



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Denkleiers • Leading Minds • Dikgopolo tsa Dihlatse

ARTIFICIAL INTELLIGENCE FOR ENHANCED MASTER DATA QUALITY MANAGEMENT IN ERP SYSTEMS

by

Carla Jacobs

17053422

Submitted in partial fulfilment of the requirements for the degree

MCom (Informatics)

in the

FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES

at the

UNIVERSITY OF PRETORIA

Study leader:

Prof A van der Merwe

Date of submission

31 August 2024

Declaration regarding Plagiarism

The Department of Informatics emphasises integrity and ethical behaviour with regard to the preparation of all written assignments.

Although the lecturer will provide you with information regarding reference techniques, as well as ways to avoid plagiarism, you also have a responsibility to fulfil in this regard. Should you at any time feel unsure about the requirements, you must consult the lecturer concerned before submitting an assignment.

You are guilty of plagiarism when you extract information from a book, article, web page or any other information source without acknowledging the source and pretend that it is your own work. This doesn't only apply to cases where you quote verbatim, but also when you present someone else's work in a somewhat amended (paraphrased) format or when you use someone else's arguments or ideas without the necessary acknowledgement. You are also guilty of plagiarism if you copy and paste information directly from an electronic source (e.g., a web site, e-mail message, electronic journal article, or CD ROM), even if you acknowledge the source.

You are not allowed to submit another student's previous work as your own. You are furthermore not allowed to let anyone copy or use your work with the intention of presenting it as his/her own.

Students who are guilty of plagiarism will forfeit all credits for the work concerned. In addition, the matter will be referred to the Committee for Discipline (Students) for a ruling. Plagiarism is considered a serious violation of the University's regulations and may lead to your suspension from the University. The University's policy regarding plagiarism is available on the Internet at <http://upetd.up.ac.za/authors/create/plagiarism/students.htm>.

I (full names & surname):	Carla Jacobs
Student number:	u17053422

Declare the following:

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this assignment is my own, original work. Where someone else's work was used (whether from a printed source, the Internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not copy and paste any information directly from an electronic source (e.g., a web page, electronic journal article or CD ROM) into this document.
4. I did not make use of another student's previous work and submitted it as my own.
5. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his/her own work.
6. *I assure that, when using IT/ AI- supported writing tools, I have listed these tools in full in a section titled: "Overview of tools", with their product name, my source of supply (e.g. URL) and information on the functions of the software used as well as the scope of use. I also declare that I have not copied any text directly from any AI tool.*



16 June 2023

Signature

Date

TABLE OF CONTENTS

1	INTRODUCTION.....	13
1.1	BACKGROUND INFORMATION.....	14
1.2	PROBLEM STATEMENT.....	16
1.3	RESEARCH QUESTIONS.....	17
1.4	RESEARCH OBJECTIVES.....	17
1.5	ASSUMPTIONS.....	17
1.6	LIMITATIONS.....	17
1.7	CONTRIBUTION.....	18
1.8	BRIEF CHAPTER OVERVIEW.....	18
2	LITERATURE REVIEW.....	21
2.1	INTRODUCTION.....	21
2.2	OVERVIEW OF MASTER DATA TERMINOLOGY.....	22
2.2.1	MASTER DATA.....	22
2.2.2	MASTER DATA OBJECTS.....	22
2.2.3	RELATIONSHIP BETWEEN MASTER DATA AND TRANSACTIONAL DATA.....	23
2.3	MASTER DATA MANAGEMENT IN ENTERPRISE RESOURCE PLANNING SYSTEMS....	23
2.3.1	MASTER DATA MANAGEMENT.....	23
2.3.2	ENTERPRISE RESOURCE PLANNING SYSTEMS.....	24
2.3.3	MASTER DATA MANAGEMENT IN ENTERPRISE RESOURCE PLANNING SYSTEMS	24
2.3.4	MASTER DATA MANAGEMENT SOFTWARE.....	25
2.3.5	DATA GOVERNANCE IN MASTER DATA MANAGEMENT.....	25
2.4	MASTER DATA QUALITY MANAGEMENT.....	25
2.4.1	DATA QUALITY MANAGEMENT.....	26
2.4.2	MASTER DATA MANAGEMENT AS A DATA QUALITY SOLUTION.....	26
2.4.3	IMPACT OF MASTER DATA QUALITY ON ORGANISATIONAL PERFORMANCE.....	26
2.4.4	FACTORS CONTRIBUTING TO INADEQUATE MASTER DATA QUALITY.....	27

2.4.5	MASTER DATA QUALITY DIMENSIONS.....	28
2.5	ARTIFICIAL INTELLIGENCE.....	29
2.5.1	NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING USED IN TEXT ANALYSIS AND CLASSIFICATION	29
2.5.2	INTEGRATION OF ARTIFICIAL INTELLIGENCE INTO ORGANISATIONAL FUNCTIONS 30	
2.5.3	ENHANCEMENTS TO MASTER DATA QUALITY MANAGEMENT THROUGH THE USE OF ARTIFICIAL INTELLIGENCE.....	30
2.6	THEORETICAL FRAMEWORK.....	31
2.7	CONCLUSION	32
3	RESEARCH METHODOLOGY.....	33
3.1	INTRODUCTION.....	33
3.2	RESEARCH DESIGN.....	34
3.2.1	Research Philosophy.....	34
3.2.2	Research Strategy – Design Science Research	34
3.3	DATA COLLECTION, SAMPLING, MEASUREMENT, AND DATA ANALYSIS CYCLES	36
3.3.1	SURVEY SAMPLE POPULATION.....	36
3.3.2	CYCLE 1 – DATA COLLECTION, SAMPLING, MEASUREMENT AND DATA ANALYSIS 36	
3.3.3	CYCLE 2 – DATA COLLECTION, SAMPLING, DATA ANALYSIS, AND MEASUREMENT 39	
3.3.4	EVALUATION OF FINDINGS.....	41
3.4	ETHICS.....	42
3.5	CONCLUSION	43
4	SYSTEMATIC LITERATURE REVIEW – IDENTIFICATION OF THE PRIMARY MASTER DATA MANAGEMENT FUNCTIONS	44
4.1	INTRODUCTION.....	45
4.2	CONCEPTUAL FOUNDATIONS	45
4.2.1	MASTER DATA MANAGEMENT SYSTEMS.....	45
4.2.2	FUNCTIONS OF MASTER DATA MANAGEMENT SYSTEMS.....	46

4.2.3	ENTERPRISE RESOURCE PLANNING SYSTEMS.....	46
4.3	REASON FOR THE STUDY.....	47
4.3.1	OBJECTIVE	47
4.3.2	RESEARCH QUESTION	47
4.4	RESEARCH METHOD	47
4.4.1	SEARCH STRATEGY	47
4.4.2	SELECTION CRITERIA	48
4.4.3	SCREENING AND SELECTION.....	48
4.4.4	Prisma Flowchart.....	49
4.4.5	Quality Assessment.....	50
4.4.6	Data Extraction.....	52
4.4.7	Data Analysis	52
4.5	RESULTS	53
4.5.1	MASTER DATA MANAGEMENT RESEARCH TRENDS.....	53
4.5.2	MASTER DATA MANAGEMENT FUNCTIONS	54
4.5.3	MASTER DATA MANAGEMENT FUNCTIONS FRAMEWORK	58
4.5.4	MASTER DATA MANAGEMENT FUNCTIONS LITERATURE TRENDS	60
4.5.5	INDUSTRIES	61
4.5.6	MASTER DATA OBJECTS.....	61
4.6	DISCUSSION.....	63
4.7	CONCLUSION	63
5	SYSTEMATIC LITERATURE REVIEW – IDENTIFICATION OF ARTIFICIAL INTELLIGENCE CAPABILITIES APPLICABLE TO MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT.....	65
5.1	INTRODUCTION.....	66
5.2	CONCEPTUAL FOUNDATIONS	67
5.2.1	ARTIFICIAL INTELLIGENCE.....	67
5.2.2	MACHINE LEARNING.....	67
5.2.3	NATURAL LANGUAGE PROCESSING.....	67

