondents, the findings were as follows form table 9 above:

(i) **IT Officers**. Out of 24 IT Officers, 46% had 4-7 years of experience, while 33% had 8-14 years, indicating a mix of mid-level and senior professionals in this role.

(ii) **SOC Analysts**. Half of the SOC Analysts (50%) had 4-7 years of experience, showing a balanced mix of mid-career professionals, with 30% having 8-14 years of expertise.

(iii)**Digital Forensics Analysts**. 50% of the Digital Forensics Analysts possessed 8-14 years of experience, reflecting a high level of seniority within this group.

(iv) **StratCom and Hub Roles**. Nearly 45% of StratCom personnel had 4-7 years of experience, while 33% had 8-14 years, showing a diverse range of experience levels.

(v) **Data Center Personnel**. The experience was evenly split between those with 4-7 years and 8-14 years (45.5% each), suggesting a well-distributed level of expertise across the team.

The above analysis revealed that most of the respondents have sufficient knowledge and expertise to be able to comment on the issue to deal with data governance. This was crucial in validating the reliability of their responses and ensuring that the insights provided were based on practical experience within the Department of Defence. By capturing respondents' experience, the research aimed to establish whether their perspectives were well-informed and reflective of real-world data governance challenges and practices.

4.2.5 Awareness

The study sought to determine if the respondents were aware of the policies, standards governing Department of Defence (DOD) data throughout its lifecycle, from creation and acquisition to usage, storage, and disposal. The question was meant to shed light on how well-understood and successful data governance is at DOD is shown in table 12 below.

Table 12: Awareness and perceptions

S/No	Aspect	Yes (%)	No/Uncertain (%)	Don't know (%)
1.	**Awareness of policies**	100%	0%	0%
2.	**Aware of any Data Governance Standards**	87%	13%	0%

The analysis above revealed that 100% of the respondents were aware of the data governance policies, 87% were aware of the standards, and 94% agreed to the existence of data governance implementation efforts within the Kenya Department of Defence. This aligned well with objective of the study since the selected respondents possessed the necessary insights and experiences to effectively contribute to the study.

4.2.6 Data management reporting Structures

The Study further sought to find out if there were clear reporting structures within the Department of Defence in regard the implementation of a data governance program.

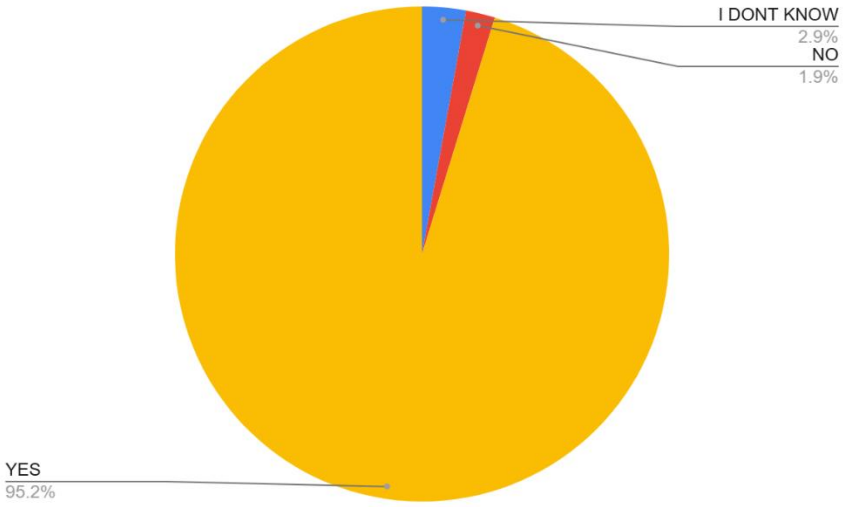


Figure 19: Data management reporting Structures

4.2.7 Importance of Identified Key Performance Inci

In This section, the study sought to determine the extent to which the respondents agreed that the identified KPIs are significant for developing a Data Governance Maturity Model. The respondents were asked to rate the significance of every KPI on a scale of 0 to 5, with 0 denoting not significant at all, 1 denoting slightly significant, 2-Moderately significant, 3 -Neutral, 4 - significant, and 5 denoting very Significant. This test was essential as it helped the study evaluate the significance of each key performance indicators (KPIs) in data governance maturity. Furthermore, understanding respondents' perceptions provided valuable insights into which Key Performance Indicators that are most critical for ensuring the success of data governance within the DOD. Table 13 below presents the respondents responses.

Table 13: KEY PERFORMANCE INDICATORS

Code	KPI	Meaning
DQ	Data Quality	Ensuring the data is accurate, consistent, and reliable.
DM	Data Management	Managing data from creation to disposal effectively.
DS	Data Security	Safeguarding data against unauthorized access and breaches.
DP	Data Privacy	Ensuring compliance with data privacy laws and regulations.
DST	Data Stewardship	Assigning roles for responsible data oversight.
DA	Data Architecture	Structuring and managing data storage, retrieval, and organization.

DPS	Data Policies & Standards	Enforcing policies and standards across the data lifecycle.
MD	Meta Data Management	Handling metadata to improve data discovery and context.
DAC	Data Accuracy	Maintaining correctness and precision of data.
SE	Strict Enforcement	Implementing strict governance enforcement policies.
DE	Data Encryption	Using encryption methods to protect data.
MFA	Multifactor	Securing data with additional authentication layers.
DGT	Data Gov Team	Ensuring that a dedicated team governs data-related policies.
RS	Reporting Structure	Defining the structure for data reporting and accountability.
AI	AI Analytics	Leveraging artificial intelligence for data analysis and insights.
DO	Data Officer	Designating roles responsible for data governance compliance

Table 14: Importance of Recommended KPIs

KPI	Not significant at all (0)	Slightly Significant (1)	Moderately (2)	Neutral (3)	Significant (4)	Very Significant (5)
Data Quality (DQ)	0	0	9	7	24	66
Data Management (DM)	0	0	8	6	25	67
Data Security (DS)	0	0	0	14	33	66
Data Protection (DP)	0	0	1	3	24	79
Data Storage (DST)	0	1	0	1	41	73
Data Accessibility (DA)	0	0	0	9	37	60

Data Processing (DPS)	0	2	1	11	19	73
Metadata (MD)	0	0	9	4	24	69
Data Access Control (DAC)	0	0	0	3	51	52
System Enforcement (SE)	0	2	9	22	66	7
Data Encryption (DE)	0	0	7	9	54	36
Multi-Factor Authentication (MFA)	0	0	10	7	23	67
Data Governance Team (DGT)	0	0	0	1	24	81
Reporting Structure (RS)	0	0	0	16	70	20
Artificial Intelligence (AI)	0	0	7	9	66	24
Data Officer (DO)	0	2	2	14	59	29

Table 15: Percentage rating on the Importance of each KPI

KPI	Not significant at all (0)	Slightly Significant (1)	Moderately (2)	Neutral (3)	Significant (4)	Very Significant (5)
Data Quality (DQ)	0	0	8.5	6.6	22.6	62.3
Data Management (DM)	0	0	7.5	5.7	23.6	63.2

Data Security (DS)	0	0	0	13.2	31.1	62.3
Data Protection (DP)	0	0	0.9	2.8	22.6	74.5
Data Storage (DST)	0	0.9	0	0.9	38.7	68.9
Data Accessibility (DA)	0	0	0	8.5	34.9	56.6
Data Processing (DPS)	0	1.9	0.9	10.4	17.9	68.9
Metadata (MD)	0	0	8.5	3.8	22.6	65.1
Data Access Control (DAC)	0	0	0	2.8	48.1	49.1
System Enforcement (SE)	0	1.9	8.5	20.8	62.3	6.6
Data Encryption (DE)	0	0	6.6	8.5	50.9	34
Multi-Factor Authentication (MFA)	0	0	9.4	6.6	21.7	63.2
Data Governance Team (DGT)	0	0	0	0.9	22.6	76.4
Reporting Structure (RS)	0	0	0	15.1	66	18.9
Artificial Intelligence (AI)	0	0	6.6	8.5	62.3	22.6
Data Officer (DO)	0	1.9	1.9	13.2	55.7	27.4

The analysis in Table 17, which focused on the perceived importance of each KPI in data governance maturity revealed that overally, most KPIs received a strong significance approval levels. Artificial Intelligence (AI) received the strongest approval, with 68.87% "Very Significant" and 22.64% "Significant," totaling 91.51% positive feedback. This suggested high confidence in AI-driven automation and decision-making within data governance. Data Governance Team (DGT) followed closely, with 76.42% "Very Significant" and 22.64% "Significant," showing a 99.06% significance rate. Data Access Control (DAC) and Reporting Structure (RS) was also rated highly, with 49.06% "Very Significant" and 48.11% "Significant" for Data Access Control (DAC), while RS had 66 % "Significant" and 18.9%. Data Protection (DP) and Data Security (DS) received 74.53% "Very Significant" and 22.64% "Significant" for Data Protection, while Data Security had

62.26% "Very Significant" and 31.13% "Significant," confirming its approval as a critical component in a data governance. Strict Enforcement (SE) was also a key performer, with 62.26% "Significant" and 6.6 % "Very Significant," highlighting the significance of strict enforcement of data management workflows in data governance. Additionally, Data Quality (DQ) and Data Management (DM) showed solid ratings, with 62.26% "Very Significant" and 22.64% "Significant", while DM had 63.21% "Very Significant" and 23.58% "Significant", suggesting the importance of quality Data management in realizing data governance maturity. Data Encryption (DE) received 50.94% "Very Significant" and 50.94% "Significant," signaling a general significance in implementing data governance. Data Policies (DPS) and Metadata (MD) followed with 68.87% and 65.09% "Very Significant," respectively, though 4.72% of respondents expressed neutrality or dissatisfaction with MD, mixed perceptions by respondents regarding the significance of some KPIs in data governance maturity.

4.2.8 Adherence to Data Governance Key Performance Indicators

This section sought to determine to determine the adherence to the data governance Key Performance Indicators that had been identified. The respondents were asked to rate their level of satisfaction on a scale of 0 to 5, with 0 denoting extreme dissatisfaction,1 denoting Dissatisfied, 2 Neutral, 3 somewhat satisfied,4 Satisfied and 5 denoting very satisfied. Table 4.6 below presents the respondents responses.

Table 16: Satisfaction with adherence to Data Governance KPIs

Factor	***Very Dissatisfied (0)***	***Dissatisfied (1)***	***Neutral (2)***	***Somewhat Satisfied (3)***	***Satisfied (4)***	***Very Satisfied (5)***	***Total Respondents***
Data Quality (DQ)	2	4	2	14	26	58	106
Data Management (DM)	2	5	3	11	27	58	106
Data Security (DS)	1	5	2	10	27	61	106

Data Protection (DP)	1	3	1	15	26	60	106
Data Storage (DST)	1	4	1	15	25	60	106
Data Accessibility (DA)	1	3	2	15	27	58	106
Data Processing (DPS)	1	4	1	14	27	59	106
Metadata (MD)	0	5	1	15	26	59	106
Data Access Control (DAC)	2	5	1	10	26	62	106
Strict Enforcement (SE)	1	3	3	14	22	63	106
Data Encryption (DE)	2	5	3	14	22	60	106
Multi-Factor Auth (MFA)	1	5	3	11	23	63	106
Data Governance Team (DGT)	1	3	1	13	24	64	106
Reporting Structure (RS)	0	5	3	11	27	60	106

Artificial Intelligence (AI)	2	3	1	10	25	65	106
Data Officer (DO)	1	5	2	14	25	59	106

Further analysis on the above data was conducted to determine the percentage of satisfaction with adherence on for each KPI. This was essential as it helped the study identify strengths and gaps in current data governance posture.

Table 17: Percentage Level of Satisfaction with the Data Governance KPIs Adherence

KPI	**% Very Dissatisfied (0)**	**% Dissatisfied (1)**	**% Neutral (2)**	**% Somewhat Satisfied (3)**	**% Satisfied (4)**	**% Very Satisfied (5)**
Data Quality (DQ)	1.89	3.77	1.89	13.21	24.53	54.72
Data Management (DM)	1.89	4.72	2.83	10.38	25.47	54.72
Data Security (DS)	0.94	4.72	1.89	9.43	25.47	57.55
Data Protection (DP)	0.94	2.83	0.94	14.15	24.53	56.60
Data Storage (DST)	0.94	3.77	0.94	14.15	23.58	56.60
Data Accessibility (DA)	0.94	2.83	1.89	14.15	25.47	55
Data Processing (DPS)	0.94	3.77	0.94	13.21	25.47	55.66
Metadata (MD)	0.00	4.72	0.94	14.15	24.53	55.66
Data Access Control (DAC)	1.89	4.72	0.94	9.43	24.53	58.49

Strict Enforcement (SE)	0.94	2.83	2.83	13.21	20.75	59.43
Data Encryption (DE)	1.89	4.72	2.83	13.21	20.75	56.60
Multi-Factor Auth (MFA)	0.94	4.72	2.83	10.38	21.70	59.43
Data Governance Team (DGT)	0.94	2.83	0.94	12.26	22.64	60.38
Reporting Structure (RS)	0.00	4.72	2.83	10.38	25.47	56.60
Artificial Intelligence (AI)	1.89	2.83	0.94	9.43	23.58	61.32
Data Officer (DO)	0.94	4.72	1.89	13.21	23.58	55.66

Table 17 above revealed that Artificial Intelligence (AI) received the highest approval, with 61.32% "Very Satisfied" and 23.58% "Satisfied," indicating strong confidence in AI capabilities. Digital Governance (DGT) and Strict Enforcement (SE) also performed well, with 60.38% and 59.43% "Very Satisfied," respectively, reflecting likely robust governance structures and efficient data workflows. Multi-Factor Authentication (MFA) was also rated highly, with 59.43% "Very Satisfied" and 21.70% "Satisfied," showing likely strong security adoption.

Data Access Control (DAC) and Reporting Structure (RS) were rated at 58.49% and 56.60% "Very Satisfied," respectively, affirming confidence in security policies. Additionally, Data Security (DS) and Data Protection (DP) scored 57.55% and 56.60% "Very Satisfied," reflecting likely strong safeguards for data integrity. Data Quality (DQ), Data Management (DM), and Data Accessibility (DA) received 54.72% and 57.55% respectively, "Very Satisfied," suggesting well-managed processes in ensuring accurate, accessible, and reliable data.

However, Data Encryption (DE), Data Processing (DPS), and Data Officer (DO) received 56.60%, 55.66%, and 55.66% "Very Satisfied," respectively, but with 4.72% dissatisfaction each, indicating potential concerns over adherence to encryption standards, processing efficiency, and

data governance leadership. Data Storage (DST) and Metadata (MD), while rated 56.60% and 55.66% "Very Satisfied," had 3.77% and 4.72% dissatisfaction, suggesting potential issues in scalability and metadata accessibility. Data Access Control (DAC) and Strict Enforcement (SE)despite strong positive ratings, had 6.61% and 2.83% dissatisfaction, indicating a gap Reporting Structure (RS) and Data Governance Team (DGT), although well-rated, showed 4.72% dissatisfaction, suggesting a gap in adherence to Reporting Structure; this underscores the need for implementing a data governance maturity model which will ensure visibility of implementation of each Key Performance Indicators.

From the analysis above its clear that the Department of Defence is aware of the Key Performance Indicators required for effective data governance. However, the analysis above shows that only 54-61% of the respondents are very satisfied that these Key Performance Indicators are being adhered to in governing data within the Department of Defence; hence the need of a data governance maturity Model that would allow DOD to assess compliance.

4.2.9 Additional Factors

In addition, the study sought to determine additional data governance aspects, perceptions and suggestions that if implemented, would improve data governance initiatives within the DoD. This was important in ensuring that the model is comprehensive. The outcomes are depicted in figure 20 below.

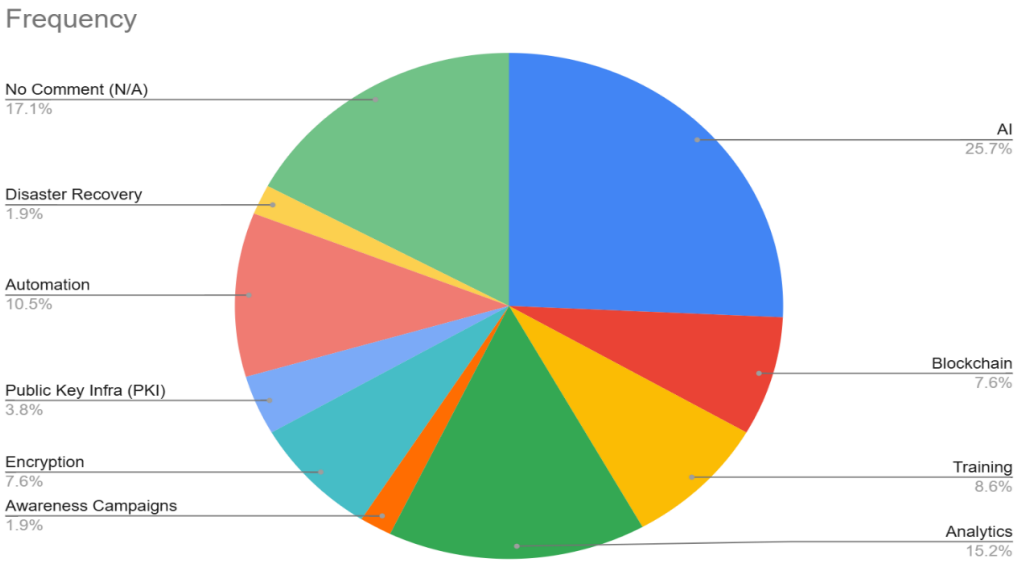


Figure 20: Additional Data Governance factors.

From the analysis above, *Artificial Intelligence (AI)* was the most suggested aspect accounting for *25.7%* of all responses, indicating a strong perception of AI's role in advancing data governance capabilities. *Analytics* followed at *15.2%,* highlighting the importance of data-driven decision-making. *Automation* at *10.5%,* reflected the need for efficiency in governance processes. Additional responses included *Training (8.6%), Blockchain (7.6%), Public Key Infrastructure (PKI) (3.8%), Encryption (1.9%), Awareness Campaigns (1.9%),* and *Disaster Recovery (1.9%).* These insights collectively underscore a strong demand for the incorporation of emerging technologies, workforce capability development, and robust security infrastructure in bolstering DoD data governance frameworks.

4.3 Findings

To achieve study objective two, which was to derive a Data Governance Maturity Model (DGMM) based on the identified Key Performance Indicators (KPIs), this section presents the application of Principal Component Analysis (PCA). PCA was instrumental in reducing the dimensionality of the KPIs by identifying patterns of correlation and grouping them into coherent underlying components. The extracted components formed the basis for constructing the maturity model, which ensured that each dimension reflects a distinct aspect of data governance.

4.3.1 Principal Component Analysis

The Principal Component Analysis (PCA) was conducted to reduce the dimensionality of 16 data governance KPIs, aiming to identify underlying components that capture the maximum variance.

4.3.2 Variables

To facilitate multivariate analysis and improve interpretability, the variables were coded as shown in the table 18 below;

Table 18: Variables

Code	KPI	Meaning
DQ	Data Quality	Ensuring the data is accurate, consistent, and reliable.
DM	Data Management	Managing data from creation to disposal effectively.
DS	Data Security	Safeguarding data against unauthorized access and breaches.
DP	Data Privacy	Ensuring compliance with data privacy laws and regulations.
DST	Data Stewardship	Assigning roles for responsible data oversight.
DA	Data Architecture	Structuring and managing data storage, retrieval, and organization.