5.2.4	DATA QUALITY MANAGEMENT	68
5.2.5	MASTER DATA MANAGEMENT	68
5.3	REASON FOR THE STUDY	68
5.3.1	OBJECTIVE	68
5.3.2	RESEARCH QUESTION	68
5.4	RESEARCH METHOD	68
5.4.1	SEARCH STRATEGY	69
5.4.2	SELECTION CRITERIA	69
5.4.3	SCREENING AND SELECTION	70
5.4.4	PRISMA FLOWCHART	71
5.4.5	QUALITY ASSESSMENT	72
5.4.6	DATA EXTRACTION	76
5.4.7	DATA ANALYSIS	76
5.5	RESULTS	76
5.5.1	RESEARCH TRENDS	76
5.5.2	ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT – INITIAL FINDINGS	77
5.5.3	ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT – CLASSIFICATION	78
5.5.4	ARTIFICIAL INTELLIGENCE CAPABILITIES LITERATURE TRENDS	80
5.6	DISCUSSION	81
5.7	CONCLUSION	82
6	EVALUATION OF THE MASTER DATA MANAGEMENT FUNCTIONS AND ARTIFICIAL INTELLIGENCE CAPABILITIES IDENTIFIED FROM LITERATURE	83
6.1	INTRODUCTION	83
6.2	EVALUATION OF IDENTIFIED PRIMARY MASTER DATA MANAGEMENT FUNCTIONS	84
6.2.1	SURVEY OUTLINE – MASTER DATA MANAGEMENT FUNCTIONS	84
6.2.2	SURVEY RESPONSES – MASTER DATA MANAGEMENT FUNCTIONS	85
6.2.3	ANALYSIS OF SURVEY FINDINGS – MASTER DATA MANAGEMENT FUNCTIONS ...	87

6.3	EVALUATION OF IDENTIFIED ARTIFICIAL INTELLIGENCE CAPABILITIES APPLICABLE TO MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT	91
6.3.1	SURVEY OUTLINE – ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT	91
6.3.2	SURVEY RESPONSES – ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT.....	92
6.3.3	ANALYSIS OF SURVEY FINDINGS – ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT	94
6.4	CONCLUSION	97
7	AI-ENABLED MDM FRAMEWORK – RESEARCH ARTEFACT.....	98
7.1	INTRODUCTION.....	98
7.2	AI-ENABLED MDM FRAMEWORK	98
7.2.1	MOST PROMINENT MASTER DATA MANAGEMENT FUNCTIONS IDENTIFIED.....	99
7.2.2	MOST PROMINENT ARTIFICIAL INTELLIGENCE CAPABILITIES IDENTIFIED FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT	102
7.2.3	LOGICAL MAPPING OF ARTIFICIAL INTELLIGENCE CAPABILITIES TO MASTER DATA MANAGEMENT FUNCTIONS	104
7.2.4	AI CAPABILITIES MAPPED TO CORRESPONDING MASTERD DATA MANAGEMENT FUNCTIONS	106
7.3	AI-ENABLED MASTER DATA MANAGEMENT FRAMEWORK.....	107
7.3.1	THE AI-ENABLED MDM FRAMEWORK AS A USABLE DSR ARTEFACT FOR INDUSTRY PROFESSIONALS	110
7.3.2	THE AI-ENABLED MDM FRAMEWORK AS A RESEARCH TOOL FOR FUTURE RESEARCH IN THE FIELD OF AI AND MDM	113
7.4	CONCLUSION	114
8	CONCLUSION	115
8.1	INTRODUCTION.....	115
8.2	FINDINGS AND ANALYSIS	116
8.3	SUMMARY OF CONTRIBUTIONS	118
8.4	FUTURE RESEARCH	118
8.5	CONCLUSION	119

9 REFERENCES..... 120

LIST OF FIGURES

Figure 1: Structure of Research Paper - Position of Chapter 1	13
Figure 2: Brief Chapter Overview.....	19
Figure 3: Structure of Research Paper - Position of Chapter 2.....	21
Figure 4: Diffusion of Innovations Theory, adopted from Rogers (2010).....	31
Figure 5: Structure of Research Paper - Position of Chapter 3.....	33
Figure 6: Design Science Research Process Cycle, based on Vaishnavi and Kuechler (2004)	35
Figure 7: Research Methodology - DSR Cycle 1 Process Steps	37
Figure 8: Research Methodology - DSR Cycle 2 Process Steps	39
Figure 9: Research Methodology - Evaluation of Findings	42
Figure 10: Structure of Research Paper - Position of Chapter 4.....	44
Figure 11: PRISMA Flowchart – Adapted from Moher et al. (2009).	49
Figure 12: MDM Research Trends.....	53
Figure 13: MDM Functions Literature Trends.....	60
Figure 14: Industries Utilising MDM Systems.....	61
Figure 15: Master Data Objects	62
Figure 16: Structure of Research Paper - Position of Chapter 5.....	65
Figure 17: PRISMA Flowchart – Adapted from Moher et al. (2009).	71
Figure 18: AI for MDM and DQM Research Trends	77
Figure 19: MDM Functions Literature Trends.....	80
Figure 20: Structure of Research Paper - Position of Chapter 6.....	83
Figure 21: Survey Responses: MDM Functions – Impact on Master Data Quality	85
Figure 22: MDM Functions – Most Human Involvement.....	86
Figure 23: MDM Functions – Most Time-Consuming	86
Figure 24: Analysis of Survey Findings: MDM Functions – Impact on Master Data Quality	87
Figure 25: Analysis of Survey Findings: MDM Functions – Requires Most Human Involvement.....	88
Figure 26: Analysis of Survey Findings: MDM Functions – Most Time-Consuming	89

Figure 27: Analysis of Survey Findings: MDM Functions – Combined Findings	90
Figure 28: AI Capabilities – Most Important for Organisations to Adopt.....	92
Figure 29: AI Capabilities – Most Easily Implemented into an MDM System	93
Figure 30: AI Capabilities – Usefulness of an AI-Enabled MDM Reference Framework	93
Figure 31: Analysis of Survey Findings: AI Capabilities – Most Important for Organisations to Adopt...	94
Figure 32: Analysis of Survey Findings: AI Capabilities – Easily Implemented into an MDM System....	95
Figure 33: Analysis of Survey Findings: AI Capabilities – Combined Findings.....	96
Figure 34: Structure of Research Paper - Position of Chapter 7	98
Figure 35: AI-Capabilities Framework	106
Figure 36: Structure of Research Paper - Position of Chapter 8.....	115

LIST OF TABLES

Table 1: SLR Databases	48
Table 2: Quality Assessment Codes	50
Table 3: Sources Included in SLR	51
Table 4: Quality Evaluation Scores	52
Table 5: MDM Functions Identified from Literature.....	57
Table 6: MDM Functions Framework	59
Table 7: Master Data Objects	62
Table 8: SLR Databases	70
Table 9: Quality Assessment Codes	72
Table 10: Sources Included in SLR.....	74
Table 11: Quality Evaluation Scores	75
Table 12: AI Capabilities Identified from Literature.....	78
Table 13: AI Capabilities for MDM and DQM - Classification.....	79
Table 14: Most Prominent MDM Functions	101
Table 15: Most Prominent AI Data Management Capabilities	103
Table 16: Logical Mapping of AI Capabilities to MDM Functions	105
Table 17: AI-Enabled MDM Framework (Design Science Research Artefact)	109

ARTIFICIAL INTELLIGENCE FOR ENHANCED MASTER DATA QUALITY MANAGEMENT IN ERP SYSTEMS

ABSTRACT

This research project investigates Artificial Intelligence (AI) capabilities that can be applied to Master Data Management (MDM) systems within Enterprise Resource Planning (ERP) environments to enhance master data quality. MDM systems are pivotal in managing high-quality, reliable master data within organisations, particularly in ERP systems where data integrity directly impacts business processes and decision-making.

Recent studies underscore the growing interest in automating MDM tasks using AI. However, challenges persist, such as managing master data from diverse sources, which often results in duplicates and inconsistencies and leads to bad business decisions. This research project addresses the importance of maintaining high-quality master data by exploring the potential of AI to address data-related challenges effectively. As AI technology advances, leveraging its capabilities holds promise for streamlining MDM processes and optimising business operations in ERP environments.

The primary research artefact is an AI-Enabled MDM Framework that categorises MDM functions to appropriate AI data management capabilities and provides suggested AI tools and methods identified from the literature. The framework was developed using a Systematic Literature Review (SLR) approach with a Design Science Research (DSR) methodology. It categorises MDM functions and tasks to appropriate AI data management capabilities.

The research artefact was created by evaluating and combining the findings from both Systematic Literature Reviews. To evaluate the findings from the Systematic Literature Reviews, a survey with corresponding questions related to the research topic was presented to data management professionals and analysed. The most prominent MDM functions were defined by classifying the MDM functions identified from the Systematic Literature Reviews according to the highest number of references in the literature and the highest percentages in the survey responses. The most prominent AI capabilities were also defined by classifying the AI data management capabilities according to the highest percentages in the survey responses and the highest number of references in the literature.

The usefulness of the research artefact as a DSR contribution is demonstrated by presenting case studies of how the framework can be applied to Master Data Management challenges. The findings can be applied to future research as a theoretical framework for research in AI and MDM, and industry professionals can also use it as a reference framework for integrating AI tools into MDM systems.

Keywords: Master Data; Master Data Management; Data Quality; Enterprise Resource Planning; Artificial Intelligence; Machine Learning; Natural Language Processing

1 INTRODUCTION

The figure below displays the layout and position of Chapter 1 in relation to the rest of the research paper.

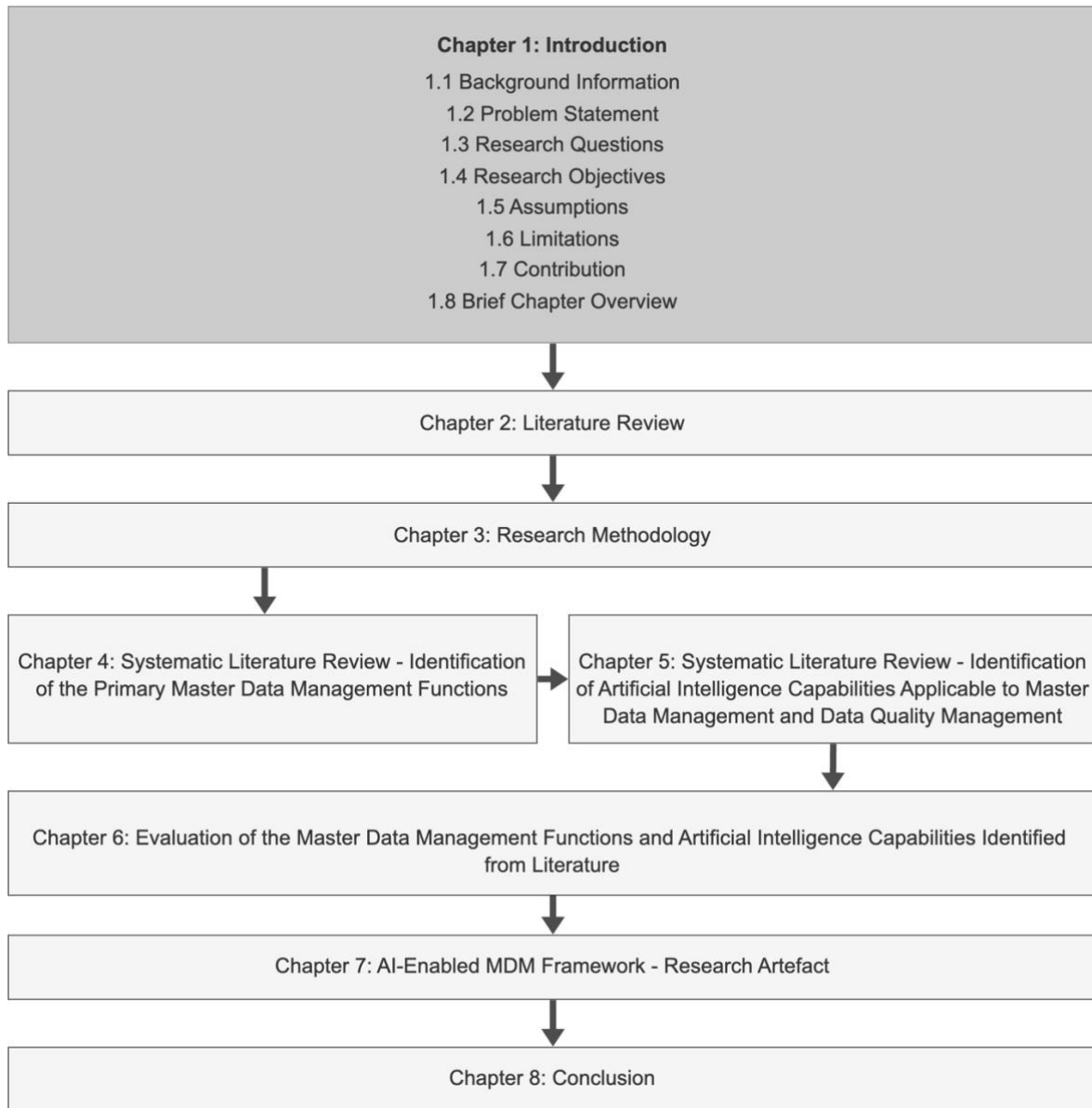


Figure 1: Structure of Research Paper - Position of Chapter 1

This research paper investigates the application of Artificial Intelligence (AI) to Master Data Management (MDM) for enhanced master data quality in an Enterprise Resource Planning (ERP) system. MDM systems are valuable tools widely adopted to maintain reliable master data within organisations. It is essential for the quality of master data in ERP systems to be adequately maintained and continually improved because bad data quality has a negative impact on business processes and reporting (Hechler et al., 2020).