DPS	Data Policies & Standards	Enforcing policies and standards across the data lifecycle.
MD	Meta Data Management	Handling metadata to improve data discovery and context.
DAC	Data Accuracy	Maintaining correctness and precision of data.
SE	Strict Enforcement	Implementing strict governance enforcement policies.
DE	Data Encryption	Using encryption methods to protect data.
MFA	Multifactor	Securing data with additional authentication layers.
DGT	Data Gov Team	Ensuring that a dedicated data governance team.
RS	Reporting Structure	Defining the structure for data reporting and accountability.
AI	AI Analytics	Leveraging artificial intelligence for data analysis and insights.
DO	Data Officer	Designating roles responsible for data governance compliance.

4.3.3 Correlation Matrix

The figure 21 below shows a Correlation Matrix which depicts the correlation coefficients between pairs of variables. In this study, the Correlation Matrix was essential in Key Performance Indicators (KPIs) analysis since it helped identify dependencies between KPIs. A KPI with Strong positive correlations with multiple others suggested that, improving that KPI would have a broad impact across the governance implementation.

		DQ	DM	DS	DP	DST	DA	DPS	MD	DAC	SE	DE	MFA	DGT	RS	AI	DO
Correlation	DQ	1.000	.855	.736	.633	.733	.611	.724	.843	.511	.407	.622	.511	.284	.322	.754	.610
	DM	.855	1.000	.737	.744	.863	.749	.779	.789	.552	.220	.431	.325	.316	.152	.715	.621
	DS	.736	.737	1.000	.586	.642	.600	.669	.679	.522	.286	.477	.351	.256	.200	.693	.568
	DP	.633	.744	.586	1.000	.632	.534	.565	.560	.586	.320	.441	.369	.363	.226	.687	.615
	DST	.733	.863	.642	.632	1.000	.721	.707	.702	.474	.189	.443	.264	.222	.228	.641	.519
	DA	.611	.749	.600	.534	.721	1.000	.951	.849	.545	.217	.394	.297	.234	.125	.652	.592
	DPS	.724	.779	.669	.565	.707	.951	1.000	.938	.504	.201	.410	.270	.201	.111	.672	.543
	MD	.843	.789	.679	.560	.702	.849	.938	1.000	.461	.257	.448	.276	.179	.166	.636	.468
	DAC	.511	.552	.522	.586	.474	.545	.504	.461	1.000	.821	.545	.646	.767	.720	.798	.860
	SE	.407	.220	.286	.320	.189	.217	.201	.257	.821	1.000	.715	.821	.658	.895	.678	.721
	DE	.622	.431	.477	.441	.443	.394	.410	.448	.545	.715	1.000	.858	.365	.618	.858	.721
	MFA	.511	.325	.351	.369	.264	.297	.270	.276	.646	.821	.858	1.000	.472	.720	.798	.860
	DGT	.284	.316	.256	.363	.222	.234	.201	.179	.767	.658	.365	.472	1.000	.552	.545	.616
	RS	.322	.152	.200	.226	.228	.125	.111	.166	.720	.895	.618	.720	.552	1.000	.574	.619
	AI	.754	.715	.693	.687	.641	.652	.672	.636	.798	.678	.858	.798	.545	.574	1.000	.921
	DO	.610	.621	.568	.615	.519	.592	.543	.468	.860	.721	.721	.860	.616	.619	.921	1.000

Figure 21: Correlation Matrix

Picking a few examples from the findings in the above 21;

- Data Quality (DQ) & Data Management (DM) → 0.855: A strong link suggests that improving data management practices will significantly enhance data quality and consistency.
- Artificial Intelligence (AI) & Data Officer (DO) → 0.921: A very strong relationship suggested role of leadership in maximizing AI's impact to drive data governance maturity.
- Data Access Control (DAC) & Strict Enforcement (SE) → 0.821: A very strong relationship suggested that strict enforcement is essential to ensuring effective control over data.

The analysis above indicated that the variables had similar underlying attributes, which presented an opportunity to select one representative variable from each pair where applicable in order to avoid redundancy.

4.3.4 PCA Scree Plot Analysis

To interpret and visualize how the total variance was distributed across principal components, the Study conducted a Screeplot Analysis where:

a. The **y-axis** (Eigenvalue) represented how much *variance* each principal component explained.
b. The **x-axis** (Component Number) showed each principal component, in order of how much variance it explains.

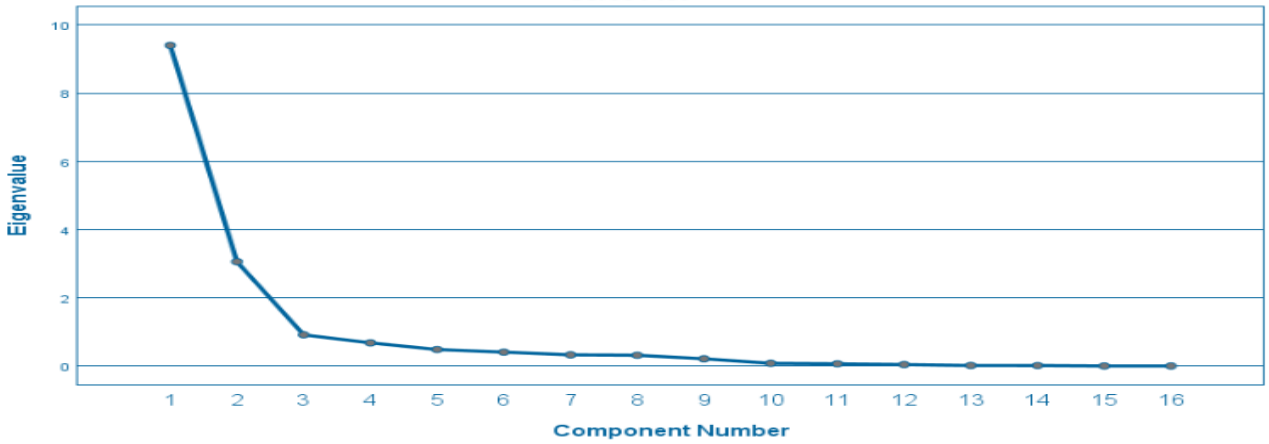


Figure 22: Scree Plot Showing Explained Variance by Components

The first principal component (DQ) captured the highest eigenvalue (about 9.5–10) in the dataset, showing that it explains the largest proportion of total variance in the dataset. Subsequent components captured progressively smaller amounts of variance.

To determine the percentage variance of the principal components, Exploratory Principal Component Analysis (EFA) was conducted as shown in figure 23 below.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.402	58.761	58.761	9.402	58.761	58.761	8.041
2	3.058	19.111	77.872	3.058	19.111	77.872	6.988
3	.914	5.711	83.584				
4	.677	4.231	87.815				
5	.482	3.012	90.827				
6	.406	2.536	93.362				
7	.326	2.039	95.401				
8	.315	1.967	97.368				
9	.210	1.315	98.683				
10	.079	.493	99.176				
11	.064	.401	99.577				
12	.042	.262	99.839				
13	.013	.083	99.922				
14	.013	.078	100.000				
15	-2.859E-15	-1.787E-14	100.000				
16	-6.313E-15	-3.945E-14	100.000				

Figure 23: Explained Variance of Data Governance KPIs Using Principal Component Analysis.

Figure 23 above revealed that principal components (***Data Quality*** and ***Data Management***) explained about *78%* of the total variance combined, suggesting that these are the strongest indicators of governance maturity and they well represented the original data set.

Lower variance in other components indicated that while they play a role, they contribute less to distinguishing governance effectiveness across organizations. This analysis suggested the need to prioritize high-variance KPIs in the data governance maturity model.

4.3.5 Communalities

To establish much of each original variable's variance was explained by the extracted components, communalities were extracted as shown in the figure 24 below. The *"*column" shows the initial communality (set to 1.000 for all variables). The "*Extraction"* column shows the proportion of

variance explained by the underlying factors after extraction in figure 23 above. Higher extraction values indicated a strong relationship between the variable and the principal factors as shown in figure 4.4 below, while lower values suggested a weaker connection.

	Initial	Extraction
DQ	1.000	.783
DM	1.000	.877
DS	1.000	.658
DP	1.000	.578
DST	1.000	.738
DA	1.000	.779
DPS	1.000	.860
MD	1.000	.841
DAC	1.000	.811
SE	1.000	.913
DE	1.000	.688
MFA	1.000	.816
DGT	1.000	.524
RS	1.000	.792
AI	1.000	.930
DO	1.000	.871

Figure 24: Communalities

(i) High communalities. Variables like Artificial Intelligence *(.930), Data Management (.877),* and *Data Officer (.871)* have high extraction values, indicating that a large percentage of their variance was explained by the extracted components.

(ii) Lower communalities. Variables such as *Data Governance Team (.524)* and *Data Protection (.578)* had lower communalities, meaning they shared less variance with the common factors and may be less influenced by them.

High communality KPIs such as Artificial Intelligence, Data Management and Data Officer revealed that they are key drivers of data governance whereas low communality KPIs suggested that they have slightly less in common with the extracted components, though still they may influence success of data governance implementation in other ways.

4.3.6 Factor Loadings

The study Utilized the Principal Components extraction method without rotation, which revealed two significant components as shown in the figure 25 below. The Overall loadings ranged from -0.511 to 0.956, indicating the correlation strength between each KPI and the two components. This provided a simplified structure for understanding the key dimensions of data governance within the dataset.

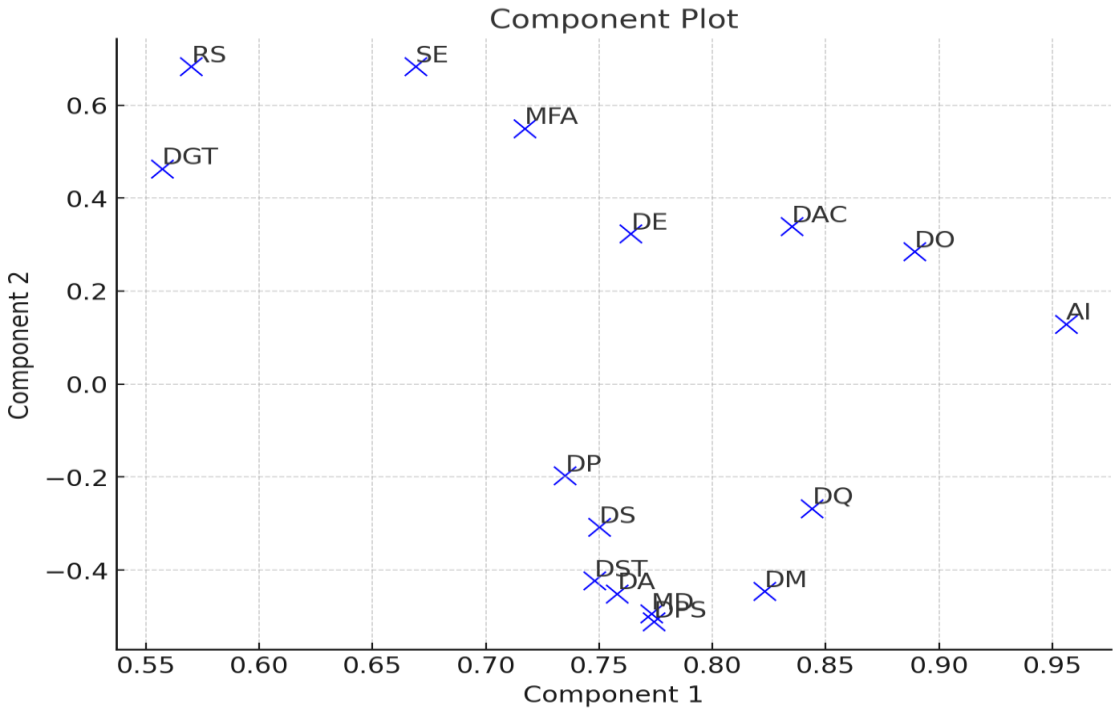


Figure 25: Component Plot

The component plot above revealed two distinct dimensions:

a) ***Component 1*** which was characterized by high positive loadings from KPIs such as DQ (Data Quality, 0.844),AI (Artificial Intelligence, 0.956), and DO (Data Officer, 0.889). This suggested a more focus towards effectiveness of data quality and overall data management practices as well as utilizing AI in data driven processes.

b) ***Component 2*** exhibited a more varied loading pattern, with Strict Enforcement (SE, 0.683) and Data Policies and Standards (DPS, -0.511) displaying an inverse relationship. The negative loading suggests that as stringent data governance measures increase, the influence of Data Policies and Standards diminishes. Furthermore, this discrepancy indicates that strict enforcement (SE) of data governance measures does not necessarily lead to stronger adherence to data policies and standards (DPS). This misalignment is likely attributed to the complex and decentralized nature of the DoD environment, where certain units operate autonomously, and security measures are sometimes implemented independently of overarching policy frameworks.

4.3.7 Component Correlation Matrix

The component correlation matrix revealed how closely related the two extracted factors are. The figure below represents the Component Correlation Matrix from a principal component analysis (PCA) with oblique rotation (Oblimin with Kaiser Normalization). From the findings, ***Diagonal*** values **(1.000)** indicated that each component correlates perfectly with itself while ***Off-diagonal*** value **(.412)** indicated a moderate positive correlation between the two components as shown in the figure 4.6 below. Is indicated that the variables are related but still have some distinctiveness.

Component	1	2
1	1.000	.412
2	.412	1.000

Figure 26: Component Correlation Matrix

4.3.8 Interpretation of Key Variables

4.3.8.1 Highly Correlated Variables

The below variables exhibited high correlation with each other as shown below;

a. DQ and DM (0.855).

c. AI and DO (0.921).

d. DAC and SE (0.821)

This suggested that the variables had similar underlying attributes, which presented an opportunity to select one representative variable from each pair where applicable in order to avoid redundancy in the model.

4.3.8.2 Key Contributors from Component Analysis

Drawing from the results of the Component Score Coefficient Matrix;

a. **Component 1:** Focus on organizing and managing data, with DM (Data Management), DPS (Data Policies and Standards), and MD (Metadata) as key contributors.
b. **Component 2:** Focus on data security and access controls, driven by DAC (Data Access Control), DE (Data Encryption), and MFA (Multi-Factor Authentication).

4.3.8.3 Variables with Moderate Correlation across Components

Artificial Intelligence (AI), Data Encryption (DE), and Data Officer (DO) exhibited more balanced correlations across different components and hence useful in capturing variety of dimensions and provide a comprehensive model.

4.3.8.4 Variables with Lower Correlations

Notably, SE and DGT had lower or negative correlations. However, in order to capture the unique aspects that are not explained by other variables, including them could add valuable diversity.

Strict Enforcement (SE) and Data Governance Team (DGT) showed low correlations with Data Architecture (DA), Data Privacy (DP), Data Stewardship (DST), and Data Security (DS). However, their significant loadings on Component 2 (0.683 and 0.463, respectively) indicated that they represent a distinct dimension of data governance, hence including them would be essential to capture unique aspects beyond the other variables.

4.3.9 Summary of Primary Variables

From the analysis above and in Table 4.5 below, the primary variables selected for the model were Data Policies and Standards (DPS, loading 0.82 on Component 1), Data Accuracy (DAC, loading 0.75 on Component 1), Data Encryption (DE, loading 0.76 on Component 2), and Multi-Factor Authentication (MFA, loading 0.72 on Component 2), as they significantly loaded on the components and covered a range of different dimensions. Strict Enforcement (SE), Reporting Structure (RS), and Data Governance Team/Office (DGT) were included to capture unique elements not covered by the primary variables. To minimize redundancy, one variable was chosen from each of the three pairs of highly correlated variables (DM and DS, DM and AI, DS and AI, all correlations above 0.7). This method guaranteed that the model remained reliable and minimized redundancy.

Table 19: Summary of Variables Selected for the Model.

Category	Selected Variables (KPIs)	Justification
Primary Variables	Data Policies (DPS), Data Accuracy. (DAC), Data Encryption (DE), Multi-Factor Authentication (MFA).	Significantly loaded on components and cover different dimensions.
Distinct Elements	Strict Enforcement (SE), Reporting Structure (RS), Data Governance Team/Office (DGT).	To include unique elements not covered by primary variables.
From Highly Correlated Pairs	Data Management (DM), Data Security. (DS), Artificial Intelligence & Analytics (AI).	Chosen from pairs to avoid multicollinearity.

As depicted in table 19 above, primary variables were chosen for strong factor loadings and dimensional coverage, distinct variables for unique aspects, and a selection from highly correlated pairs to avoid multicollinearity. This aligned with the study objective of identifying necessary data governance KPIs for deriving a comprehensive data governance maturity model for Kenya department of Defence.

4.4 Derivation of the Mathematical Model

To assign maturity level scores, the study used the Hybrid Maturity Model in table 4.6 below, which was derived from the Capability Maturity Model Integration (CMMI) and Gartner Maturity Model, and benchmarking with the guiding principles of ISO/IEC 38500-1.

Table 20: Hybrid Maturity Level Model.

Maturity Level	Description	Score (X)
Unaware/None	No formal governance processes.	0
Aware	Basic recognition of governance needs.	1
Defined	Formal processes are in place, but not standardized.	2
Managed	Processes are managed, measured, and controlled.	3
Optimized	Processes are continuously improved based on metrics.	4

Mature	Fully institutionalized Data governance and is a strategic capability.	5

From the table 20 above, the Hybrid maturity model distinguishes itself through its specific focus on data governance maturity. This offers a more granular and detailed progression compared to the broader process-oriented CMMI and the general data-centric Gartner models. This specialization is particularly beneficial for the department of Defence context, where precise evaluation and improvement of data governance practices are critical for operational effectiveness, security, interoperability, and data-driven decision-making.

The following Weighted Composite Index Model modeling equation was used to compute weights necessary for computing the Maturity Index **Y**, in this case the Data Governance Maturity Index (DGMI).

$$\mathbf{Y = W_1V_1 + W_2 V_2 + \ldots + WnVn}$$

Where,

Y = Data Governance Maturity Index. (MI)

W = Weights

V = Data Governance Key Performance Indicators (KPIs)/ (User assessment Score per question)

n = Number of assessment questions

Assuming that the coefficients of the assessment questions are constant, so that $\mathbf{W=W_1=W_2=\ldots W,}$ Consequently, **W** will be the weight, thus

$$\boldsymbol{MI = W\ V_1 + WV_2 + WV_3 + \ldots + WVn.}$$

Equation 2: Mathematical Model

Since W is common,

$$\boldsymbol{MI = W\ (V1 + V2 + V3 + \ldots Vn)}$$

This study had 16 questions that directly assessed the Data Governance KPIs (Variables), but after Principal Component Analysis, 10 KPIs were selected therefore case, n=10 and the maximum score that the user could have in a scale of 0 to 5 was; 5*10 = 50.

Putting this back into the previous equation, then

MI =V1/50 + V2 /50 + V3 /50+ … V10 /50

Therefore,

MI = 1/50(V1 + V2 + V3 + … V10)

Hence W = 1/50 = 0.002

The relevant weight for the DGMI model was 0.002, as shown on the results above.