Research suggests that organisations are eager to automate time-consuming tasks related to MDM, like data de-duplication and data harmonisation, to allow employees to focus on activities that yield greater business insights (Dutta et al., 2021). Leveraging AI tools can have significant advantages in data processing and data management. Researchers Dutta et al. (2021) have explored solutions such as automated data harmonising, which uses dictionary matching, machine learning, and fuzzy text similarity. Other researchers have investigated using AI tools, such as neural networks, to improve the identification of duplicate records in MDM systems (Zong et al., 2017).

Overall, organisations are facing significant challenges in governing increasing levels of data (Krieger & Schorr, 2019). Managing master data from multiple sources can lead to duplicate, inaccurate, and inconsistent data, which can lead to bad business decisions (Haneem, Ali, et al., 2017; Vilminko-Heikkinen & Pekkola, 2017). Therefore, maintaining high-quality master data is crucial for organisations that rely on digital processes, impacting business process efficiency, decision-making, and revenue (Riesener et al., 2022).

More research is necessary on this critical topic of master data quality management. AI capabilities have the potential to effectively address data-related challenges and subsequently optimise business processes for organisations using ERP software (Dutta et al., 2021). As AI advances, there are increasing opportunities for data process automation. Therefore, it is vital to research these technologies' applications in improving MDM systems' data quality.

1.1 BACKGROUND INFORMATION

Master data refers to crucial business information used across multiple applications, systems and business processes in an organisation (Jaksic et al., 2017). Managing master data can be challenging due to large data volumes and isolated datasets (Krieger & Schorr, 2019). Inaccurate data can also arise from problems such as system user interface issues and the substantial time and effort required for data maintenance activities such as data profiling tasks (Krieger & Schorr, 2019; Riesener et al., 2022). Fortunately, Master Data Management (MDM) software solutions exist to deal with these challenges by providing a platform for data stewards to manage consistent definitions of master data and distribute it to relevant business units (Maedche, 2010).

MDM tools have become more mature, and the MDM market is primarily dominated by industry leaders in ERP solutions, like SAP, Oracle, and IBM (Fernando & Haddela, 2017). Many enterprises have adopted these tools in their IT departments to solve the problem of low master data quality. They use MDM tools to manage data creation, maintenance and other data quality activities and processes (Hechler et al., 2020).

Hechler et al. (2020) state that MDM systems provide organisations with functionality such as de-duplication, data governance, and data stewardship. Furthermore, Prokhorov and Kolesnik (2018) argue

that using these tools increases transparency, facilitates data distribution between systems, and eliminates duplicate records.

The motivation for using MDM systems is to maintain a source of consistent, accurate and relevant master data used in different business applications and units within the organisation (Hikmawati et al., 2021). However, managing master data in an MDM system presents numerous data quality challenges, such as incomplete, inaccurate, irrelevant and duplicate master data records (Nulhusna et al., 2022).

According to researchers, a lack of clear data ownership, ambiguous data governance policies and standards, and inconsistent data cleansing efforts are significant reasons for poor quality master data (Silvola et al., 2011). Researchers have identified data quality dimensions as a guideline for maintaining high-quality data, highlighting dimensions such as accuracy, completeness, and consistency.

Significant progress has been made in Natural Language Processing tools over the past decade. Despite some remaining issues, integrating and using Natural Language Processing tools, such as Chatbot technology, have gained popularity in many organisations (Galanis et al., 2021). As these tools become more sophisticated, many organisations integrate AI tools such as Machine Learning and Natural Language Processing into their information systems.

Natural Language Processing and Machine Learning have been used in different scenarios, such as automatically classifying user requirements in software development and categorising helpdesk tickets with Machine Learning. Using these tools can shorten processing time and prevent mistakes due to human error (Miliano et al., 2020; Shehadeh et al., 2021). A study by Gantayat et al. (2019) suggests that the future of user interfaces in business applications, including Enterprise Resource Planning systems, will rely on natural language interactions.

Li and Xu (2022) investigated using AI-enabled technologies in Customer Relationship Management (CRM), which they found to significantly impact customer experience and improve customer loyalty. Additionally, Hechler et al. (2020) found that integrating Artificial Intelligence into MDM systems can substantially alleviate the workload of data stewards. This can be accomplished by utilising the AI tool to assist with configuring fuzzy matching engines to improve the accuracy of matching duplicate records in MDM systems and to aid data stewards with repetitive administrative tasks.

Combining AI tools like NLP and ML to analyse natural language, find patterns, and use the findings to predict new information could automate several data quality management activities. Integrating AI tools into MDM systems allows organisations to improve master data quality in their enterprise resource planning systems, enhancing business process efficiency and decision-making (Riesener et al., 2022).

1.2 PROBLEM STATEMENT

With the increasing volume and variety of data and the growing regulations organisations must comply with, maintaining master data quality has become costly and tedious for many (Hechler et al., 2020; Schäffer & Leyh, 2017). This is due to several factors, including poor data quality management, complex Enterprise Resource Planning (ERP) landscapes, and unclear data ownership (Maedche, 2010; Silvola et al., 2011; Vilminko-Heikkinen & Pekkola, 2019).

Master Data Management (MDM) gained prominence in 2009, coinciding with the emergence of the Big Data movement (Haneem, Azmi, et al., 2017). Since then, organisations have adopted MDM systems to combat data quality issues and enhance overall data quality in ERP systems. However, data quality remains challenging due to the increasing amount of data created and the complexities of managing master data from multiple sources (Haneem, Azmi, et al., 2017).

According to researchers such as Ibrahim et al. (2021), organisations face severe master data quality problems due to technological, managerial, organisational, stakeholder, and external factors. Inadequate management of master data quality can result in severe consequences for an organisation, such as decisions based on inaccurate data and inefficient business processes (Hikmawati et al., 2021; Schäffer & Leyh, 2017).

Gröger (2021) argues that quality data forms the foundation of AI enablement in industrial enterprises. ERP systems are negatively impacted by data management issues, such as extracting source data several times due to heterogeneous data landscapes (Gröger, 2021). Therefore, researchers must explore new approaches and methodologies for managing master data to increase data quality.

Although integrating AI with MDM systems to enhance the overall quality of master data is an approach suggested by some researchers, such as Hechler et al. (2020), it has not been extensively researched. According to them, AI in MDM systems can help rapidly integrate data sources and reduce the workload of data stewards. A deeper examination of the topic is necessary to contribute to the existing literature and guide data practitioners on adopting AI in data management functions.

The primary purpose of this study is to discover how AI capabilities can be utilised in MDM systems to improve the quality of master data within an ERP system. The goal is for AI researchers and data management researchers to use the findings as a guide for adopting AI tools in MDM systems. Furthermore, the research can be used to determine the overall impact that AI will have on data management functionality.

This research will also enable data professionals to understand the usefulness and value that can be added to an organisation's ERP systems by integrating AI tools with their MDM systems for higher-quality data.

1.3 RESEARCH QUESTIONS

Primary Research Question:

- Which Artificial Intelligence capabilities can be applied to Master Data Management functions to enhance the quality of master data in an Enterprise Resource Planning system?

Secondary Research Questions:

- What are the primary functions of a Master Data Management system in an organisation utilising Enterprise Resource Planning systems?
- What are the primary capabilities of Artificial Intelligence that can be applied to Master Data Management functions and Data Quality Management functions?

The first secondary research question is addressed in the Systematic Literature Review carried out in Chapter 4 of the research paper, and the second secondary research question is addressed in the Systematic Literature Review conducted in Chapter 5 of the research.

The primary research question is addressed in Chapter 7 of the research. Chapter 7 presents the study's primary research contribution in the form of a design science research artefact.