To compute the MI as a percentage factor,

MI = 0.02 (V1+ V2+V4 +…. V10) * 100

Hence,

MI= 0.02 (V1+V2+V3+…V10) %

Equation 3: Percentage Maturity Index

4.4.1 DGMI Mathematical Model

Achieving a specific Maturity Index indicated the organization's overall posture and the percentage of Data Governance maturity it represents. Below is the equation for computing Data Governance Maturity Index (DGMI).

DGMI= MI%

4.4.2 Model Scenarios

Below demonstrations depict the three model scenarios.

4.4.2.1 Best Case

The optimal situation is reached when the total of the ten Data Governance KPIs scores equals 50. That is, DGMI = V1 + V2 +…V10 = 50
By substituting back in equation 3,

MI= 0.002 (V1+V2+V3+…V10) %

MI = (0.02*50) %

MI = 100%

This depicts that the organization's Data Governance program is fully integrated with continuous improvement (mature) in line with ISO/IEC 38500-1 guiding principles.

4.4.2.2 Average Case

The typical situation is in the middle, with the organization 50% mature and 50% in the process of establishing data governance. The typical case scenario on a scale of 0 to 5, as used in this study, is when the organization being evaluated received an average score of 2.5 on each of the 10 assessment questions, or a total score of 40, which leans towards a neutral score.

MI= 0.02 (V1+V2+V3+…V10) %

MI = (0.02 25) %*

Average Scenario Equation: *MI = (0.02 *25) % =**50%***

4.4.2.3 Worst Case

The worst-case scenario is the converse of the best-case scenario whereby the assessment scores depict that the organization are least mature in terms for as long as ISO/IEC 38500-1 guiding principles is concerned; wherefore MI is below average. Example: Assuming that every KPI being assessed gets an average score of 1.

Therefore,

MI = (0.02 *10) %

MI = 20%

4.4.2.4 Threshold summary

Below in table 21 is a summary of the maturity thresholds and the assessment scale.

Table 21: Maturity Thresholds

Level	Maturity Level	Score	DGMI (%)	Description
Level 1	Unaware	0	0%	Organization has no formal data governance.
Level 2	Aware	1	20%	Basic awareness of data governance exists.
Level 3	Defined	2	40%	Data governance processes are defined but not fully operational.

Level 4	Managed	3	60%	Data governance processes are actively managed and followed.
Level 5	Optimized	4	80%	Processes are optimized for efficiency and effectiveness.
Level 6	Mature	5	100%	Data governance is fully integrated and continually improved.

4.4.2.5 Assessment Scale

A range of scores from 0 to 80, or percentages from 0% to 100%, are depicted by the DGMI scale below.

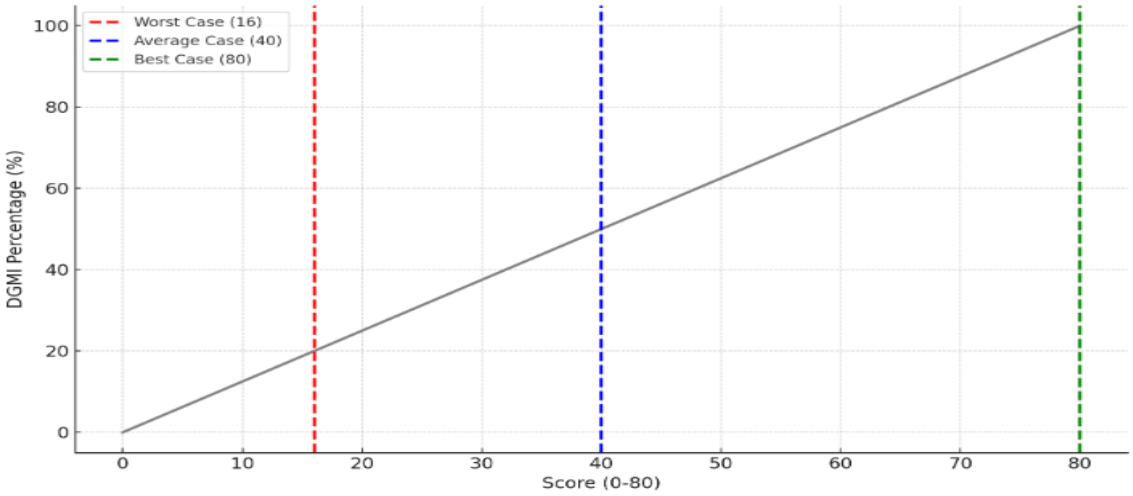


Figure 27: Maturity Assessment Scale

With the green zone denoting a mature data governance program, the blue zone representing an average level of implementation, and the red zone denoting a critical stage where the data governance program needs immediate attention and corrective action.

4.5 Analysis Discussions

In this section, the study discussed the Exploratory Principal Component Analysis, factor extraction, factor retention, factor rotation, interpretation and concluded by describing the architecture of the model and its validation.

4.5.1 Exploratory Principal Component Analysis

Exploratory Principal Component Analysis (EFA) played a crucial role in developing Data Governance Maturity Model (DGMM) by grouping relevant KPIs and uncovering the dimensions that define data governance maturity. EFA conducted in this study followed a structured approach

to identify the underlying factors related to data governance performance. The Principal Component Analysis aimed to reduce the complexity of numerous KPIs (Key Performance Indicators) into a manageable set of factors that could inform the development of a Data Governance Maturity Model (Costello & Osborne, 2019).

The correlation matrix provided in the table 22 below, Exploratory Principal Component Analysis (EFA) served as a critical tool for understanding the relationships among variables in the context of the Data Governance Maturity Model for the Kenya Department of Defence. The matrix showcased the interdependencies among various dimensions, such as Data Quality (DQ), Data Management (DM), Data Security (DS), and others, which highlighted their potential contributions to underlying latent factors.

The correlation values in the matrix ranged from weak to strong, with most variables demonstrating significant positive correlations. For instance, the relationship between Data Quality (DQ) and Data Management (DM) showed a strong correlation of 0.855, reflecting the intrinsic link between managing data effectively and ensuring its quality. Similarly, from table 22 below, Data Accessibility (DA) and Data Policies and Standards (DPS) exhibited a near-perfect correlation of 0.951, likely due to the overlapping components within these dimensions. On the other hand, weaker correlations, such as between Reporting Structure (RS) and Data Management (DM) (0.152), suggested minimal interaction between these dimensions, potentially due to their distinct roles within the maturity model framework

Table 22: Communalities

Communalities		
	Initial	**Extraction**
DQ	1.000	.783
DM	1.000	.877
DS	1.000	.658
DP	1.000	.578
DST	1.000	.738
DA	1.000	.779
DPS	1.000	.860
MD	1.000	.841
DAC	1.000	.811
SE	1.000	.913
DE	1.000	.688
MFA	1.000	.816
DGT	1.000	.524
RS	1.000	.792
AI	1.000	.930
DO	1.000	.871
Extraction Method: Principal Component Analysis.		

4.5.4 Initial Communalities

Initially, all variables had communalities set to 1.000, reflecting the assumption that each variable contributes its full variance to the analysis. This baseline enabled the extraction process to determine how much of this variance remained represented in the final factor structure.

4.5.4.1 Extraction Communalities

The extraction communalities from table 4.8 above indicated the proportion of variance for each variable that was explained by the latent factors. Some variables exhibited high communalities, signifying strong representation within the factor model. For example, Artificial Intelligence **(AI)** had the highest communalities value at 0.930, demonstrating that nearly all its variance was explained by the extracted factors. Similarly, Strict Enforcement (SE) showed a value of 0.913, indicating that it plays a central role in defining the maturity model.

Other variables, such as Data Policies and Standards (DPS) at 0.860 and Data Access Control (DAC) at 0.811, also demonstrated significant contributions, suggesting they are integral to the underlying constructs of the maturity model. Moderate communalities values, such as those for Data Quality (DQ) at 0.783 and Data Management (DM) at 0.877, highlighted their substantial influence on the factors while allowing room for further refinement.

On the other hand, several variables showed lower communality values, suggesting that they might be impacted by more dimensions than those found in the current factor structure. Data Governance Team (DGT), for instance, had the lowest value (0.524), indicating that it likely represented elements that the extracted components did not fully capture.

4.5.5 Analysis of Total Variance Explained in Principal Component Analysis

Table 23 below presents a detailed summary of the proportion of total variance in the dataset explained by each extracted component during the Principal Component Analysis (PCA). This analysis was fundamental in determining the number of latent factors to retain by identifying the components that meaningfully captured the underlying structure of the Data Governance Maturity Model (DGMM). Only components with eigenvalues greater than 1 were considered significant contributors, which ensured that the retained factors represented the most impactful dimensions of data governance maturity.

Table 23: Principal Component Analysis - Total Variance Explained

Total Variance Explained							
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
Component	Total	% Of Variance	Cumulative %	Total	% Of Variance	Cumulative %	Total
1	9.402	58.761	58.761	9.402	58.761	58.761	8.041
2	3.058	19.111	77.872	3.058	19.111	77.872	6.988
3	.914	5.711	83.584				
4	.677	4.231	87.815				
5	.482	3.012	90.827				
6	.406	2.536	93.362				
7	.326	2.039	95.401				
8	.315	1.967	97.368				
9	.210	1.315	98.683				
10	.079	.493	99.176				
11	.064	.401	99.577				
12	.042	.262	99.839				
13	.013	.083	99.922				
14	.013	.078	100.000				
15	-2.859E-15	-1.787E-14	100.000				
16	-6.313E-15	-3.945E-14	100.000				
Extraction Method: Principal Component Analysis.							
a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.							

4.5.5.1 Initial Eigenvalues

From table 4.9 above, the eigenvalues in the initial solution indicate the amount of variance captured by each component. The first component had an eigenvalue of 9.402, accounting for 58.761% of the total variance, making it the dominant factor. This high percentage indicated that the first component captured a significant proportion of the shared variance among variables.

The second component had an eigenvalue of 3.058, contributing 19.111% of the variance from table 4.9 above. Together, these two components explained a cumulative 77.872% of the total variance, demonstrating their importance in summarizing the dataset. The remaining components, starting with the third, had eigenvalues less than 1 and contributed progressively smaller amounts, with the third component explaining only 5.711% and subsequent components adding even less. These diminishing contributions suggested that additional components provided limited explanatory power hence were not retained in the final factor solution.

4.5.5.2 Extraction Sums of Squared Loadings

After the extraction process, the sums of squared loadings confirmed that only the first two components were to be retained. These components collectively explained 77.872% of the variance as shown in table 4.9 above, validating their impact in capturing the unobservable aspects of data governance practices as well as underlying relationships among the variable.

4.5.5.3 Rotation Sums of Squared Loadings

The rotation process redistributed the variance across components to enhance interpretability without altering the total explained variance. Following rotation, the variance for the first component decreased to 8.041 as shown from table 4.9 above, while the variance explained by the second component increased to 6.988. This redistribution improved the balance of explanatory power between the two components, making the factor model easier to interpret. The rotation ensured that each factor contributed significantly to the explanation of variance, reducing the dominance of any single component and allowing for a more complex understanding of the data.

The analysis highlighted the strong contributions of the first two components, which dominated the variance explanation with a combined total of 77.872%. This finding suggested that the dimensions of data governance maturity were effectively summarized by these two factors.

Notably, the remaining components contributed minimal variance, with the third component explaining just 5.711% from table 4.9 above, and the subsequent components adding progressively less.

4.5.5.4 Implications for the Maturity Model

With the first two components accounting for 77.872% of the total variance from table 4.9 above, the model exhibited strong dimensionality and robustness. This indicated that the essential aspects of data governance maturity were effectively represented by focusing on these two latent factors. Additionally, the rotation process enhanced the interpretability of the model, allowing a balanced representation of the factors and ensuring that each contributed significantly to the overall explanation of variance.

4.5.6 Analysis of the Scree Plot for Principal Component Analysis

The scree plot displayed in figure 28 below provided a visual representation of the eigenvalues corresponding to the components derived during the Principal Component Analysis. This tool was essential for identifying the optimal number of components to retain in the analysis, based on the "elbow criterion."

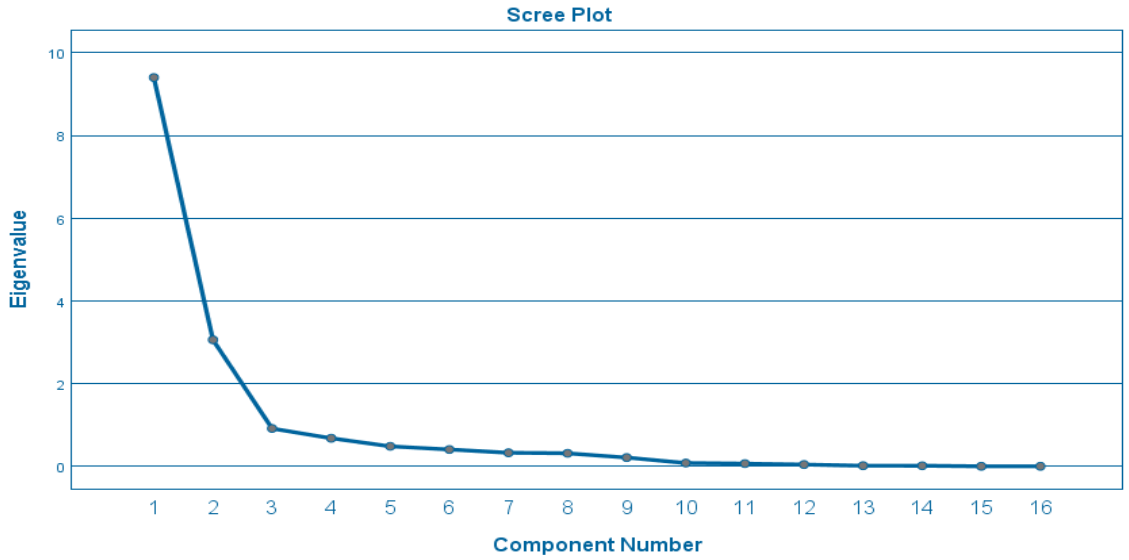


Figure 28: Principal Component Analysis - Scree Plot

The scree plot in figure 28 above, showed a steep decline in eigenvalues from Component 1 to Component 3, followed by a more gradual leveling off from Component 4 onward. The first component had the highest eigenvalue, which was consistent with its dominant contribution to the total variance, as discussed in the total variance explained table 28 above. The second component also had a high eigenvalue, signifying its substantial explanatory power. Together, these two components explained a significant portion of the dataset's variance, as evidenced by the steep drop after these initial components.

Beyond the second component, the slope of the plot flattened considerably. This leveling off suggested that the remaining components contributed only marginally to the variance and was likely to represent noise or less meaningful dimensions of the data. The "elbow" of the plot, where the slope visibly changed, appeared at Component 3, indicating the point where the additional components ceased to contribute substantially to explaining the variance. From the figure 28 above, the study observed the following:

i) Dominance of Initial Components: The steep decline from Component 1 to Component 2 demonstrated the dominance of these components. Component 1 had the largest eigenvalue, reflecting its role as the primary factor capturing the largest proportion of shared variance.
ii) Flattening of Subsequent Components: The gradual flattening of the curve beyond Component 3 confirmed that the explanatory power of additional components diminished significantly. The eigenvalues for these components were below 1, indicating that they contributed less variance than an individual variable.
iii) Elbow Criterion: The "elbow" at Component 3 suggested that only the first two components were to be retained, as they captured the meaningful variance within the dataset. Components beyond this point did not provide substantial additional explanatory power.

The scree plot strongly supports the retention of two components for the factor model. This aligned with the earlier analysis of total variance explained from table 28 above, where the first two components accounted for a cumulative 77.872% of the variance. Retaining only these two components ensured that the analysis remained concise, focusing on the most critical dimensions of the dataset while minimizing noise.

4.5.7 Analysis of the Component Matrix for Principal Component Analysis

The component matrix in Table 24 below provided essential insights into the relationships between variables and the two extracted components in the Principal Component Analysis. The factor loadings in the matrix represented the correlation between each variable and the underlying components. This helped the study to identify which variables are aligned with a particular latent factor, and hence support accurate grouping and interpretation.

Table 24: Principal Component Analysis - Component Matrix

Component Matrix[a]		
	Component	
	1	**2**
DQ	.844	-.267
DM	.823	-.446
DS	.750	-.308
DP	.735	-.197
DST	.748	-.423
DA	.758	-.451
DPS	.774	-.511
MD	.773	-.494
DAC	.835	.339
SE	.669	.683
DE	.764	.323
MFA	.717	.550
DGT	.557	.463
RS	.570	.683
AI	.956	.129
DO	.889	.285
Extraction Method: Principal Component Analysis.		
a. 2 components extracted.		

4.5.7.1 Component 1

Most variables had high positive loadings on this component, indicating their strong association with it as shown in table 4.10 above. Notably, Artificial Intelligence (AI) had the highest loading (0.956), suggesting that it was a dominant contributor to this component and played a central role

in the Data Governance Maturity Model. Similarly, Data Quality (DQ) (0.844), Data Management (DM) (0.823), and Data Policies and Standards (DPS) (0.774) exhibited significant loadings, emphasizing their importance in the shared variance captured by Component 1. The variables associated with Component 1 appeared to represent structural or technical aspects of data governance, such as quality assurance, management processes, and data handling systems. These dimensions formed the backbone of the maturity model.

4.5.7.2 Interpretation of Component 2

The second component (Component 2) captured additional variance not explained by Component 1 and highlighted dimensions that are distinct yet complementary. Strict Enforcement (SE) showed the highest loading on Component 2 (0.683), underscoring its importance in defining this factor. Strict Enforcement (SE) appeared to represented an operational and compliance aspect of data governance.

4.5.7.3 Variables with Dual Contributions

Some variables contributed meaningfully to both components, reflecting their multifaceted roles in the maturity model. For instance, Data Access Control (DAC) contributed to Component 2 (0.339) while maintaining a strong association with Component 1 (0.835). Multi-Factor Authentication (MFA) had moderate loadings on both Component 1 (0.717) and Component 2 (0.550). Similarly, Data Governance Team (DGT) had 0.557 on Component 1 and 0.463 on Component 2 while Data Encryption (DE) had 0.764 on Component 1 and 0.323 on Component 2 as shown in in table 4.10 above. This dual relationship underscores their relevance in both technical and operations aspects of data governance.

4.5.7.4 Implications for the Maturity Model

The pattern matrix underscored the multifaceted nature of the Data Governance. Component 1 focused on technical aspects such as Data Quality, Artificial Intelligence, Analytics, Data Management systems (technology), processes and standards. Component 2 focused on the operational aspects such as Strict Enforcement (SE) or compliance, for effective implementation of data governance and continuous improvement. The presence of cross-loading variables, such as Artificial Intelligence (AI) and Data Governance Team (DGT), highlighted the integrated nature of data governance.

4.5.8 Component Correlation Matrix

The component correlation matrix in table 25 below offered more valuable insights into the relationship between the two extracted components. The correlation between Component 1 and Component 2 was 0.412 which indicated a moderate positive relationship. This suggested that while the components captured distinct dimensions of the data governance, they were not entirely independent and share some degree of overlap.

Table 25: Principal Component Analysis - Component Correlation Matrix

Component Correlation Matrix		
Component	**1**	**2**
1	1.000	.412
2	.412	1.000
Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization.		

The moderate correlation reflected the interdependence between the technical and operational dimensions of data governance as captured by Component 1 and Component 2). Technical elements such as data management, quality, and analytics provided the foundational position which supports operational data governance practices. Similarly, operational elements generated feedback and insights aligned to enhancing operational data governance practices. This assessment aligned with real-world practices governance implementation strikes balance between operations and technical.

4.5.9 Analysis of the Component Score Covariance Matrix for Principal Component Analysis

The component score covariance matrix in table 26 below provided a quantitative measure of the relationship between the two components, highlighting the extent to which the components vary together, their interdependence and shared variance between them. In this analysis, the diagonal entries represent the variances of the individual components, while the off-diagonal entries reflected the covariance between them.

Table 26: Principal Component Analysis - Component Score Covariance Matrix

Component Score Covariance Matrix		
Component	**1**	**2**
1	1.170	.824
2	.824	1.170
Extraction Method: Principal Component Analysis.		
Rotation Method: Oblimin with Kaiser Normalization.		
Component Scores.		

4.5.9.1 Variance of Individual Components

The variances for both components were 1.170, as shown in table 26 above, as seen from the diagonal entries. This value indicated that each component explained a relatively equal amount of variance in the dataset. The similarity in variances showed that both components played a comparable role in representing the underlying structure of the data. Component 1, focused on technical and structural dimensions, while Component 2, on operational and evaluative aspects, contributed equally to the overall data governance. This balance was critical for ensuring that the maturity model captures a comprehensive view of data governance.

4.5.9.2 Covariance Between Components

The covariance between Component 1 and Component 2 was 0.824 as shown in table 26 above, indicating a substantial positive relationship. This value reflected the degree to which the components share variability, signifying their interdependence within the data governance model. The strong covariance highlighted the interconnected nature of technical and structural dimensions in data governance.

The balanced variances emphasized the importance of a unified approach to data governance in that both components should be implemented and refined simultaneously. Neglecting one dimension may weaken the overall data governance program as the two components are closely linked.

4.5.10 Summary

This section developed a mathematical model to assess data governance maturity within the Kenya Department of Defence (DoD). By leveraging Principal Component Analysis, key performance indicators (KPIs) were systematically identified to ensure a balanced and accurate assessment while minimizing redundancy. A Hybrid Maturity Model was formulated by integrating elements

from Capability Maturity Model Integration (CMMI) and Gartner, with an additional "Maturity" level introduced for a more precise evaluation, aligning with ISO/IEC 38500-1 principles. The Data Governance Maturity Index (DGMI) was calculated using a weighted Weighted Composite Index Model, with a standardized weight of 0.002 per KPI, resulting in a structured scoring framework from 0% (Unaware) to 100% (Mature). The model was validated through best-case, average, and worst-case scenarios, demonstrating its effectiveness in providing a standardized, quantifiable measure of data governance maturity.

4.6. Instantiation and Empirical Testing of the Research Artefact

To achieve study objective three, which was to implement a prototype model, a Data Governance Maturity model prototype was developed using a rapid prototyping approach. The following processes were applied to guide the development process:

4.6.1 Requirements Gathering Process

The requirements for the system were collected based on the Data Governance KPIs observed within the Kenya DOD, together with the emerging technologies.

4.6.2 Quick Design Process

A rapid conceptual design of the system's database and modules was created using flow diagrams, as summarized in Figure 36. The design of the different modules is depicted in Figure 29 below.

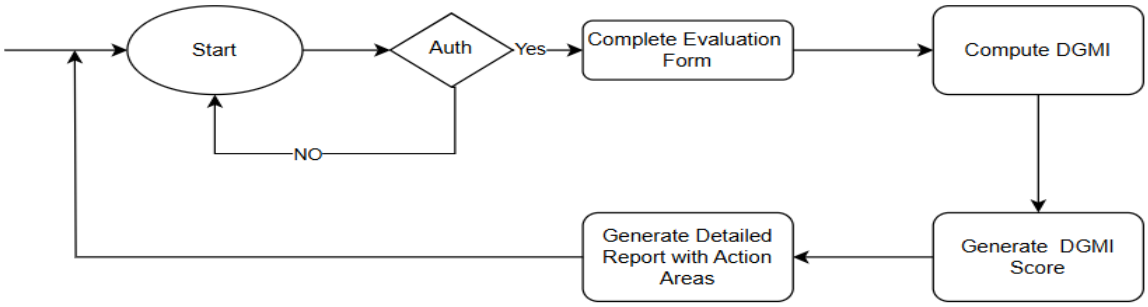


Figure 29: Prototype Flow Diagram

4.6.3 Coding

Python Django was used as the server-side framework to develop the prototype. Django's built-in functionality was utilized for the interactive elements, while CSS3 was employed for styling the user interface. The database was managed using PostgreSQL as the engine.

4.6.4 Prototype Evaluation

The prototype was evaluated using FEDS framework, as outlined in Section 3.2, comparing it against the gathered system objectives. This evaluation method guided the ongoing refinement process.

4.6.5 Testing & Refinement

The code was iteratively refined as system requirements were met. Once a satisfactory version of the system was achieved, it was deployed, allowing real users to register and test all functionalities. Feedback received during this phase informed further development and refinement of both the model and the system.

4.6.6 System Architecture

The central architecture consists of the following key components: User Authentication Module. This module makes sure that the system is only accessible by authorized users who have registered credentials.

4.6.7 User Registration

The User Registration Module is the gateway to the system, where users register by entering their full name, organization, username, email address, and password. System capabilities are only accessible to registered users.

4.6.8 Password Management Module

This module controls the verification and complexity criteria for passwords. It ensures adherence to data security standards by employing MD5 Hashing to safeguard user passwords at the database level.

4.6.9 User Session Handling Module

To ensure security and accountability, this module manages user sessions by setting and tracking sessions when users log in, keeping an eye on actions during active sessions, and ending sessions when users log out.

4.6.10 Maturity Assessment Module

This module pulls test questions from the database and displays them using a structured format. From the DB, Users is prompted with a Data Governance implementation Status quo question that targets to assess the implementation of certain KPIs.

4.6.11 Output Module

This module presents the interactive graphical results of evaluations that have been submitted. Additionally, users can print or download comprehensive DGMI score report that offer suggestions on actions areas for improving the maturity levels of the Data Government.

4.6.12 Core Logic

This module responds to user requests, interprets input, and produces output. Additionally, it communicates with the database, computes the organization's Data Governance Maturity Index, and outputs the results.

4.6.13 Database

The system has one database with manages important tables. Users Table. This table Complements the authentication module by storing all registered user information and confirming user access. All assessment questions, threshold scores, and related suggestions used to gauge the maturity of data governance are stored in the Assessment Questions table. The Questions_Categories table arranges questions based on particular Data Governance KPIs. The User_Assessment table keeps track of finished tests and adds timestamps to indicate when they are submitted.

4.6.14 System UML Diagrams

The DGMI prototype has a number of players, each with distinct duties and responsibilities to guarantee the system's effective operation, security, and upkeep. The main players are: User. System users are the individuals or groups to access the platform to perform various tasks and functions related to Data Governance Maturity assessment. These users typically interact with the system for data entry, analysis, and reporting.

System Administrator. The system administrator is responsible for the management, maintenance, and security of the system. The administrator oversees the technical aspects of the system, ensuring that it operates smoothly and that data governance metrics are accurately captured and reported. Administrator allows a higher level of access and play a critical role in system oversight and security enforcement.

4.6.15 Class Diagram

The class diagram in the figure 30 below represents the system's static structure, showing the classes, attributes, and methods involved in the prototype design. It identifies relationships like associations and aggregations among components such as User Authentication, Historical Data, and graphic user interface elements. Figure 30 was crucial in visualizing the modular architecture of the system, ensuring a clear separation of responsibilities between components. It helped streamline the design by defining methods for user registration, data calculations, and trend analysis while maintaining scalability and clarity.

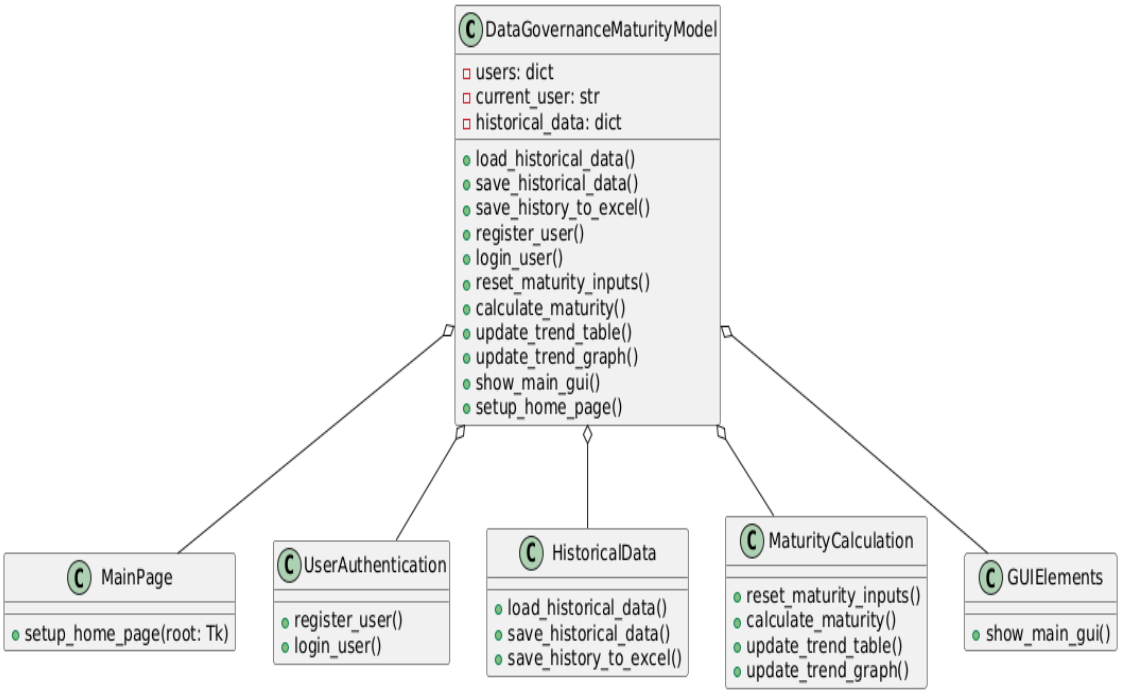


Figure 30: Class Diagram

4.6.16 Activity Diagram

The activity diagram in figure 31 below illustrates the flow of processes within the system, from user login and registration to calculating maturity levels and saving historical data. It highlights the conditional flow (e.g., login success or failure) and the sequential execution of tasks. This diagram was useful for understanding the user journey and system interactions, ensuring all major workflows were accounted for. It helped refine the system's logical flow and ensured smooth transitions between activities like input processing and report generation.

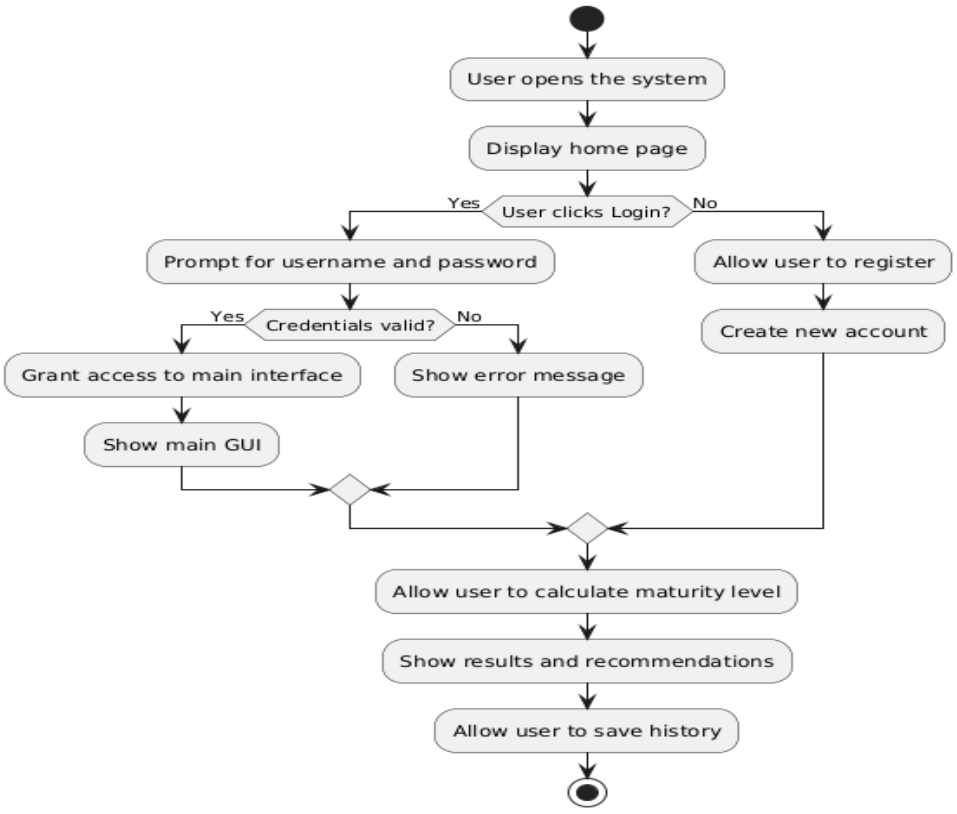


Figure 31: Activity Diagram

4.6.17 Use Case Diagram

The use case diagram in figure 32 below outlines the interactions between users (actors) and the system's functionalities, such as logging in, registering, calculating maturity levels, and exporting history. By categorizing tasks into user roles (e.g., User and Admin), it clarified system requirements and user expectations. This diagram was instrumental in capturing high-level functionality and validating that all essential features were included, ensuring alignment with user needs during prototype development.

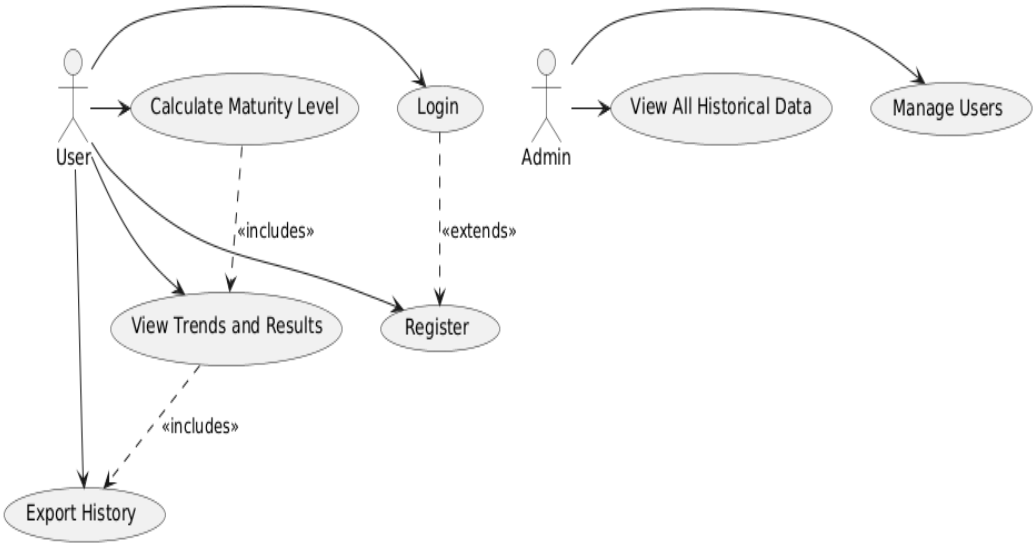


Figure 32: Use Case Diagram

4.6.18 Sequence Diagram

The sequence diagram shows the interaction between system components over time, detailing the order of method calls and responses as show in the figure 33 below. For example, it describes how user inputs flow through components like Authentication and Maturity Calculation, with results saved in Historical Data. This visualization ensured that dependencies and method execution were logically sequenced, which helped prevent runtime issues and facilitated efficient component integration during implementation.

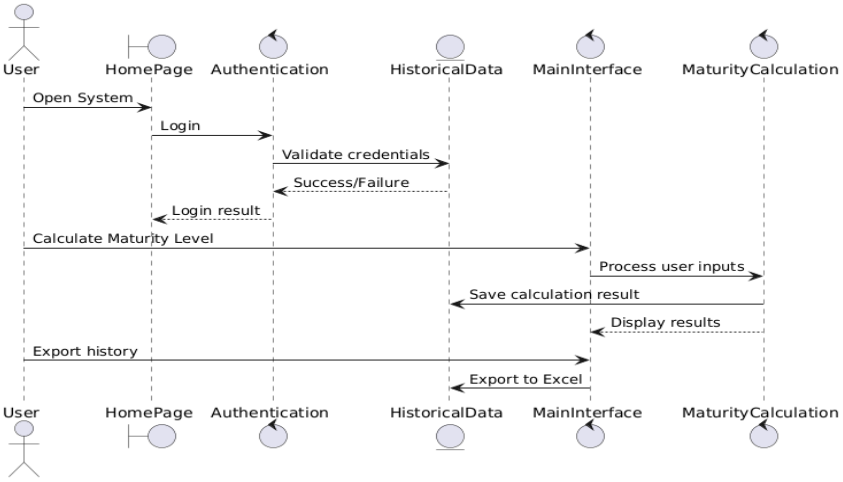


Figure 33: Sequence Diagram

4.6.19 System User Interface

This section depicts various user interfaces across various prototype's functionalities;