1.4 RESEARCH OBJECTIVES

The primary objective of the research is to investigate how Artificial Intelligence capabilities can enhance Master Data Management functions to improve the quality of master data within an Enterprise Resource Planning system.

1.5 ASSUMPTIONS

Assumptions regarding the value and accuracy of the literature used in the Systematic Literature Reviews (SLR) are made. The researcher assumes that the information and findings of the literature papers included in the SLR have been peer-reviewed and are of high quality.

In the evaluation phase of the research, the survey respondents are assumed to be truthful and honest and that the findings from the research survey are trustworthy.

1.6 LIMITATIONS

The research is limited to the methodology of the research strategy chosen, as described in Chapter 3. Design Science Research (DSR) has been selected as the strategy for adopting a pragmatist research paradigm. The research focuses on creating a helpful artefact in the form of an *AI-Enabled Master Data Management (MDM) framework*. The artefact will provide practitioners and researchers with a framework that outlines the capabilities of Artificial Intelligence (AI) to assist with specific MDM functions and

increase data quality in enterprise resource planning systems. The research will not investigate specific AI tools but the general capabilities of emerging AI tools, as stated in the literature.

The researcher's access to research participants who work in data management also limits the research. Master Data Management (MDM) is a specialised subject in which few IT industry professionals possess extensive expertise. It is important to note that the data management professionals who were asked to complete the research survey mainly have an SAP data consulting background. This could limit the generalisability of the research outcome because the MDM functionality provided by SAP is only universal to some MDM vendor solutions. Nonetheless, SAP is ranked as one of the most prominent MDM software tools, and having conducted the study with experts in the SAP-specific data management field is of high value.

1.7 CONTRIBUTION

The primary contribution of the research will be an evaluated and refined framework that maps AI capabilities to Master Data Management (MDM) functions. The purpose is to discover MDM functionalities that may benefit from the application or integration of AI tools for organisations that maintain master data in ERP systems.

Integrating AI with an MDM tool can help increase an organisation's data quality, improve business processes, and achieve greater overall efficiencies and reporting (Riesener et al., 2022). This artefact contribution, the *AI-Enabled Master Data Management Framework*, will present practitioners with a tool for adopting AI technologies in data management.

It will provide academics with a potential solution for bad data quality in ERP systems in the form of a framework with specific MDM functions mapped to the relevant AI capabilities for each function. The framework will also shed light on the complexity of adopting AI data management. This framework may also serve as a base for future research by identifying the necessary skills data management practitioners must cultivate to ensure their skills and expertise remain relevant.

1.8 BRIEF CHAPTER OVERVIEW

The diagram below provides a broad overview of the content in each chapter of the research, which is structured into eight Chapters.

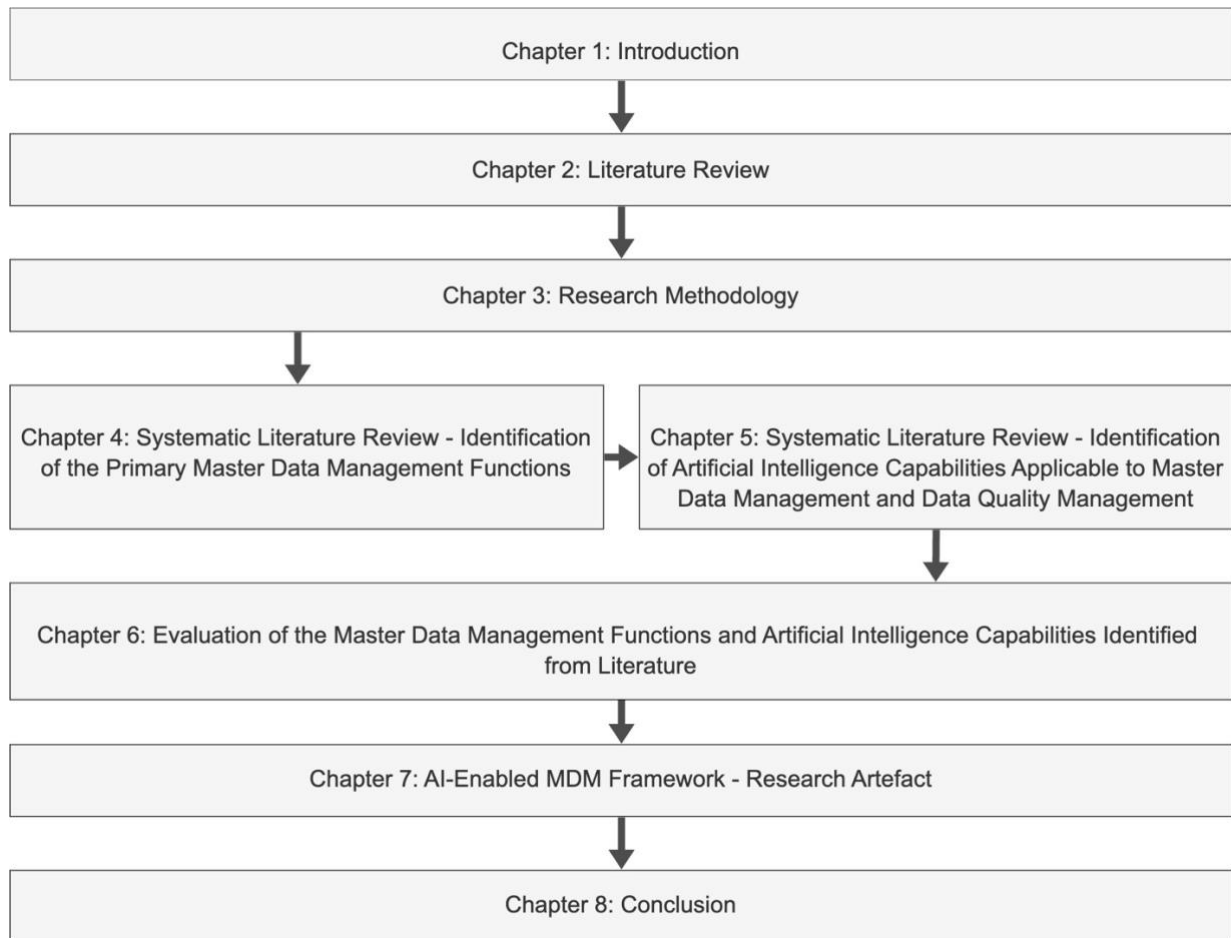


Figure 2: Brief Chapter Overview

Chapter 1 provides the background information, problem statement, research questions and objectives, research assumptions, and limitations and ends with the expected contribution of the research. Chapter 2 contains the literature review of research papers on applying Artificial Intelligence (AI) applications to Master Data Management (MDM) systems to improve the quality of master data within an Enterprise Resource Planning (ERP) system. It provides an overview of the master data terminology, MDM in ERP, Master Data Quality Management, and Artificial Intelligence and describes the theoretical framework chosen for the research topic.

Chapter 3 outlines the research methodology, describes the research design, and outlines the research strategy and philosophy chosen. Furthermore, the chapter specifies and graphically depicts the two respective cycles of data collection, sampling, measurement, and data analysis and concludes with the ethical considerations of the research. Chapter 4 comprises the first Systematic Literature Review (SLR) to identify the primary MDM functions from current academic literature. It outlines the conceptual foundations of the SLR, the reasons for the study, the research method and discusses the SLR results.