4.6.19.1 Home Page

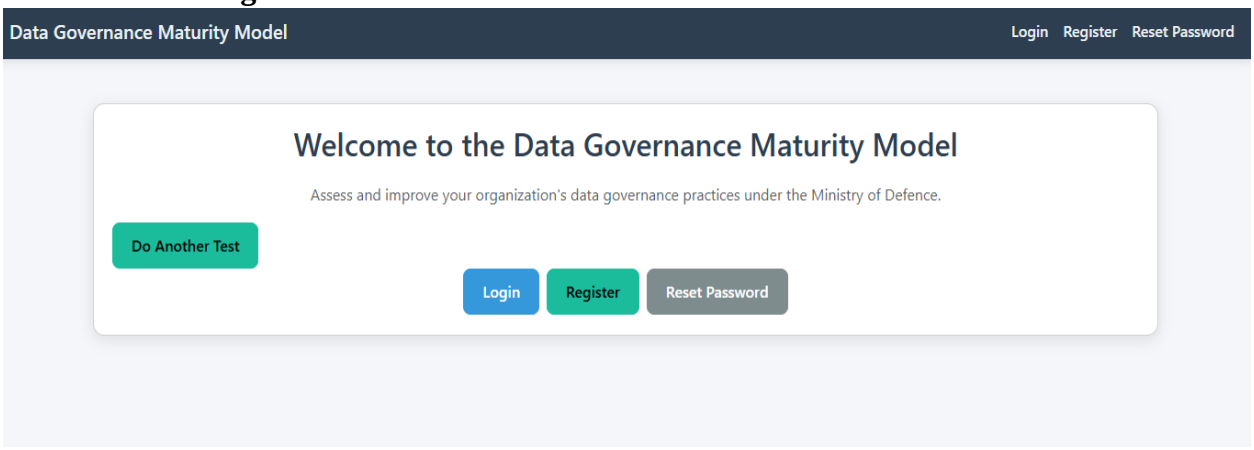


Figure 34: Systems Home Page

4.6.19.2 User Registration

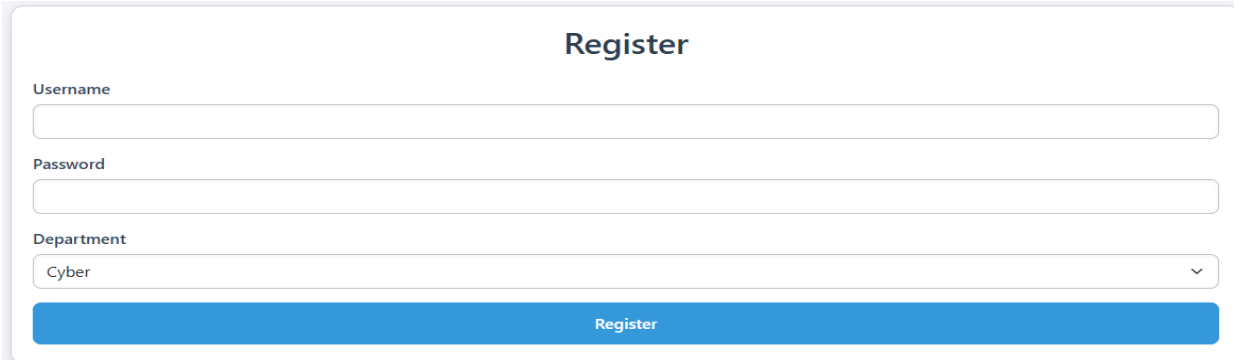


Figure 35: User Registration

4.6.19.3 User Login GUI

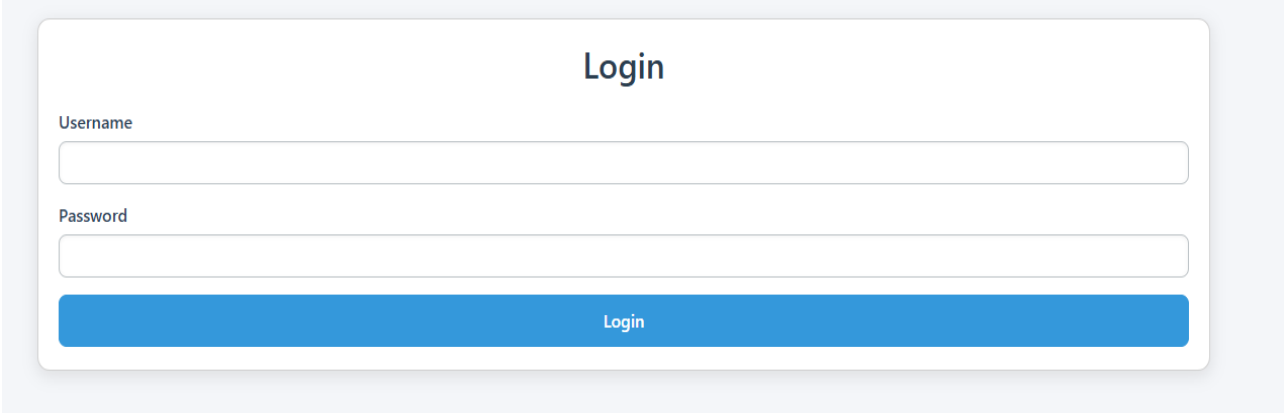


Figure 36: U

4.6.19.4 DGMM_ Evaluation Page

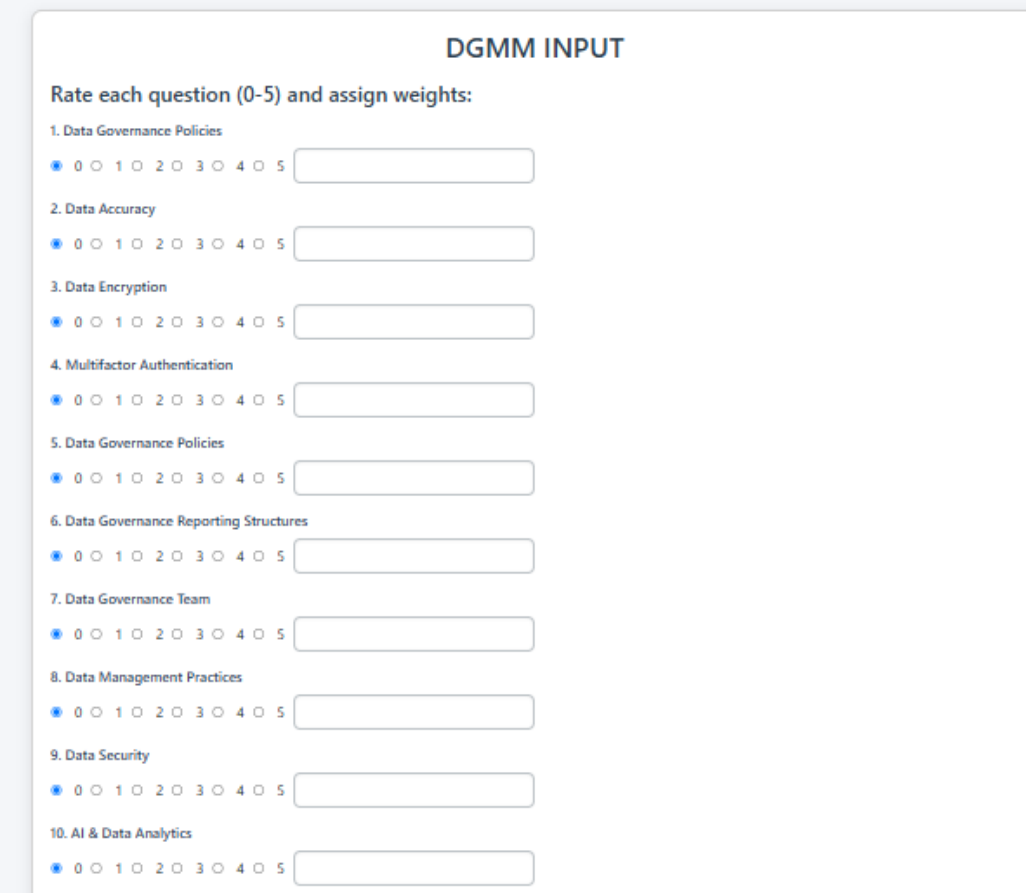


Figure 37: Evaluation page

4.6.19.5 Key Functionalities

Figure 38: Evaluation Example

4.6.19.6 Model output Scale

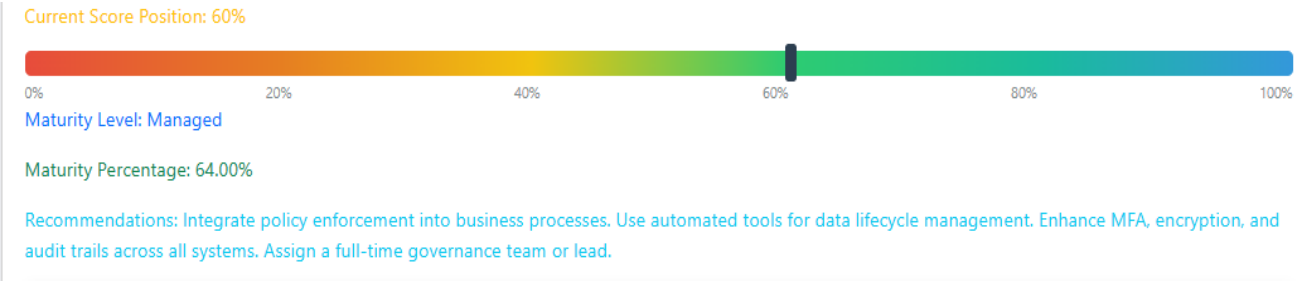


Figure 39: Model output scale

4.6.19.7 Trend Analysis Historical Evaluation

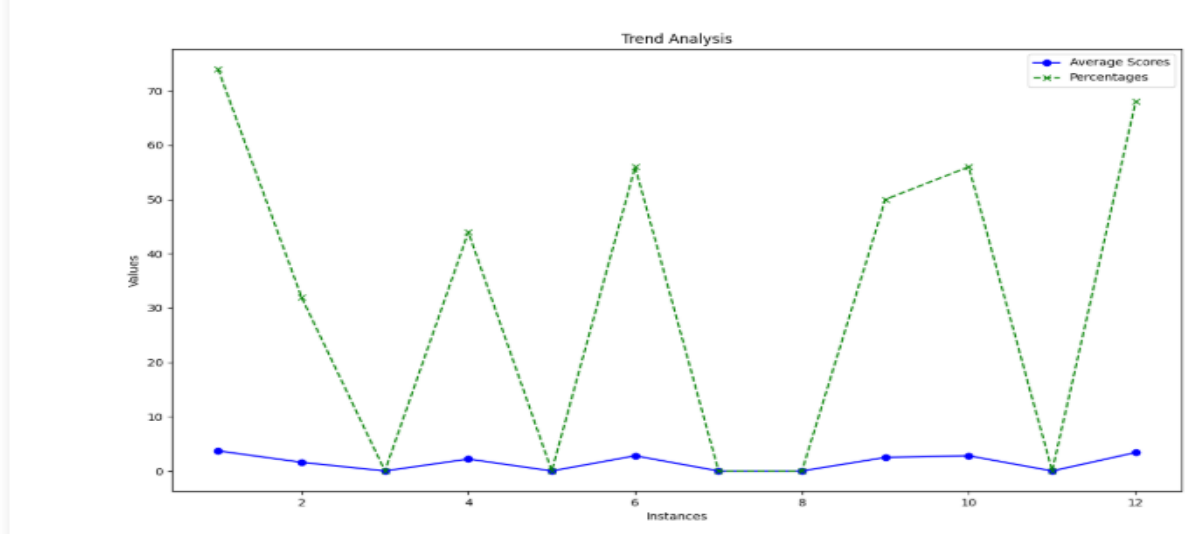


Figure 40: Trend Analysis Historical Evaluation Graph

4.6.19.8 Department Summary

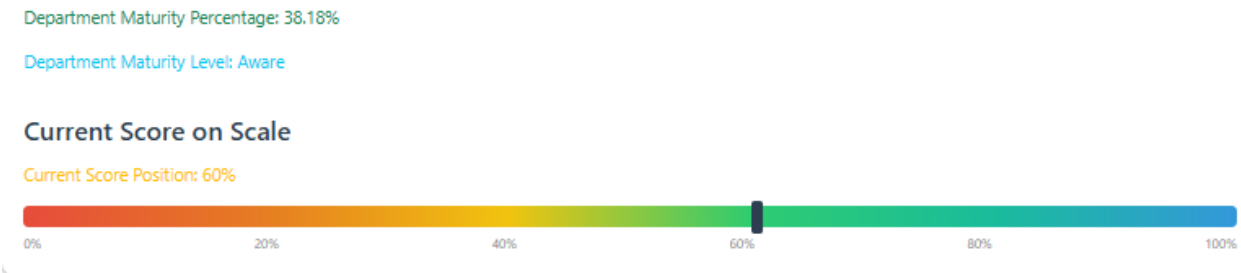


Figure 41: Department summary maturity

4.6.19.9 Department Maturity Comparison

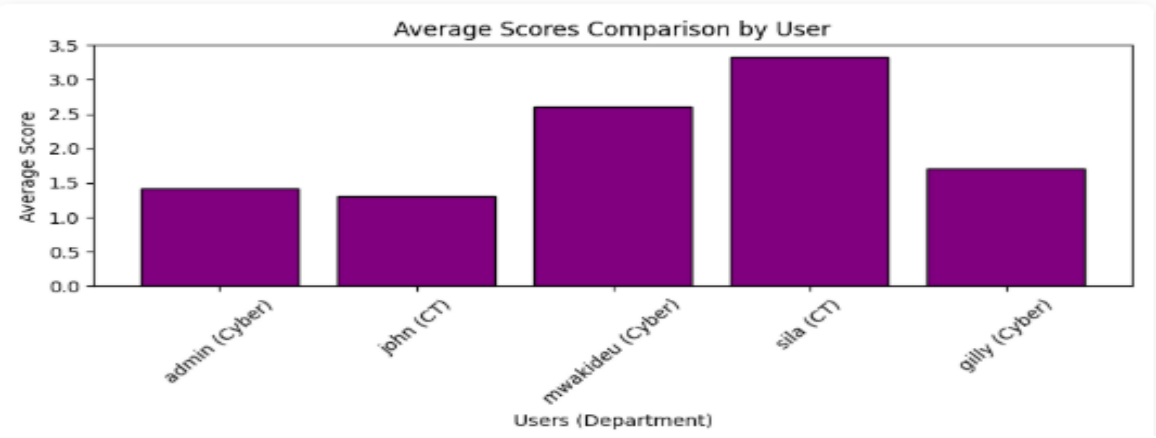


Figure 42: Maturity Comparisons

4.7 Model Validation

To address Study Objective Four, which was to validate the Data Governance Maturity Model at the Kenya Department of Defence, this study employed an empirical validation technique. (UX Glossary, 2025), *Empirical validation involves the act of testing and confirming the efficacy of design decisions through the collection and analysis of observable user data. This covers how users engage with the tool, what activities they complete successfully or unsuccessfully, and how they feel about the experience*. Questionnaires were issued to eleven (11) select Users who hold managerial positions to test the model within the department and give their feedback.

4.7.1 Model Usability

According to (Hubble, 2025), this involves users hands-on interaction with a product, encouraging them to share their thoughts and provide feedback. It is especially effective in identifying major issues related to the product's ease of use, clarity, and overall user experience from the perspective of the intended audience. The study sought to validate users experience when conducting a maturity assessment with the tool. The respondents were asked to rate on a scale of 1-5***, (****Very Difficult****,*** *Difficult****,*** *Neutral/Neither easy nor Difficult****,*** *Easy****,*** *Very Easy)* how easy it was to complete a full data governance assessment using the model.

Table 27: Model usability users' feedback

Rating	Frequency
Very Easy	8
Neutral	1
Easy	2
Difficult	0
VeryDifficult	0

From table 27 above, eight (8) respondents rated the Model usability to be very easy, one (1) rated it Neutral to mean neither easy nor difficult while two (2) rated it easy. This was translated to 72.7% very easy, 18.2% easy and 9.1% Neutral respectively as shown in 43 below. This analysis showed that majority of the select respondents experienced ease of use while interacting with the model. One respondent wasn't sure which suggested potential areas that require further improvements.

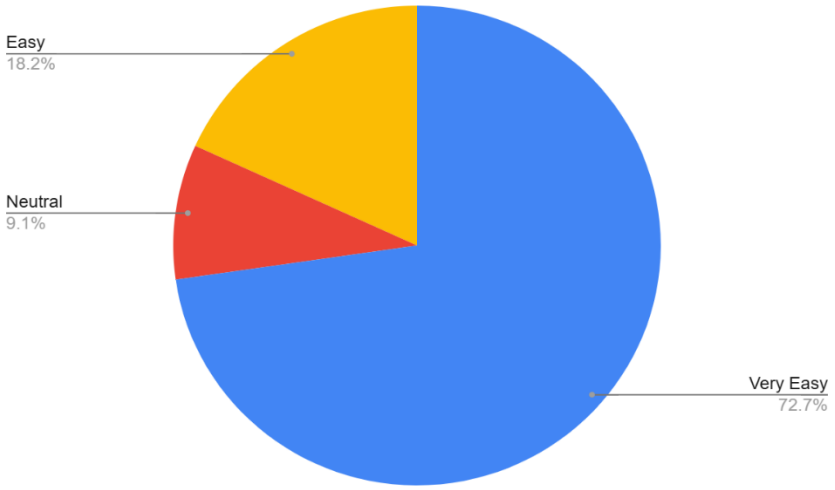


Figure 43: Model Usability

4.7.2 Model Effectiveness

The study sought to determine whether users could finish a basic task such as sign up,sign in and conduct an assessment without outside help. This aimed at validating how intuitive and self-explanatory the model was in real-world applications.

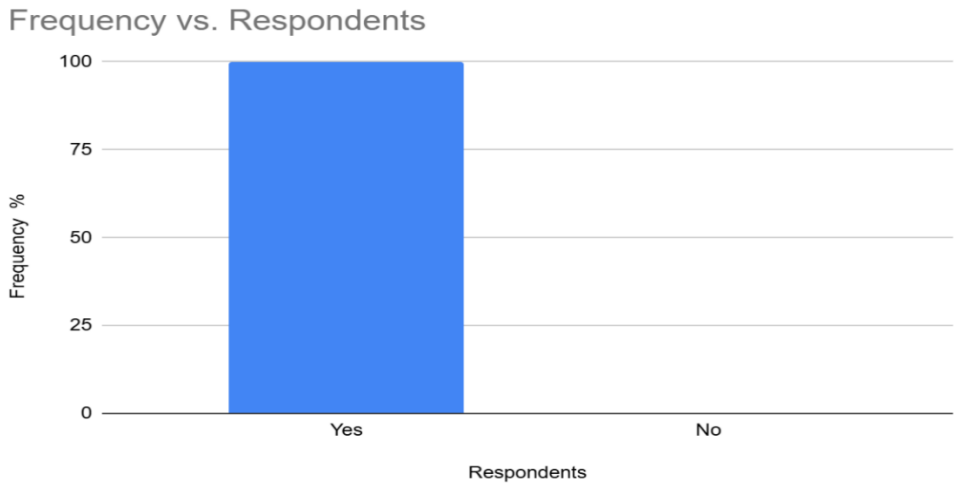


Figure 44: Model Effectiveness

From Figure 44 above, every respondent (100%) was able to use the model to complete simple tasks, such as calculating the data governance maturity score. This confirmed that while performing routine evaluations, end users will find the model to be very easy to use and intuitive.

4.7.3 Models Perceived Value

The model sought to establish which specific features users found most useful or valuable during the assessment process. To achieve this, respondents were asked an open-ended question aimed at identifying components of the model that contributed most to their overall experience and perceived effectiveness. This helped highlight the model's strengths from the users' perspective and informed areas for potential enhancement.

Table 28: perceived value

Feature	Frequency (%)
Analytics	70
UI/UX	10
Calculator	10
Most Parts	10

Analysis on table 28 above confirmed that 70% of the respondents identified the analytics feature as the most valuable aspect of the model, indicating its strong impact on user experience and decision support. Other features such as the user interface, calculator, and overall model structure were also appreciated, though to a lesser extent, each receiving 10% of the responses.

4.7.4 Model Perceived Accuracy

The study sought to validate the model's ability to reflect the Kenya Department of Defence (DOD) data governance. Respondents were asked to rate, on a scale from 0 (Not at all) to 4 (Completely), the extent to which they believe the model accurately captured the maturity level of DODs' data governance practices.

Table 29: Model Accuracy

Rating	Freq %
Not at All	0
Slightly	0
Moderately	0
Very	45.5
Completely	54.5

The analysis above indicated strong confidence in the model's accuracy, with 54.5% of respondents stating it completely reflects the maturity of data governance within the Kenya Department of Defence, and 45.5% indicating it reflects it very well. No respondents rated the model as moderately, slightly, or not at all accurate, suggesting that users overwhelmingly perceive the model as a reliable representation of their data governance maturity.

4.7.5 Model Face Validity

To validate the model's face validity specifically, whether it accurately measured data governance maturity based on users perception and real world applicability, the study asked respondents whether the results generated by the model aligned with their own understanding or experience of their department's data governance status. This Yes/No/Not sure question helped evaluate the congruence between the model's output and user expectations, offering insight into how well the model reflects real-world conditions. One hundred percent (100%) of the respondents indicated that the model's findings matched their perception of the data governance situation in their department. This model's strong face validity suggested that users found it to be accurate and credible.

4.7.6 Models Perceived Limitations

In order to determine the perceived constraints of the model, the participants were asked to provide methods for improving the model's capacity to assess the maturity of data governance. This aided the study in determining possible areas for model enhancement based on actual user experiences and gathering user opinions regarding any restrictions, holes, or inefficiencies in the model's current architecture. The purpose of the collected insights was to improve the model's overall efficacy, usefulness, and relevance in subsequent iterations.

Table 30: Models perceived Limitations

Question	*Feedback*
What suggestions do you have to improve the model's ability to evaluate data governance maturity?	*Looks okay,* *All good,* *Automate more,* *Good Work,* *All good,* *Automation,* *Automate imput,* *Integrate with our audit tools,* *not sure ,* *Good starting point. Can be advanced more.*

Analysis from table 30 above indicated that the model received mostly positive user response, with five (5) respondents expressing pleasure with comments like "**Good work**," "**All good**," and "**Looks okay**." Three (3) respondents expressed a wish to decrease manual processes by highlighting the need for better automation, especially in the area of data input. One (1) user also

suggested integrating the model with already-existing audit tools. Notably, one (1) user was **not sure**, which suggested a potential need for more precise instructions or user support, while another user indicated that the model was a **good place to start** that could be expanded.

4.8 Summary

In conclusion, the study was able to successfully identify necessary Data Governance Key Performance Indicators, derive, implement, and validate the model within the Kenya Department of Defence. Through the identification of relevant performance indicators and the application Principal Component Analysis, the modes was derived. Using rapid prototype development methodology, a functional tool was developed to facilitate practical use, allowing users to assess, interpret, and visualize their maturity levels with ease. User feedback confirmed the model's usability, accuracy, and alignment with real organizational experiences, while also suggesting enhancements such as automation and system integration.

CHAPTER FIVE
5. SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter summarizes the study's main conclusions based on the analysis and findings covered in the preceding chapter. It provides an overview of the main outcomes in relation to the study objectives, highlights the study's contribution to knowledge and practice, and outlines conclusions derived from each objective. The chapter also offers practical recommendations and suggestions for future research, based on insights gained through the model's development, implementation, and validation.

5.2 Summary and Findings

5.2.1 Introduction to Summary

This section covers the main objectives of the study which was to: *Identify* What are the Key Performance Indicators required for measuring data governance implementation within the Kenya Department of Defence, *Derive the Data Governance Maturity Model using the Key Performance Indicators Identified, Implement the Data Governance Model* and lastly *Validate the Model using real-life scenarios at the Kenya Department of Defence*.

5.2.2 Key Performance Indicators necessary for Developing a Data Governance Maturity Model

The first objective of this study was to identify the Key Performance Indicators required for measuring data governance implementation within the Kenya Department of Defence. After a thorough analysis of the literature on data governance, this study identified sixteen (16) Key Performance Indicators (KPIs) as listed in Table 4.3 that are necessary for determining the implementation of a Data Governance program. These KPIs include Data Quality, Data Management, Data Security, Data Privacy, Data Stewardship, Data Architecture, Data Policies & Standards, Metadata Management, Data Accuracy, Strict Enforcement, Data Encryption, Multifactor Authentication, Data Governance Team, Reporting Structure, and AI Analytics. Notably, certain KPIs stood out as more critical for achieving the minimum Data Governance Maturity, specifically Data Security, Data Quality, AI & Analytics, Metadata Management, and Data Accuracy.

5.2.3 Data Governance Maturity Model

The Second objective of the study was to derive the Data Governance Maturity Model The Data Governance maturity mathematical model was developed using linear mathematical approach. To reduce the dimensionality of the data, Principal Component Analysis was conducted which led the study to narrow down to sixteen Key Performance Indicators that proved significant in Data Governance maturity. The weights of the variables (KPIs) were derived using regression analysis while the variable scores were determined using the hybrid maturity model in line with the ISO/IEC 38500-1 principles.

5.2.4 Data Governance Maturity Model Prototype / Tool

The third objective of the study was to implement a functional prototype of the Data Governance Maturity Model. The prototype was developed as a web application to ensure ease of access, cross-platform compatibility, and centralized management for Department of Defence personnel. This section outlines the key components that constitute the design and implementation of the system prototype. It provides an overview of the technology stack, architectural and security considerations, deployment environment, and the core functionalities supported by the system. Presenting these elements in structured tables enhances clarity, supports technical documentation standards, and improves readability for both developers and evaluators.

Table 31: Technology Stack Overview

Component	Technology Used	Justification
Backend Framework	Python Django	Robust, secure, scalable, supports rapid development, and provides built-in admin and ORM.
Frontend	HTML, CSS, JavaScript, Bootstrap	Ensures responsive, professional UI design; supports seamless interaction and cross-device compatibility.
Database	PostgreSQL	Reliable, strong performance, handles complex queries efficiently, and supports ACID compliance.
Visualization Tools	matplotlib, Chart.js	Enables dynamic generation of analytical graphs and visual dashboards for trend analysis.

Table 32: Architecture and Security Measures

Category	Description
System Architecture	Modular three-tier architecture separating Presentation (UI), Business Logic (Django), and Data (PostgreSQL).
User Authentication & Authorization	Secure session-based authentication with enforced user roles.
Password Security	Passwords hashed using bcrypt—a modern, computationally expensive, brute-force-resistant algorithm.
Data Security Controls	Role-based access control, input sanitization, and CSRF protection on all forms to prevent SQL injection and cross-site attacks.

Table 33: Deployment Environment

Component	Technology / Tools	Role in Deployment
Web Server	Gunicorn (WSGI)	Handles Django application execution and HTTP requests.
Reverse Proxy	Nginx	Manages static files, load handling, and routes requests to Gunicorn.
Database Server	PostgreSQL	Hosted on the same secured VPS for data storage and retrieval.
Security Configurations	Firewall, SSL/TLS certificate (HTTPS)	Protects the server and encrypts all data in transit.

Table 34: Core System Functionalities

Functionality	Description
Secure User Registration & Login	Implements protected access through authenticated user accounts.
Interactive Assessment Form	Allows users to rate the 10 core Data Governance KPIs.
Real-time DGMI Calculation	Automatically computes the Data Governance Maturity Index and assigns maturity level.
Dynamic Dashboard	Displays results, historical trend analysis, interactive charts, and departmental comparisons.
Automated Recommendations	Generates tailored insights based on the computed maturity level.
Data Export	Users can download their assessment history in Excel format for extended analysis.

5.2.5 Model Validation

The fourth objective of the study was to validate the Model at the Kenya Department of Defence. To validate the model, actual users within the department conducted real-world Data governance

maturity assessments. During these tests, users interacted with the model where they conducted sample Data Governance Maturity assessments comparing multiple scenarios of Data Governance program. Furthermore, using a questionnaire, the study sought to determine if the users found the Model to be effective in assessing the maturity of a Data Governance program. These tests validated the Models dependability and efficacy in a real-world setting, showcasing its ability to evaluate the maturity of data governance and provide useful insights derived from actual user data. This ensured that the model fulfilled its design goals and performed as anticipated in a real-world operational environment.

5.3 Contribution to Knowledge and Practice.

This section discusses how the research contributes to data governance knowledge and practice.

5.3.1 Contribution to knowledge

This study enriches the academic field by offering a measurable, testable, and generalizable theory of data governance maturity that future researchers can build upon, critique, or extend across sectors and geographies.

5.3.2 Contribution Practice

This study offered a structured approach for assessing and quantifying the maturity of data governance practices within organizations. The Data Governance Maturity Model demonstrated the feasibility of deploying a maturity model within a real-world context and also lays the foundation for broader institutional adoption and continuous improvement in data governance maturity evaluation.

5.4 Conclusion Per objective

5.4.1 Objective 1: *What are the KPIs required to measure data governance implementation within the Kenya Department of Defence?*

The study successfully identified Key Performance Indicators as depicted in Table 4.3 crucial for developing a comprehensive Data Governance Maturity Model for evaluating Data Governance Maturity Levels within the Kenya Department of Defense and any other related Ministries, Departments and Agencies.

5.4.2 Objective 2: *How can the model be derived?*

The current research was able to spell out an effective and statistically justified Data Governance Maturity Model (DGMM). The model was developed in a multi phased method, which combined both the theoretical and empirical methodologies. To find the latent structures and remove

multicollinearity in the identified Key Performance Indicators (KPIs), the dimensionality of these variables were reduced applying the Principal Component Analysis (PCA) as the first step of the analysis of the collected data. This analysis step allowed grouping of similar indicators into different governance dimensions as People, Processes, Technology, Organizational and Emerging factors.

They were then able to calculate the composite scores of each dimension using the factor loadings and the mechanisms of assigning relative importance to the factors were made using regression. The quantitative method used in this approach was the assurance that the model would not just record the presence of governance elements but the impact, they had on the overall maturity, and the differing magnitude.

The model that comes up is scalable, adaptable, and replicable, which makes it suitable to apply in the Kenya Department of Defence and other ministries, departments, and agencies (MDAs). Above all, the model offers an orderly and evidence-based system to evaluate, track and improve data governance procedures in complex organisational settings. This was a major breakthrough in the conceptualization of data governance theory into a quantifiable system which can not only be used in the development of strategic plans but also throughout the monitoring of licensed performances within public sector establishments.

5.4.3 Objective 3: How can the DGMM Model tool be implemented?

The Data Governance Maturity Model prototype was developed using an iterative rapid-prototyping approach.. Tkinter was utilized as the graphical user interface and it was written in Python as the primary programming language along with Postgres as a backend data storage and retrieval mechanism. This way of building the model allowed the project to receive continuous feedback that was user-friendly, thus maintaining stability with the functionality as well as the operational needs.

The formulated instrument allows entering self-assessment information according to a defined dimensional framework, and this produces data governance maturity index. This index is then graphically represented and reported in forms that are manageable and in forms that are implementable. Through this, the prototype will provide an interactive and reachable form of data

governance practices monitoring, assessment, and improvement procedures of Kenya Department of Defence.

5.4.4 Objective 4: How can the model be validated?

The proposed Data Governance Maturity Model (DGMM) was validated by way of testing it in reality wherein the real users, that is, the data management staff, will use the prototype system to make sample data governance maturity checks. These findings show the reliability, usefulness, and feasibility of the model. Notably, the validation has indicated that the DGMM has performed rather as predicted in a controlled environment but also given the same results when implemented in real life settings. Additionally, the exercise also established that the DGMM has met its entitlement in the field of design and that it can be used as a dependable tool in the evaluation and direction of data governance maturity within the departments of the Kenyan department of defence as well as other similar entities.

5.4.5 Summary

Driven by the lack of a strictly empirical basis to quantify the maturity of data governance environment within Kenya Department of Defense, the authors have come up with the Data Governance Maturity Model (DGMM) in order to address this research methodology gap. The instrument produced gave the Defense sector an opportunity to have a systematic way of measuring organizational maturity in data governance and as well the instrument gave the system a way of making specific, strategic interventions to be made to strengthen the institutions data governance program. Besides clearly demonstrating value in regard to the Kenya Defense establishment, the DGMM provides a transferable example that other government ministries, departments, and agencies (MDAs) can utilize to evaluate and optimise their data control.

5.5 Conclusions

5.5.1 Conclusion on Objective One: Identification of Key Performance Indicators (KPIs)

The study successfully identified and validated a set of sixteen Key Performance Indicators (KPIs) critical for measuring data governance maturity within the Kenya Department of Defence. Through a synthesis of global literature and empirical validation within the defence context, it was established that while universal KPIs like Data Quality and Data Security are foundational, their implementation is uniquely characterized by defence-specific imperatives such as Strict Enforcement, AI for situational awareness, and interoperability within a joint-force command structure. The Principal Component Analysis (PCA) further refined this set, confirming that a core

group of ten KPIs—including Data Policies, Data Encryption, and the Data Governance Team—statistically encapsulate the essential dimensions of maturity in this environment.

5.5.2 Conclusion on Objective Two: Derivation of the Data Governance Maturity Model (DGMM)

A robust and quantitatively derived Data Governance Maturity Model (DGMM) was successfully developed. The model is not a subjective checklist but is grounded in empirical data. The application of PCA provided a statistically sound method for dimensionality reduction, ensuring the model's construct validity. The resultant mathematical formula, the Data Governance Maturity Index (DGMI), offers a transparent and replicable method for computing a precise maturity score. The accompanying six-level Hybrid Maturity Model, which integrates the structure of CMMI with defence-specific relevance, provides a clear and staged roadmap for progressive improvement.

5.5.3 Conclusion on Objective Three: Implementation of the DGMM Prototype

The DGMM was successfully operationalized through a functional web-based prototype. Developed using the Rapid Application Development methodology, the prototype effectively translates the theoretical model into a practical tool. It enables users to conduct assessments, automatically calculates the DGMI, generates tailored recommendations, and visualizes historical trends. User acceptance testing confirmed the prototype's high usability, intuitiveness, and practical value as a decision-support tool for DOD personnel.

5.5.4 Conclusion on Objective Four: Validation of the Model

The DGMM and its prototype were rigorously validated within the Kenya DOD, demonstrating both practical utility and scientific rigor. The validation process confirmed high face validity, with users affirming the model's accuracy in reflecting their operational reality. Furthermore, quantitative measures, including an excellent System Usability Scale (SUS) score and a 100% task success rate, provide strong evidence of the artifact's effectiveness, usability, and relevance in solving the identified problem.

5.6 Contribution to Knowledge and Practice

This study makes a tripartite contribution. To knowledge, it offers one of the first empirically-grounded data governance maturity models tailored for a defence context, addressing a significant gap in the literature. Methodologically, it demonstrates the rigorous application of DSR and PCA in developing a maturity model, moving beyond qualitative checklists. In practice, it provides the Kenya DOD with a validated, ready-to-use tool to objectively benchmark its data governance

maturity, identify gaps, and guide strategic investment, thereby enhancing data-driven decision-making and national security.

5.7 Critical Reflection on Limitations

While the study achieved its objectives, its limitations must be acknowledged. Firstly, the generalizability of the findings is context-bound. The model was specifically designed and validated for the Kenya DOD; its direct application to other sectors or even other national defence forces may require adaptation to account for different regulatory environments and operational cultures. Secondly, the validation, while robust, was conducted with a specific, expert user group within a single organization. Broader validation across a more diverse set of stakeholders within the DOD could further strengthen the model's credibility. Finally, the model's input mechanism relies on user self-assessment, which carries an inherent risk of perception bias. Future iterations could be strengthened by integrating objective, system-generated data to supplement subjective ratings.

5.8 Prioritized and Actionable Recommendations

5.8.1 Immediate Recommendations for the Kenya Department of Defence

Institutionalize the DGMM: Integrate the DGMM prototype into the annual strategic planning and audit cycles of the DOD. Mandate its use across key directorates to establish a standardized baseline and track progress annually.

Formalize Data Governance Roles: Use the model's findings to formally constitute and empower the Data Governance Team (DGT) and define clear data stewardship roles within the existing chain of command, as highlighted by the model's emphasis on organizational factors.

5.8.2 Strategic Recommendations for Future Research

Enhance Model Objectivity (High Priority): Future research should focus on integrating the DGMM with existing DOD systems (e.g., audit logs, SIEM tools) to automate the collection of evidence for certain KPI scores, thereby reducing reliance on self-reporting and increasing assessment objectivity.

Conduct Longitudinal Studies: Apply the DGMM over a 3-5 year period within the DOD to study maturity progression, identify common improvement pathways, and validate the long-term impact of the model on operational effectiveness.

Explore Generalizability: Investigate the adaptation and application of the DGMM framework to other Kenyan government ministries and agencies (MDAs), particularly those in the national security cluster, to foster a whole-of-government approach to data governance.

REFERENCES

Abdi, H., & Williams, L. J. (2010, July). Principal component analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data Governance: A conceptual framework, structured review, and research agenda. *International journal of information management*, *49*, 424-438.

Aiello, M., Esposito, G., Pagliari, G., Borrelli, P., Brancato, V., & Salvatore, M. (2021). How does DICOM support big data management? Investigating its use in medical imaging community. *Insights into Imaging*, *12*(1), 1-21.

Aiken, P. (2016). Experience: Succeeding at data management—BigCo attempts to leverage data. *Journal of Data and Information Quality (JDIQ)*, *7*(1-2), 1-35.

Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring big Data Governance frameworks. *Procedia Computer Science*, *141*, 271-277.

Al-Dossari, H., & Sumaili, A. A. AData Governance MATURITY ASSESSMENT: A CASE STUDY OF SAUDI ARABIA.

Al-Dossari, H., & Sumaili, A. A. AData Governance MATURITY ASSESSMENT: A CASE STUDY OF SAUDI ARABIA.

Alhassan, I., Sammon, D., & Daly, M. (2016). Data Governance activities: an analysis of the literature. *Journal of Decision Systems*, *25*(sup1), 64-75.

Alhassan, I., Sammon, D., & Daly, M. (2019). Critical success factors for Data Governance: a theory building approach. *Information Systems Management*, *36*(2), 98-110.

Alhassan, I., Sammon, D., & Daly, M. (2019, April 3). Critical Success Factors for Data Governance: A Theory Building Approach. *Information Systems Management*, *36*(2), 98–110. https://doi.org/10.1080/10580530.2019.1589670

Al-Ruithe, M., & Benkhelifa, E. (2017). *Cloud Data Governance Maturity Model*. Paper presented at the Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing.

Biggs, R., De Vos, A., Preiser, R., Clements, H., Maciejewski, K., & Schlüter, M. (2021, July 29). *The Routledge Handbook of Research Methods for Social-Ecological Systems*. Routledge.

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018, January 1). *A Review on Sample Size Determination for Cronbach's Alpha Test: A Simple Guide for Researchers*. The Malaysian Journal of Medical Science; University of Science Malaysia. https://doi.org/10.21315/mjms2018.25.6.9

Burcu, A. (2000). A comparison of two data collecting methods: interviews and questionnaires. *Hacettepe Univ J Educ, 18*, 1-10.

Ceruti, M. G. (2003). Data management challenges and development for military information systems. *IEEE Transactions on Knowledge and Data Engineering*, *15*(5), 1059-1068.

Chan, L. L., & Idris, N. (2017, October 31). Validity and Reliability of The Instrument Using Exploratory Principal Component Analysis and Cronbachâ™s alpha. *International Journal of Academic Research in Business and Social Sciences*, *7*(10). https://doi.org/10.6007/ijarbss/v7-i10/3387

CMMI. (n.d.). https://www.umsl.edu/~sauterv/analysis/6840papers_f12/Powell/index.html

Cupoli, P., Earley, S., & Henderson, D. (2014). Dama-dmbok2 framework. *Dama International*.
Cupoli, P., Earley, S., & Henderson, D. (2014). Dama-dmbok2 framework. *Dama International*.