Chapter 5 comprises the second Systematic Literature Review (SLR) undertaken to identify AI capabilities applicable to MDM functions from current academic literature. It outlines the conceptual foundations of

the SLR, the reasons for the study, the research method and discusses the SLR results. Chapter 6 serves as the evaluation phase of the study. It evaluates and analyses the findings of both systematic literature reviews conducted in Chapters 4 and 5 and includes the results of the research survey carried out to establish which MDM functions have the most significant impact on master data quality, which require the most human involvement, and which are the most time-consuming. It also includes the research survey results applicable to establishing which of the AI capabilities identified are the most important for organisations to adopt and which AI capabilities would be most easily implemented into the MDM functions extracted from the literature.

Chapter 7 presents the primary research artefact, the *AI-Enabled MDM Framework*. This chapter outlines the steps taken to produce the final research artefact. It demonstrates its usefulness by presenting four case study examples of how it can help solve master data quality issues in an MDM system. Chapter 8 concludes with the findings and analysis of the research and provides a summary of the research contributions along with future research suggestions.

2 LITERATURE REVIEW

The figure below displays the layout and position of Chapter 2 in relation to the rest of the research paper.

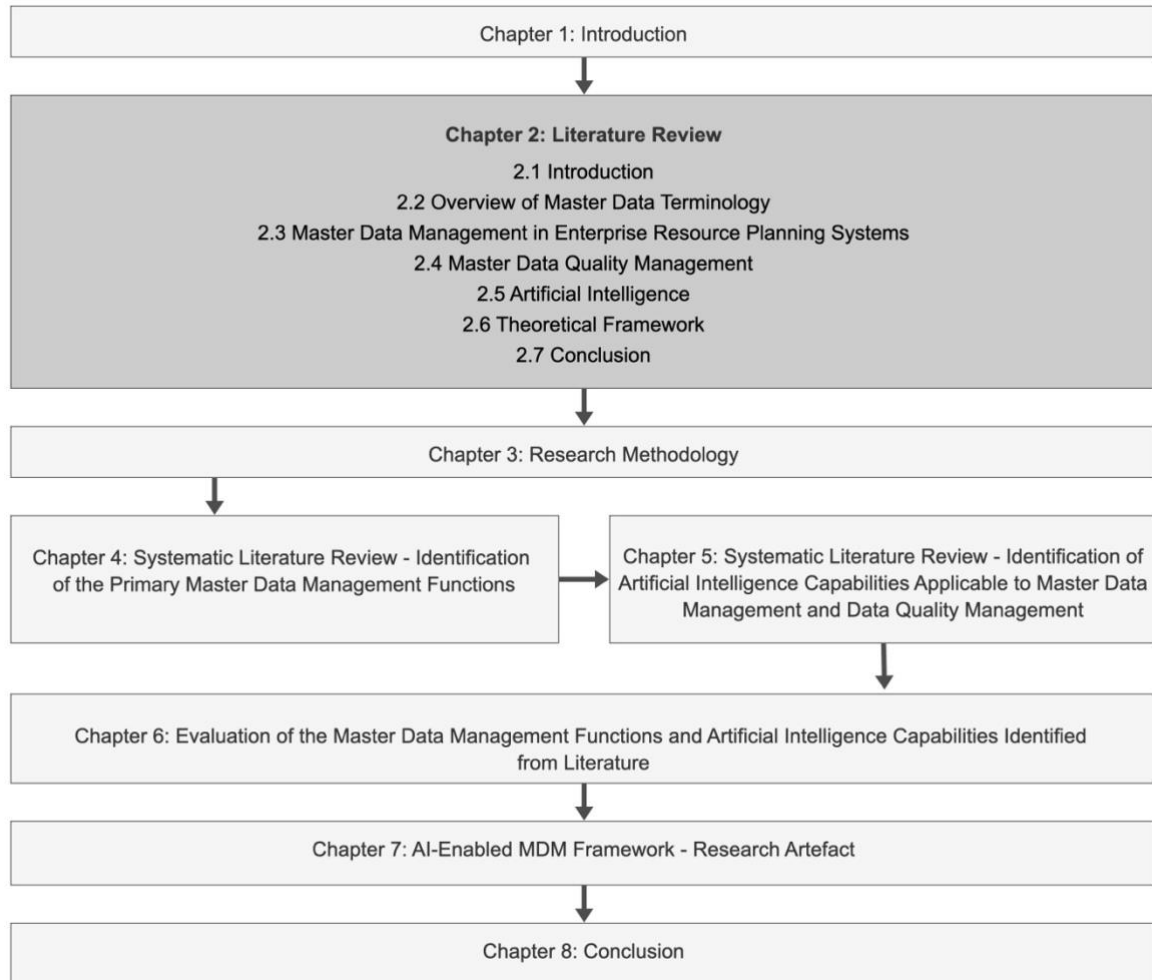


Figure 3: Structure of Research Paper - Position of Chapter 2

2.1 INTRODUCTION

This literature review examines master data quality in Master Data Management (MDM) systems, significantly impacting business process efficiency and accurate reporting and decision-making in organisations (Riesener et al., 2022). Master data refers to crucial business entities, including customers, suppliers, and products, and MDM systems are employed to ensure consistent and high-quality definitions of this critical data (Maedche, 2010).

The literature indicates several barriers to achieving high-quality master data, such as complex enterprise systems landscapes (Maedche, 2010), data redundancies (Gröger, 2021), and increasing regulatory and legal provisions (Otto et al., 2012).

Although research on using Artificial Intelligence (AI) tools to enhance data quality is limited, some researchers, such as Hechler et al. (2020), have investigated the potential of integrating AI with MDM functions to improve the quality of master data. Their findings suggest that AI can significantly reduce manual processing for data stewards, simplify MDM configuration, and detect duplicate values in the system.

2.2 OVERVIEW OF MASTER DATA TERMINOLOGY

This section aims to define the fundamental terminologies relevant to master data. First, the definition of master data is established, followed by an introduction of the primary master data objects identified in the existing research literature. Furthermore, terminologies applicable to master data, such as Transactional and Reference Data, are defined.

2.2.1 MASTER DATA

Master Data is an essential set of data objects created by business processes and commonly used among most organisational departments (Knapp & Lutz, 2021). An organisation's product or material, customer, and vendor or supplier information are examples of data widely referred to as Master Data (Schäffer & Leyh, 2017). This data performs various analytical and operational tasks in different business units (Fernando & Haddela, 2017).

Master Data is utilised in several systems, applications, and business processes within the organisation and contains critical business entities with corresponding attributes (Jaksic et al., 2017). It is created once and repeatedly used over time, with occasional modifications to the data at infrequent intervals (Knolmayer & Röthlin, 2006). Knolmayer and Röthlin (2006) It also explains that procedures which ensure the adherence of the data captured in the organisation to the business rules are employed to regulate the distribution of Master Data.

2.2.2 MASTER DATA OBJECTS

As per the study by Knapp and Lutz (2021), there is a shared consensus among data professionals regarding the types of data objects that fall under the category of master data. These data objects primarily include Customer, Vendor, and Item or Product master data. Knolmayer and Röthlin (2006) confirm this by stating that master data in ERP systems can represent Customers, Suppliers, and Products.

Entities and their corresponding attributes classify master data objects. Depending on the type of data stored in each attribute, these attribute values can have different data types, such as Numeric, String, Boolean, or Date (Knapp & Lutz, 2021).

2.2.3 RELATIONSHIP BETWEEN MASTER DATA AND TRANSACTIONAL DATA

Knapp and Lutz (2021) emphasise the difference between Master Data and Transactional Data. Master Data represents business objects; it is regularly created and can be used repeatedly in business processes. In contrast, Transactional Data is created during business processes, but it is only used during a single business process and is subsequently utilised for analytics or documentary purposes.

2.3 MASTER DATA MANAGEMENT IN ENTERPRISE RESOURCE PLANNING SYSTEMS

This section examines the significance of Master Data in an Enterprise Resource Planning (ERP) system. As noted in a literature review by Haneem, Ali, et al. (2017), interest in Master Data Management research has surged since the onset of the significant data era in 2004. Wang et al. (2009) underscore the role of master data in daily business operations, particularly in information systems such as Customer Relationship Management (CRM) and ERP systems.

2.3.1 MASTER DATA MANAGEMENT

Master Data Management (MDM) is defined by Fernando and Haddela (2017) as a method of maintaining high-quality, reliable and consistent data among data sources. An organisation adopts it to maintain a consolidated view of its master data. Master Data Management is a method of bridging the silos between different information systems and organisational units (Vilminko-Heikkinen & Pekkola, 2017). Prokhorov and Kolesnik (2018) define Master Data Management as the tools and processes used in the cycle of defining and managing the core data of a company.

Jaksic et al. (2017) state that MDM software provides business users in an organisation with a single reference point of the master data objects and attributes relevant to them in their daily business processes. MDM is not only a technology but refers to a function within an organisation that stores standard definitions of master data, which are shared across different business units (Haneem, Ali, et al., 2017).

According to a literature review by Haneem, Ali, et al. (2017), organisations with various data sources use MDM to solve the problem of isolated data sources controlled in silos across different business units. Data warehouse and data mart tools have been utilised to fulfil the need for enterprise data integration. The drawback of these technologies is that they fail to manage organisations' real-time data validation requirements. Thus, MDM has emerged as a better solution for real-time data validation of data dispersed across the organisation's data sources (Haneem, Ali, et al., 2017).

2.3.2 ENTERPRISE RESOURCE PLANNING SYSTEMS

Enterprise Resource Planning (ERP) systems represent a unified view of an enterprise's business units and describe a company's value chains (Maedche, 2010). They are crucial business enablers for decision-making and daily operations in an organisation (Knolmayer & Röthlin, 2006).

ERP systems support business activities by unifying scattered information from separate organisational units, such as finance, human resources, and sales (Knolmayer & Röthlin, 2006). The primary objective of an ERP system is to manage transactional and master data, allowing business users and processes to use the same data (Maedche, 2010).

Standardisation of master data objects in ERP systems has matured, but some challenges have surfaced as ERP landscapes become more complex in large organisations (Maedche, 2010). It is no longer unusual for large companies to use several ERP systems uniquely configured to support individual business processes best. These separate systems create the need for more coordination between a company's different information systems and ERP systems (Knolmayer & Röthlin, 2006).

2.3.3 MASTER DATA MANAGEMENT IN ENTERPRISE RESOURCE PLANNING SYSTEMS

Fernando and Haddela (2017) have categorised different technical Master Data Management approaches, among which the ERP-centric approach stands out. This approach is implemented by organisations that depend on ERP systems for their day-to-day business processes and rely on Master Data Management systems to maintain master data used in these processes. The interdependency between these two systems makes it crucial to ensure that the MDM and ERP system processes overlap to ensure data integration (Fernando & Haddela, 2017).

As the need for coordination between different information systems and instances of ERP systems grows, large companies often use multiple instances of ERP (Knolmayer & Röthlin, 2006). Knolmayer and Röthlin (2006) have highlighted that business decisions made using erroneous data could have unfavourable consequences for the organisation, and inadequate master data will subsequently harm the quality of transactional data in the ERP system. Ensuring the high reliability of master data is crucial for an organisation to make well-informed business decisions (Riesener et al., 2022).

Maedche (2010) researched an ERP-centric MDM approach and identified three primary methods for using MDM within an ERP system. These methods include collaborative authoring MDM, operational MDM, and analytical MDM. Collaborative authoring MDM focuses on creating and synchronising master data to support processes such as data stewardship and management of a product catalogue. Operational MDM involves utilising master data in various operational systems. At the same time, analytical MDM focuses on maintaining cleansed master data for use in analytical environments such as Business Intelligence platforms and Data Warehouses (Maedche, 2010).

2.3.4 MASTER DATA MANAGEMENT SOFTWARE

Prominent software vendors like SAP, IBM and Oracle have developed and offer comprehensive Master Data Management software solutions (Fernando & Haddela, 2017). These solutions include SAP NetWeaver MDM, Oracle MDM, IBM InfoSphere MDM, and Informatica MDM (Haneem, Ali, et al., 2017).

Using one of the abovementioned tools to demonstrate their capabilities and advantages, according to Prokhorov and Kolesnik (2018), the Master Data Management component in the SAP MDM tool helps manage relationships with customers by increasing transparency, effortlessly distributing data between systems and eliminating duplicate records. It ensures accurate data exchange and reduces costs across the supply chain, even in heterogeneous, distributed applications. It improves data integrity and optimises a company's decision-making processes (Prokhorov & Kolesnik, 2018).

Although most MDM offerings have matured, some MDM solutions do not provide data profiling capabilities, which are used during the initial stages of MDM, such as data quality assessment (Haneem, Ali, et al., 2017). Due to this gap, Haneem, Ali, et al. (2017) suggest that partnership growth between MDM software providers and vendors that provide data quality tools is evident.

2.3.5 DATA GOVERNANCE IN MASTER DATA MANAGEMENT

Data Governance is a fundamental aspect of Master Data Management, as established in a research study by Vilminko-Heikkinen and Pekkola (2019). The study addresses the technical and organisational aspects of managing master data, which requires leadership and authority. They define Data Governance as the structures and processes that protect and enhance data as a critical asset in an organisation.

According to Haneem, Ali, et al. (2017), an organisation must establish a Data Governance team comprising business and information technology staff with defined roles and responsibilities to ensure good Data Governance in an MDM system. This team should take responsibility for critical data governance factors in MDM, such as managing data elements and entities, ensuring information policy adherence, and maintaining high-quality master data.

2.4 MASTER DATA QUALITY MANAGEMENT

Master data has become increasingly complex, posing significant challenges for companies to maintain it with high quality (Riesener et al., 2022). Zong et al. (2017) note that poor data quality is one of the leading causes of ERP project failure, emphasising the critical role of high-quality master data in providing organisations with a competitive advantage.

Master data is crucial in an organisation, and to improve master data quality, it is essential to implement an MDM solution that ensures the data is consistent, accurate, and current (Nulhusna et al., 2022).

2.4.1 DATA QUALITY MANAGEMENT

Data Quality Management (DQM) is an organisational function that aims to verify that data is fit for use and satisfies the expectations of various data users. It involves continuously planning, implementing, and monitoring data quality improvement strategies and activities to increase the overall quality of an organisation's systems (Nulhusna et al., 2022).

2.4.2 MASTER DATA MANAGEMENT AS A DATA QUALITY SOLUTION

Researchers widely acknowledge the importance of data quality management in an MDM system. Nulhusna et al. (2022) consider data quality crucial to an organisation's success. Riesener et al. (2022) claim that high-quality master data provides valuable insights into process performance. Organisations need to prioritise the development of processes and tools to ensure a high level of master data quality across different business units, and they can achieve this by using Master Data Management tools (Schäffer & Leyh, 2017).