Data Protection Act - OFFICE OF THE DATA PROTECTION COMMISSIONER KENYA. (2021, November 8). OFFICE OF THE DATA PROTECTION COMMISSIONER KENYA - ODPC Kenya. https://www.odpc.go.ke/dpa-act/

David Patón-Romero, J., Baldassarre, M. T., Rodríguez, M., & Piattini, M. (2019, December). Maturity model based on CMMI for governance and management of Green IT. *IET Software*, *13*(6), 555–563. https://doi.org/10.1049/iet-sen.2018.5351

Dirner, M., Yuan, E., Blalock, J., & VA, C. I. O. A. (2007). Realizing the Army Net-Centric Data Strategy (ANCDS) in a Service Oriented Architecture (SOA). *Crosstalk*, *7*.

DoD CIO(IM), O. N. US ARMY NET-CENTRIC STARATEGY.

Duncan, A. D. (2021). Over 100 Data and Analytics Predictions through 2025.

Ekundayo, T., Bhaumik, A., Chinoperekweyi, J., & Khan, Z. (2023, July 17). The Impact of Open Data Implementation on Entrepreneurship Ability in Sub-Saharan Africa. *Human Behavior and Emerging Technologies*, *2023*, 1–11. https://doi.org/10.1155/2023/7583550

Harland, P. E., & Uddin, Z. (2014). Effects of product platform development: fostering lean product development and production. *International Journal of Product Development*, *19*(5/6), 259. https://doi.org/10.1504/ijpd.2014.064881

Herselman, M., Wayi, N., & Olaitan, O. (2019). A Data Governance Maturity Evaluation Model for government departments of the Eastern Cape province, South Africa. *South African Journal of Information Management*, *21*(1), 1-12.

Hofmann, S., Müller, O., & Rossi, M. (2020, December 1). *Designing for Digital Transformation. Co-Creating Services with Citizens and Industry*. Springer Nature. http://books.google.ie/books?id=pR8MEAAAQBAJ&pg=PA407&dq=Wolff,+Clemens+%26+K%C3%BChl,+Niklas+%26+Satzger,+Gerhard.+(2020).+Engineering+Industrial+Service+Systems:+Design+and+Evaluation+of+System-Oriented+Service+Delivery.+10.1007/978-3-030-64823-7_39.&hl=&cd=1&source=gbs_api

Johnsen, F. T., Zieliński, Z., Wrona, K., Suri, N., Fuchs, C., Pradhan, M., . . . Dyk, M. (2018). *Application of IoT in military operations in a smart city*. Paper presented at the 2018 International Conference on Military Communications and Information Systems (ICMCIS).

Kaushik, V., & Walsh, C. A. (2019, September 6). *Pragmatism as a Research Paradigm and Its Implications for Social Work Research*. Social Sciences. https://doi.org/10.3390/socsci8090255

Kevins, J., & Brian, K. (2022). Defining Data Protection in Kenya: Challenges, Perspectives and Opportunities. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4270712

Lauff, C. A., Kotys-Schwartz, D., & Rentschler, M. E. (2018, March 23). What is a Prototype? What are the Roles of Prototypes in Companies? *Journal of Mechanical Design*, *140*(6). https://doi.org/10.1115/1.4039340

Laybats, C., & Davies, J. (2018). GDPR: Implementing the regulations. *Business Information Review, 35*(2), 81-83.

MacFeely, S., Me, A., Fu, H., Veerappan, M., Hereward, M., Passarelli, D., & Schüür, F. (2022). Towards an international Data Governance framework. *Statistical Journal of the IAOS, 38*(3), 703-710.

MoALFC. (2022). MoALFC-Data-Governance-Framework-2022.

Moffat, J. (2010, January 1). *Complexity Theory and Network Centric Warfare*. DIANE Publishing.

Nadal, S., Jovanovic, P., Bilalli, B., & Romero, O. (2022). Operationalizing and automating Data Governance. *Journal of big data, 9*(1), 1-31.

Orfanus, D., De Freitas, E. P., & Eliassen, F. (2016). Self-organization as a supporting paradigm for military UAV relay networks. *IEEE Communications letters, 20*(4), 804-807.

PANKOWSKA, M. (2020). Big Data Governance. *WHO RUNS THE WORLD: DATA*, 93.

PANKOWSKA, M. (2020). Big Data Governance. *WHO RUNS THE WORLD: DATA*, 93.

Permana, R. I., & Suroso, J. S. (2018). *Data Governance maturity assessment at PT. XYZ. Case study: data management division.* Paper presented at the 2018 International Conference on Information Management and Technology (ICIMTech).

Permana, R. I., & Suroso, J. S. (2018). *Data Governance maturity assessment at PT. XYZ. Case study: data management division.* Paper presented at the 2018 International Conference on Information Management and Technology (ICIMTech).

Prasetyo, H. N. (2016). A review of Data Governance maturity level in higher education. *Jurnal Ilmiah Teknologi Infomasi Terapan, 3*(1).

Redmond, S., Wilcox, S. L., Campbell, S., Kim, A., Finney, K., Barr, K., & Hassan, A. M. (2015). A brief introduction to the military workplace culture. *Work, 50*(1), 9-20.

Rivera, S., Loarte, N., Raymundo, C., & Domínguez-Mateos, F. (2017). *Data Governance Maturity Model for Micro Financial Organizations in Peru.* Paper presented at the ICEIS (3).

Rosenthal, M. (2016). Qualitative research methods: Why, when, and how to conduct interviews and focus groups in pharmacy research. *Currents in pharmacy teaching and learning, 8*(4), 509-516.

Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior, 51*, 1068-1076.

Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3205040

Tsang, S., Royse, C., & Terkawi, A. S. (2017, January 1). *Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine*. Saudi Journal of Anaesthesia; Medknow. https://doi.org/10.4103/sja.sja_203_17

t-Test, Chi-Square, ANOVA, Regression, Correlation. . . (n.d.). https://datatab.net/statistics-calculator/reliability-analysis/cronbachs-alpha-calculator?example=Cronbachs_Alpha

Venable, J., Pries-Heje, J., & Baskerville, R. (2016, January). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to design science research. *Design science research. Cases*, 1-13.

Voss, W. G. (2019). Cross-border data flows, the GDPR, and Data Governance. *Wash. Int'l LJ, 29*, 485.

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), 1*(1), 1-27.

Wende, K., & Otto, B. (2007). A contingency approach to Data Governance.

Were, V., & Moturi, C. (2017). Toward a Data Governance model for the Kenya health professional regulatory authorities. *The TQM Journal, 29*(4), 579-589.

Were, V., & Moturi, C. (2017). Toward a Data Governance, model for the Kenya health professional regulatory authorities. *The TQM Journal, 29*(4), 579-589.

Yebenes, J., & Zorrilla, M. (2019). Towards a Data Governance framework for third generation platforms. *Procedia Computer Science, 151*, 614-621.

Yebenes, J., & Zorrilla, M. (2019). Towards a Data Governance framework for third generation platforms. *Procedia Computer Science, 151*, 614-621.

Zhang, N., & Yuan, Q. (2016). An overview of Data Governance. *Economics Paper*, *December*.

Zúñiga, D. V., Cruz, R. K., Ibañez, C. R., Dominguez, F., & Moguerza, J. M. (2018). *Master data management maturity model for the microfinance sector in Peru.* Paper presented at the Proceedings of the 2nd international conference on information system and data mining.

Mwai, J. K., Makokha, J. W., & Karume, S. M. (2020). A framework for integration of Web 3.0 and social media technologies in government portals for personalized integrated service delivery. *International Journal of Computer Trends and Technology*, *68*(5), 66–73. https://doi.org/10.14445/22312803/ijtt-v68i5p114

Fincham, J. E. (2008). Response rates and responsiveness for surveys, standards, and the Journal. *American Journal of Pharmaceutical Education*, *72*(2), 43. https://doi.org/10.5688/aj720243

UX Glossary. (2025, May 7). *Empirical Validity / Empirical Validation Meaning & Definition of Terms | UX Glossary*. https://www.uxglossary.com/terms/empirical-validity-empirical-validation/

Hubble. (2025, May 6). A comprehensive guide to usability testing for beginners. *Hubble*. https://www.usehubble.io/blog/usability-testing-guide

APPENDICIES

a) QUESTIONNAIRE

Dear Respondent

I am a Master of Science in Cyber Security student at the Cooperative University of Kenya, and I am working on a research project titled "A Data Governance Maturity Model: A Case of Kenya DOD." This is to sincerely request that you answer the questions listed below as honestly as possible. Please keep in mind that the information you supply will only be used for this academic research and will be kept anonymous according to stringent ethical standards.

Thank you for agreeing to participate in this research.

Data Governance: *A set of practices, processes, policies, guidelines, and standards that an organization implements to manage their data assets efficiently and responsibly*

gillygathogo2496@gmail.com Switch account

Not shared

* Indicates required question

1. **What position do you currently hold at your current working station?**

_____

2. **How long have you worked in your current position ?**

_____

3. **SECTION 2.** *
1. Are you aware of policies, standards governing DOD data throughout its lifecycle, from creation and acquisition to usage, storage, and disposal.

Check all that apply.

- [ ] Yes
- [ ] No
- [ ] I Don't Know

4. 2. If you answered YES to question 1 above . Do think the policies and standards being implemented are able to ensure that data is managed effectively throughout its lifecycle, from creation and acquisition to usage, storage, and disposal within the DOD? *

Check all that apply.

- [ ] Yes
- [ ] No
- [ ] I Don't Know

5. 3. Are there Clear reporting structure for DOD data governance issues to ensure accountability and effective Data Governance? *

Check all that apply.

- [ ] Yes
- [ ] No
- [ ] I Don't Know

6. 4. Based on your experience in handling data within your department at the DoD, What is your overall satisfaction with the adherence to the of key aspects of data governance. Using a scale of 1 to 5 (where 1 is very dissatisfied and 5 is very satisfied), please rate how the following aspects of data governance have been adhered to—from creation and acquisition to usage, storage, and disposal - *

a. **Data Quality**: Ensuring that data is accurate, consistent, and reliable for decision-making and analysis.

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

7. *

b. **Data Management**: Establishing processes for data storage, retrieval, and maintenance to meet organizational needs and regulatory requirements.

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

8. *

c. **Data Security**: Implementing measures to protect data from unauthorized access, alteration, or loss.

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

Activ

9. *

d. **Data Privacy**: Ensuring that data is collected, processed, and used in compliance with privacy laws and regulations, and that individuals' privacy rights are respected.

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

10. *

e. **Data Stewardship**: Assigning responsibility for data management to individuals or teams within the organization, who are accountable for its quality, integrity, and security.

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

11. *

f. **Data Architecture**: Designing and implementing data structures and systems to support the organization's data management objectives.

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

12. g. **Data Governance Policies and Standards**: Establishing guidelines and best practices for data management and usage, which are enforced throughout the organization. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

13. h.**Metadata Management**: Capturing and managing metadata (data about data) to facilitate data discovery, understanding, and governance. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

Indicate to what extent you agree with the following statements about Data Governance Operations within the DOD

14. 5. Data accuracy is vital for the reliability and effectiveness of DoD data governance operations. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

15. 6. Strict enforcement of data access control policies would enhance CIA (Confidentiality, Integrity,Authenticity) of sensitive DoD data. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

16. 7.Using data security mechanisms such as use of passwords, & encryption in sharing documents between departments would and improving overall DOD data governance. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

17. 8. Multi-factor authentication for accessing classified DOD data systems would significantly enhances security of critical Data. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

18. 9. A dedicated Data Governance team to oversee implementation of data management policies would be key in enhancing data governance within the DoD. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

19. 10.Having reporting structure for DOD data governance issues would ensures accountability and effective Data Governance? *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

20. 11. AI-driven data analytics and automation in data governance processes is essential for keeping pace with evolving data management practices & overall improvement of Data governance within the DoD. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

21. 12. Data Governance Officer/office role is essential for overseeing data management policies, ensuring compliance with regulations, and driving data governance initiatives within the DoD. To what extent do you agree that this Officer/office would contribute to overall success governance data within the DOD. *

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

22. 13. Policies aligned to International Data governance standards are essential in overall success of DOD Data governance. *

Check all that apply.

- ☐ YES
- ☐ No

23. 14. List any other factors not listed/mentioned above that you think if implemented would improve data governance initiatives within the DoD *

b). INFORMED CONSENT TO PARTICIPATE IN A RESEARCH STUDY

This Informed Consent Form is for study participants who we are inviting to participate in research entitled **'Data Governance Maturity Model: A Case of The Kenya Department of Defence'**

1. Gilly G Gathogo
   Gillygathogo2496@gmail.com

   0726530060
2. Prof. Simon Karume
   skarume@cuk.ac.ke

   0722499397
3. Dr Josphat Karani
   jkarani@kyu.ac.ke
   0724 323 148

You will receive a copy of the entire Informed Consent Form.

This Informed Consent Form Contains Two Parts:

1. Information Sheet (to share research details with you).
2. Certificate of Consent (need signatures if you accept to participate).

You will be given a copy of the full Informed Consent Form

This Informed Consent Form Has Two Parts:

1. Information Sheet (to share information about the research with you).
2. Certificate of Consent (for signatures if you agree to take part).

PART I: INFORMATION SHEET

Introduction

My name is Gilly Gathogo, and I am currently pursuing a Master's degree in Cyber Security at the Cooperative University of Kenya. My research focuses on determining the level of data governance maturity within Kenya's Department of Defence. Your participation in this study is intended to help outline potential functional and non-functional needs for the proposed model. Your selection as a possible participant is based on your Data Management experience. Before deciding whether or not to participate in this study, please consult with anybody you trust.

Purpose of the research

The purpose of my research is to develop a comprehensive Data Governance Maturity Model specifically tailored to the context of the Kenya Department of Defence. This model aims to assess and enhance the organization's data governance practices by identifying both functional and non-functional requirements. Through this research, we seek to establish a framework that will enable

effective data management, security, and compliance within the department, ultimately contributing to the overall maturity and resilience of its data governance processes.

Type of Research Intervention

This research will use questionnaires and guided interviews to determine the projected model's functional and non-functional requirements.

Participant selection

We are inviting individuals with knowledge in data management to participate in the study.

Voluntary Participation

Your participation in this research is entirely voluntary. You can choose whether or not to participate.

Duration

This research is projected to take months. During that period, you will contribute to defining the probable functional and non-functional requirements at the start of the research and also participate in model validation at the end of the study.

Benefits

Your participation is likely to help us find an answer to the study question(s) as well as benefiting future generations in advancing this study.

Confidentiality

The information gathered throughout this study endeavor will be kept strictly secret. Information about you obtained throughout the research will be kept confidential and only the researchers will have access to it. It will not be shared or supplied to anybody other than Professor Simon Karume and Dr. Josphat Karani, who are members of this study team.

Sharing the Results

The knowledge that will be gained from conducting this research will be shared with you at our last meeting before it is made freely available to the public. No confidential information will be exposed. Following the final conference, the findings will be published so that others can benefit from our research.

Right to Refuse or Withdraw

You are not required to participate in this research if you do not choose to do so. You have the option to withdraw from the study at any time. It is your choice, and all of your rights will be upheld.

Who to Contact

If you have any questions, please ask them now or later, even after the study has begun. If you have any questions later, please contact any of the following:

1. Prof. Simon Karume
   skarume@cuk.ac.ke

0722499397

2. Dr Josphat Karani
   jkarani@kyu.ac.ke

PART II: CERTIFICATE OF CONSENT

I've read the information above. I had the opportunity to ask questions regarding it, and all of them were addressed adequately. I willingly agree to take part in this investigation as a participant.

Name _____ Signature _____Date _____

Day/month/year

I witnessed a proper reading of the permission document to the possible participant, and the individual was given the opportunity to ask questions. I confirm that the individual willingly gave their consent.

Name _____ Signature _____Date _____

Day/month/year

Statement by the researcher

I have accurately read the information sheet to the possible participant and, to the best of my abilities, ensured that the following will be done.

1. Participation in this study is entirely voluntary.

2. All information acquired throughout this survey will be kept strictly secret.

3. At the last conference, researchers will discuss their findings before making them public.

I confirm that the participant had the opportunity to ask questions concerning the study, and that all of the questions were answered appropriately and to the best of my ability. I confirm that the individual was not pressured into providing consent, and that the consent was given freely and voluntarily.

A copy of this ICF has been given to the participant.

Name _____Signature _____Date _____

Day/month/year

c). APPROVAL OF RESEARCH SUPERVIROS

THE CO-OPERATIVE UNIVERSITY OF KENYA
P.O. BOX 24814-00502. KAREN. NAIROBI KENYA.
TELEPHONE: (020)2430127. 2679456.8891401. FAX (020)-8891410.
BOARD OF POSTGRADUATE STUDIES

3rd October, 2024

To: Mr. Gilly Gathogo
Professor Simon Karume
Dr. Josphat Karani

Dear Sirs,

RE: APPROVAL OF RESEARCH SUPERVISORS

Following the receipt of post graduate students research supervisors recommendation lists from the respective school postgraduate committees, subsequent deliberations and approval by the University Board of Postgraduate Studies in its meeting held on 8th August, 2024, I am happy to inform you that **Mr. Gilly Gathogo** of registration number **MCSC01/6012/2022** has been assigned the faculty members listed hereunder to supervise him in his master's research thesis/project:

Order	Name	Email	Phone Number
First Supervisor	Professor Simon Karume	skarume@cuk.ac.ke	0722499397
Second Supervisor	Dr. Josphat Karani	jkaraniw@gmail.com	0724323148

The student is advised to contact his supervisors immediately in order to embark on the research work. I wish all of you the best in this rigorous academic endeavour.

D. K. Muthoni

Director – BPS

CC: DVC- ACDRI, Registrar – ACDRI, Dean – SCM, CoD – DCSIT

CUK is ISO 9001:2015 CERTIFIED

d). AUTHORITY TO CONDUCT RESEARCH

MINISTRY OF DEFENCE
INTER OFFICE MEMO

From : HOD/ICT

TO : All HODs

Ref No. MOD/12/14/1A

Date: 6th October, 2024

RE: AUTHORITY TO CONDUCT ACADEMIC RESEARCH
CAPT GILLY GITAHI GATHOGO – MCS01/6012/2022

Authority is hereby granted to the above-name to carry out research on Data Governance Maturity for a period of one (1) month from the date of this letter. The officer is currently undertaking Masters Programme in Cyber Security at the Co-operative University of Kenya.

Kindly accord him the necessary assistance.

Kenneth Radull
HEAD OF ICT

e). NACOSCI LICENSE

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION

Ref No: ٧٩٣٥٨٥ Date of Issue: ٢٠ /May /٢٠٢٥

RESEARCH LICENSE

This is to Certify that Mr.. Gilly Gitahi Gathogo of The Cooperative University of Kenya. has been licensed to conduct research as per the provision of the Science. Technology and Innovation Act. ٢٠١٣ (Rev. ٢٠١٤) in Nairobi on the topic: Data Governance Maturity Model- A Case Study for The Kenya Ministry of Defence for the period ending : ١٠ /May /٢٠٢٦.

License No: NACOSTI /P /٢٥ /٤١٧٢٨٥١

٧٩٣٥٨٥

Applicant Identification Number

Deputy Director
NATIONAL COMMISSION FOR SCIENCE. TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document Scan the QR Code using QR scanner application.

See overleaf for conditions

THE SCIENCE, TECHNOLOGY AND INNOVATION ACT, ٢٠١٣ (Rev. ٢٠١٤)
Legal Notice No. ١٠٨: The Science, Technology and Innovation (Research Licensing) Regulations, ٢٠١٤

The National Commission for Science, Technology and Innovation, hereafter referred to as the Commission, was the established under
Science, Technology and Innovation Act ٢٠١٣ (Revised ٢٠١٤) herein after referred to as the Act. The objective of the Commission shall be to
regulate and assure quality in the science, technology and innovation sector and advise the Government in matters related thereto.

CONDITIONS OF THE RESEARCH LICENSE

١. The License is granted subject to provisions of the Constitution of Kenya, the Science, Technology and Innovation Act, and other relevant laws, policies and regulations. Accordingly, the licensee shall adhere to such procedures, standards, code of ethics and guidelines as may be prescribed by regulations made under the Act, or prescribed by provisions of International treaties of which is a signatory to.
٢. The research and its related activities as well as outcomes shall be beneficial to the country and shall not in any way;
   i. Endanger national security
   ii. Adversely affect the lives of Kenyans
   iii. Be in contravention of Kenya's international obligations including Biological Weapons Convention (BWC), Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), Chemical, Biological, Radiological and Nuclear (CBRN).
   iv. Result in exploitation of intellectual property rights of communities in Kenya
   v. Adversely affect the environment
   vi. Adversely affect the rights of communities
   vii. Endanger public safety and national cohesion
   viii. Plagiarize someone else's work
٣. The License is valid for the proposed research, location and specified period.
٤. Neither the license nor any rights thereunder are transferable.
٥. The Commission reserves the right to cancel the research at any time during the research period if in the opinion of the Commissio research is not implemented in conformity with the provisions of the Act or any other written law.
٦. The Licensee shall inform the relevant County Director of Education, County Commissioner and County Governor before commencement of the research.
٧. Excavation, filming, movement, and collection of specimens are subject to further necessary clearance from relevant Governmer Agencies.
٨. The License does not give authority to transfer research materials.
٩. The Commission may monitor and evaluate the licensed research project for the purpose of assessing and evaluating compliance the conditions of the License.
١٠. The Licensee shall submit one hard copy, and upload a soft copy of their final report (thesis) onto a platform designated by the Commission within one year of completion of the research.
١١. The Commission reserves the right to modify the conditions of the License including cancellation without prior notice.
١٢. Research, findings and information regarding research systems shall be stored or disseminated, utilized or applied in such a mani may be prescribed by the Commission from time to time.
١٣. The Licensee shall disclose to the Commission, the relevant Institutional Scientific and Ethical Review Committee, and the relevant national agencies any inventions and discoveries that are of National strategic importance.
١٤. The Commission shall have powers to acquire from any person the right in, or to, any scientific innovation, invention or patent of strategic importance to the country.
١٥. Relevant Institutional Scientific and Ethical Review Committee shall monitor and evaluate the research periodically, and make a re of its findings to the Commission for necessary action.

National Commission for Science, Technology and
Innovation (NACOSTI),
Off Waiyaki Way, Upper Kabete,
P. O. Box ٣٠٦٢٣ - ٠٠١٠٠ Nairobi, KENYA
Telephone: ٠٢٠٤٠٠٧٠٠٠, ٠٧١٣٧٨٨٧٨٧, ٠٧٣٥٤٠٤٢٤٥
E-mail: dg@nacosti.go.ke
Website: www.nacosti.go.ke

f). SYSTEMS CODE SNIPPET

```
import tkinter as tk
from tkinter import messagebox, ttk, filedialog
import matplotlib.pyplot as plt
from matplotlib.backends.backend_tkagg import FigureCanvasTkAgg
import pandas as pd
import json
import os

# Global variables
users = {"admin": "password123"}
current_user = None
historical_data = {}

# Load and save historical data
def load_historical_data():
    global historical_data
    if os.path.exists("historical_data.json"):
        with open("historical_data.json", "r") as file:
            historical_data = json.load(file)

def save_historical_data():
    with open("historical_data.json", "w") as file:
        json.dump(historical_data, file)

# Save history to an Excel file
def save_history_to_excel():
    if current_user in historical_data:
        history = historical_data[current_user]
        data = [{
            "Instance": i + 1,
            "Responses": h[0],
```

```
            "Average Score": h[1],
            "Percentage": h[2],
            "Maturity Level": h[3]
        } for i, h in enumerate(history)]

        df = pd.DataFrame(data)
        file_path = filedialog.asksaveasfilename(defaultextension=".xlsx", filetypes=[("Excel files", "*.xlsx")])
        if file_path:
            df.to_excel(file_path, index=False)
            messagebox.showinfo("Success", f"History saved to {file_path}")
    else:
        messagebox.showinfo("Info", "No historical data available to save.")

# Register user
def register_user():
    def register():
        username = username_entry.get()
        password = password_entry.get()
        if username in users:
            messagebox.showerror("Error", "Username already exists.")
        else:
            users[username] = password
            historical_data[username] = []
            save_historical_data()
            messagebox.showinfo("Success", "User registered successfully!")
            register_window.destroy()

    register_window = tk.Toplevel(root)
    register_window.title("Register")
    tk.Label(register_window, text="Username:").pack(pady=5)
    username_entry = tk.Entry(register_window)
    username_entry.pack(pady=5)
```

```
    tk.Label(register_window, text="Password:").pack(pady=5)
    password_entry = tk.Entry(register_window, show="*")
    password_entry.pack(pady=5)
    tk.Button(register_window, text="Register", command=register).pack(pady=10)

# Login user
def login_user():
    def login():
        global current_user
        username = username_entry.get()
        password = password_entry.get()
        if username in users and users[username] == password:
            current_user = username
            messagebox.showinfo("Success", f"Welcome, {username}!")
            login_window.destroy()
            show_main_gui()
        else:
            messagebox.showerror("Error", "Invalid username or password.")

    login_window = tk.Toplevel(root)
    login_window.title("Login")
    tk.Label(login_window, text="Username:").pack(pady=5)
    username_entry = tk.Entry(login_window)
    username_entry.pack(pady=5)
    tk.Label(login_window, text="Password:").pack(pady=5)
    password_entry = tk.Entry(login_window, show="*")
    password_entry.pack(pady=5)
    tk.Button(login_window, text="Login", command=login).pack(pady=10)

# Reset inputs
def reset_maturity_inputs():
    for response_var in response_vars:
```

```
        response_var.set("0")

# Calculate maturity
def calculate_maturity():
    try:
        responses = [int(selected.get()) for selected in response_vars]
        if any(r < 0 or r > 5 for r in responses):
            raise ValueError("Values must be between 0 and 5.")

        average_score = sum(responses) / len(responses)
        maturity_percentage = (average_score / 5) * 100

        maturity_levels = ["Unaware", "Aware", "Defined", "Managed", "Optimized", "Mature"]
        maturity_level = maturity_levels[int(average_score)]

        recommendations = {
            "Unaware": "Establish basic awareness and define data governance policies. Consider appointing a data governance leader and setting simple policies.",
            "Aware": "Start implementing standard practices and basic security measures. Train staff on data management principles.",
            "Defined": "Focus on formalizing data governance structures and practices. Establish a dedicated data governance team.",
            "Managed": "Enhance enforcement and monitoring of policies. Conduct regular audits and involve stakeholders in governance.",
            "Optimized": "Leverage advanced tools and analytics to optimize governance. Integrate AI-driven solutions for proactive management.",
            "Mature": "Maintain leadership in data governance and innovate continually. Explore cutting-edge technologies and industry partnerships."
        }
        comment = recommendations[maturity_level]

        if current_user not in historical_data:
```

```
            historical_data[current_user] = []

        historical_data[current_user].append((responses, average_score, maturity_percentage,
maturity_level, comment))
        save_historical_data()

        result_label.config(text=f"Maturity Level: {maturity_level}")
        percentage_label.config(text=f"Maturity Percentage: {maturity_percentage:.2f}%")
        update_trend_table()
        update_trend_graph()
    except ValueError as e:
        messagebox.showerror("Error", f"Invalid input: {e}")

# Update trend table
def update_trend_table():
    for row in trend_table.get_children():
        trend_table.delete(row)

    if current_user in historical_data:
        for idx, entry in enumerate(historical_data[current_user]):
            _, avg_score, percentage, level, comment = entry
            trend_table.insert("", "end", values=(idx + 1, f"{avg_score:.2f}", f"{percentage:.2f}%", level,
comment))

# Update trend graph
def update_trend_graph():
    if current_user in historical_data:
        scores = [entry[1] for entry in historical_data[current_user]]
        percentages = [entry[2] for entry in historical_data[current_user]]
        instances = range(1, len(scores) + 1)

        ax.clear()
```

```
        ax.plot(instances, scores, marker="o", label="Average Scores", color="blue")
        ax.plot(instances, percentages, marker="x", linestyle="--", label="Percentages", color="green")
        ax.set_title("Trend Analysis of Historical Evaluations")
        ax.set_xlabel("Evaluation Instances")
        ax.set_ylabel("Values")
        ax.legend()
        ax.grid(True)
        canvas.draw()

# Show main GUI
def show_main_gui():
    global response_vars, result_label, percentage_label, trend_table, ax, canvas

    main_window = tk.Toplevel(root)
    main_window.title("Data Governance Maturity Model")
    main_window.configure(bg="lightblue")

    questions = [
        "How are Data Governance Policies implemented?",
        "Ensuring data accuracy through standard practices?",
        "Encryption for sensitive data?",
        "Multi-factor authentication for critical systems?",
        "Enforcing data governance policies?",
        "Data Governance reporting structure?",
        "Classifying the Data Governance Team?",
        "Current data management lifecycle practices?",
        "Data security level?",
        "Leveraging AI and Data Analytics?"
    ]

    tk.Label(main_window, text="Rate each question from 0 (Unaware) to 5 (Mature):", font=("Arial", 14),
bg="lightblue").grid(row=0, column=0, columnspan=7, pady=10)
```

```
    response_vars = []
    for idx, question in enumerate(questions):
        # Add question label
        tk.Label(main_window, text=f"Q{idx + 1}: {question}", anchor="w", bg="lightblue", font=("Arial",
12)).grid(row=idx + 1, column=0, sticky="w", padx=5)

        # Add radio buttons for each question
        response_var = tk.StringVar(value="0")
        response_vars.append(response_var)
        for i in range(6):
            tk.Radiobutton(main_window, text=str(i), variable=response_var, value=str(i),
bg="lightblue").grid(row=idx + 1, column=i + 1, padx=5)

    # Buttons for actions
    button_frame = tk.Frame(main_window, bg="lightblue")
    button_frame.grid(row=len(questions) + 1, column=0, columnspan=7, pady=10)

    tk.Button(button_frame, text="Calculate Maturity Level", command=calculate_maturity, bg="blue",
fg="white").grid(row=0, column=0, padx=10)
    tk.Button(button_frame, text="Save History to Excel", command=save_history_to_excel, bg="green",
fg="white").grid(row=0, column=1, padx=10)

    result_label = tk.Label(main_window, text="", font=("Arial", 12), bg="lightblue", fg="green")
    result_label.grid(row=len(questions) + 2, column=0, columnspan=7, pady=5)

    percentage_label = tk.Label(main_window, text="", font=("Arial", 12), bg="lightblue", fg="green")
    percentage_label.grid(row=len(questions) + 3, column=0, columnspan=7, pady=5)

    # Trend Analysis Table
    trend_table_frame = tk.Frame(main_window, bg="lightblue")
    trend_table_frame.grid(row=len(questions) + 4, column=0, columnspan=7, padx=10, pady=10)
```

```
    trend_table = ttk.Treeview(trend_table_frame, columns=("Instance", "Avg Score", "Percentage",
"Level", "Comment"), show="headings")
    trend_table.heading("Instance", text="Instance")
    trend_table.heading("Avg Score", text="Avg Score")
    trend_table.heading("Percentage", text="Percentage")
    trend_table.heading("Level", text="Level")
    trend_table.heading("Comment", text="Comment")
    trend_table.pack(fill="both", expand=True)

    # Trend Graph
    fig, ax = plt.subplots(figsize=(6, 4))
    canvas = FigureCanvasTkAgg(fig, master=main_window)
    canvas.get_tk_widget().grid(row=len(questions) + 5, column=0, columnspan=7, padx=10, pady=10)

# Setup home page with the specified layout
def setup_home_page(root):
    root.configure(bg="darkgreen")

    # Title: Republic of Kenya Ministry of Defence
    tk.Label(
        root,
        text="REPUBLIC OF KENYA MINISTRY OF DEFENCE",
        font=("Arial", 16, "bold"),
        bg="darkgreen",
        fg="white"
    ).pack(pady=10)

    # Subtitle: DGMM - System
    tk.Label(
        root,
        text="DGMM - System",
```

```
        font=("Arial", 14, "bold"),
        bg="darkgreen",
        fg="white"
    ).pack(pady=5)

    # Buttons for Login and Register
    tk.Button(
        root,
        text="Login",
        command=login_user,
        bg="blue",
        fg="white",
        font=("Arial", 12)
    ).pack(pady=10)

    tk.Button(
        root,
        text="Register",
        command=register_user,
        bg="green",
        fg="white",
        font=("Arial", 12)
    ).pack(pady=10)

load_historical_data()

root = tk.Tk()
root.title("Data Governance Maturity Model")
setup_home_page(root)
root.mainloop()
```

g). MODEL VALIDATION FORM

Model Validation

* Indicates required question

1. 1. On a scale of 1-5, Rate How easy was it for you to complete a full data governance assessment using the model? *

   Mark only one oval.

   1 2 3 4 5

2. 2. Were you able to successfully generate a maturity score and interpret the results without external help? *

   Mark only one oval.

   - Yes
   - No

3. 3. Which part of the model did you find most useful or valuable? *

4. 4. To what extent do you believe the model provides an accurate reflection of your organization's data governance maturity level? (0: Not at all –1: Slightly –2: Moderately –3: Very – 4:Completely) *

Mark only one oval.

0 1 2 3 4

5. 5. After using the model, do the results align with your own understanding or experience of your department's data governance status? Yes / No *

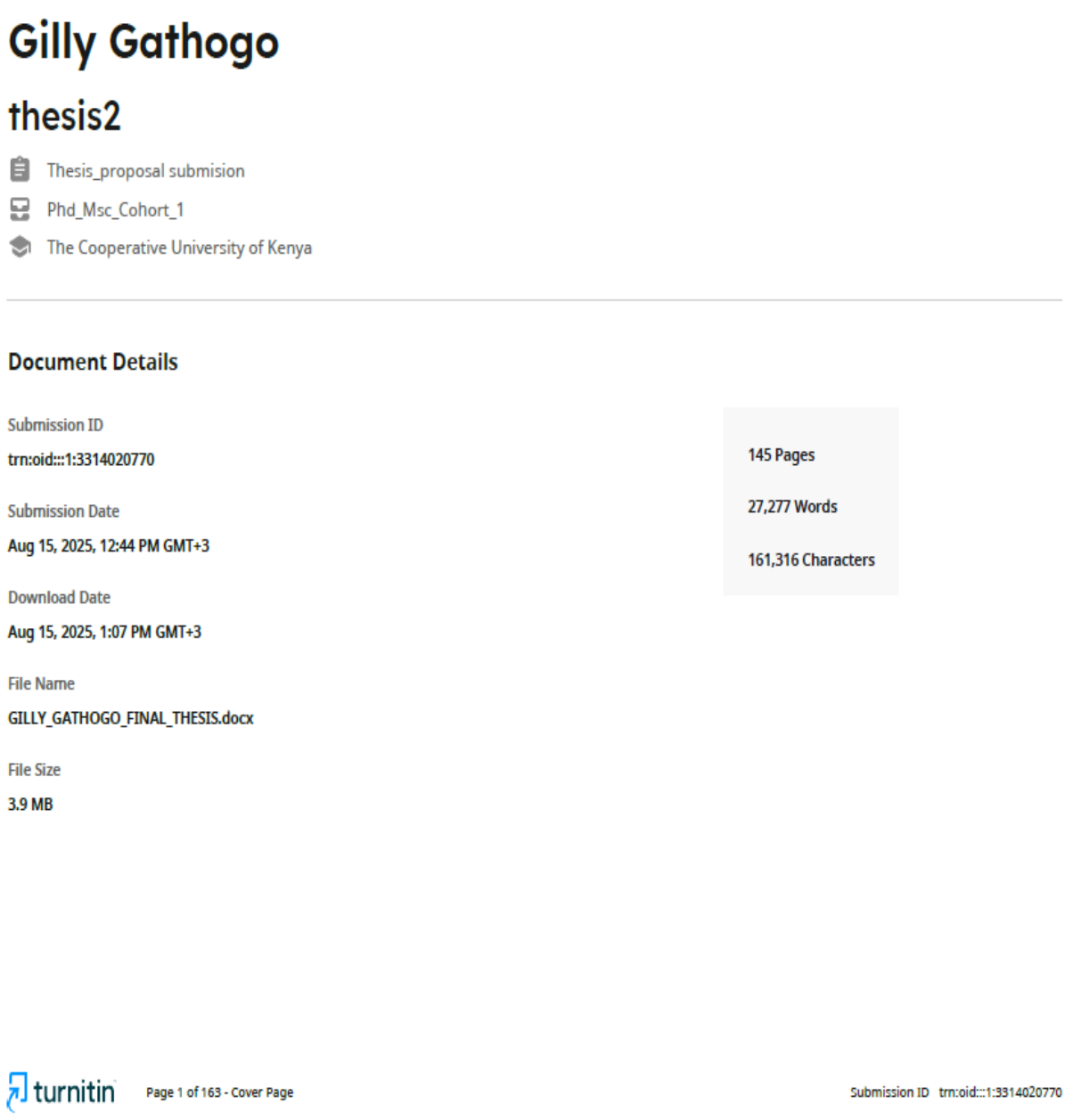
Check all that apply.

- [ ] Yes
- [ ] No
- [ ] Not Sure
- [ ] Other: _____

6. 6. What suggestions do you have to improve the model's ability to evaluate data governance maturity? *

_____

h). SIMILARITY AND AI CONTEST TEST REPORTS



Gilly Gathogo

thesis2

Thesis_proposal submision

Phd_Msc_Cohort_1

The Cooperative University of Kenya

Document Details

Submission ID

trn:oid:::1:3314020770

Submission Date

Aug 15, 2025, 12:44 PM GMT+3

Download Date

Aug 15, 2025, 1:07 PM GMT+3

File Name

GILLY_GATHOGO_FINAL_THESIS.docx

File Size

3.9 MB

145 Pages

27,277 Words

161,316 Characters

turnitin Page 1 of 163 - Cover Page Submission ID trn:oid:::1:3314020770

16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Exclusions

- 1 Excluded Source

Match Groups

- 422 Not Cited or Quoted 15%
  Matches with neither in-text citation nor quotation marks
- 40 Missing Quotations 1%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

13% Internet sources
12% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Gilly Gathogo

thesis2

Thesis_proposal submision

Phd_Msc_Cohort_1

The Cooperative University of Kenya

Document Details

Submission ID

trn:oid:::1:3314020770

Submission Date

Aug 15, 2025, 12:44 PM GMT+3

Download Date

Aug 15, 2025, 1:07 PM GMT+3

File Name

GILLY_GATHOGO_FINAL_THESIS.docx

File Size

3.9 MB

145 Pages

27,277 Words

161,316 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.