Master Data Management can reduce data quality issues like redundancies and inconsistencies, as Haneem, Ali, et al. (2017) stated. One approach to lowering redundant data records in an ERP system is to integrate all the information of a master data object, such as a Vendor, into a single master data record, resulting in a "single record of truth" and a better view of the critical master data in the organisation.

Knapp and Lutz (2021) suggest creating a golden record for each master data object, a single record representing the object's most accurate and complete version. This approach can reduce duplicates in an ERP system and ensure accurate and high-quality data.

Jaksic et al. (2017) highlight the standard functionalities in an MDM system that contribute to overall master data quality, such as removing duplicates, standardising data, and establishing custom business rules that ensure the data created in the system is correct and of high quality.

2.4.3 IMPACT OF MASTER DATA QUALITY ON ORGANISATIONAL PERFORMANCE

The increasing digitalisation in society and businesses poses a significant challenge for organisations to keep up with a constantly changing business environment. Schäffer and Leyh (2017) have identified data as the foundation of the digital economy, given its crucial role in improving business process efficiency. Furthermore, they found that the efficiency of inter-organisational collaboration largely depends on an organisation's level of data quality.

Nulhusna et al. (2022) also noted the impact of data quality on organisational success. Data quality can lead to inefficiencies, revenue reduction, poor decision-making, and customer satisfaction. As a result, any organisation with digitised processes heavily relies on the quality of its master data (Riesener et al., 2022).

Master data quality is essential for organisations that rely on digitised processes (Riesener et al., 2022). Poor-quality master data can hinder organisational performance by causing employees to double-check proposed information or rework existing products instead of searching for the existing product data in the Master Data Management system (Riesener et al., 2022).

The study by Schäffer and Leyh (2017) also found that 80% of companies believe bad data quality harms their business success. However, the same study also found that 84% of companies consider managing high data quality very costly. Therefore, organisations must balance the cost of maintaining high-quality data with its benefits to their business processes and outcomes.

2.4.4 FACTORS CONTRIBUTING TO INADEQUATE MASTER DATA QUALITY

According to research by Ibrahim et al. (2021), which aims to pinpoint the factors contributing to bad data quality, there are various reasons for inadequate master data. They suggest that data quality policies and standards, data quality assessment, governance, and continuous improvement determine master data quality. Another fundamental issue in managing master data is a lack of data ownership, unclear data governance practices, and inconsistent data cleansing efforts (Silvola et al., 2011).

Unstructured and inconsistent processes, complex architectures, and data fragmentation are vital challenges organisations face in managing high-quality master data (Vilminko-Heikkinen & Pekkola, 2017). Additionally, Riesener et al. (2022) have identified problems with system user interfaces as a significant reason why erroneous data is created in a Master Data Management system.

Furthermore, Haneem, Ali, et al. (2017) state that managing data from multiple sources can result in data quality issues such as inaccuracy, inconsistency, and duplication. Vilminko-Heikkinen and Pekkola (2017) support this claim by stating that unreliable analytics based on duplicate, incomplete, and inaccurate data can lead to wrong business decisions. To address these challenges, implementing a Master Data Management system can help consolidate critical data commonly used across business units and cleanse the data through de-duplication activities (Haneem, Ali, et al., 2017).

In the study by Zong et al. (2017), duplication was identified as a key data quality issues arising in Master Data Management systems. They aimed to develop an approach for de-duplicating material records in ERP systems. During their research, they identified some of the reasons why ERP systems contain duplicate material records. They highlighted the inefficiency of traditional duplicate detection mechanisms. They also revealed that inconsistent naming of the same material by different business users and the lack of distinct identifiers for material records in the system are some of the reasons for duplicate material records (Zong et al., 2017). Organisations with a growing number of products and product variants are particularly vulnerable to this data quality issue (Zong et al., 2017).

2.4.5 MASTER DATA QUALITY DIMENSIONS

The quality standard in an MDM system can be described using data quality dimensions. Maintaining high-quality master data is essential to enhancing the overall data quality in an ERP system. Knapp and Hasibether (2011) identify accuracy, completeness, and consistency as data quality criteria for maintaining high data standards.

Several researchers have identified different data quality dimensions that are considered the most important for determining the level of master data quality in an organisation. Knapp and Hasibether (2011) identify data quality dimensions as Uniqueness, Uniformity, Objectivity, Interpretability, Actuality, Consistency, Completeness, Accuracy, and Relevance. Nulhusna et al. (2022) state that data quality dimensions are subject to users' unique needs, but Completeness, Accuracy, and Relevance are the most commonly used dimensions.

Additionally, Riesener et al. (2022) define data quality dimensions as Consistency, Completeness, and Correctness. Herrmann et al. (2022) conducted their research with the data quality dimensions of Accuracy, Consistency, Completeness, Timeliness and Appropriate amount of data. However, they exclude the Timeliness dimension from their study for simplicity. Finally, Schäffer and Leyh (2017) state that the most critical data quality dimensions are Consistency, Integrity, Availability, Timeliness, and Accuracy.

Each of the data quality dimensions, as described in the study done by Knapp and Hasibether (2011) is defined below:

- **Uniqueness:** For a data object or entity to be considered unique, it must only exist in a single record.
- **Uniformity:** This refers to the data object or entity always being interpreted similarly.
- **Objectivity:** Data is considered objective when the information in the data record does not change based on the person who entered it.
- **Interpretability:** Interpretability means that the data content can be interpreted without supplementary or additional knowledge,
- **Actuality:** Actuality means that the data is current and timely.
- **Consistency:** The data must be logically coherent.
- **Completeness:** Entities must be completely defined according to the relevant requirements.
- **Accuracy:** Accuracy refers to how precisely the data is defined and how representative it is of the information it describes.
- **Relevance:** The relevance dimension states that data should not contain any unnecessary information.

These data quality criteria dimensions can be used to determine whether the master data maintained in an ERP system is high quality and can drive accurate business insights and reporting.

The three data quality dimensions Riesener et al. (2022) discussed are Consistency, Correctness, and Completeness. The Consistency dimension measures the level of contradictory data within a dataset. Correctness estimates how accurately the data represents reality, and the level of correctness can be determined by comparing master data with data from validated datasets. Completeness assesses the amount of missing data in a master data record (Riesener et al., 2022).

For master data records to be considered high quality, the data attributes should be as complete and descriptive as possible within the context of what the master data record is representing (Wang et al., 2009).

2.5 ARTIFICIAL INTELLIGENCE

Artificial Intelligence is growing rapidly and revolutionising numerous industries such as manufacturing, automotive, and machine building by significantly improving business processes (Gröger, 2021). Cybersecurity researchers such as Chmielarz and Pabian (2020) have studied the effectiveness of applying Artificial Intelligence to data and information management systems to safeguard an organisation's information against cybercrime attacks. Other topics that have been researched include the integration of Artificial Intelligence and database management systems (Wannalai & Mekruksavanich, 2019).

Chmielarz and Pabian (2020) claim that employees and managerial staff commonly lack an understanding of Artificial Intelligence's impact on organisational activities. They also observe underlying scepticism among the employees and managerial staff about AI-based solutions, which could hinder the adoption of Artificial Intelligence in organisations.

Further research is necessary to determine how Artificial Intelligence can be integrated into organisations to enhance current business processes, eliminate more manual tasks, and support organisations in achieving higher competitive advantages.

2.5.1 NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING USED IN TEXT ANALYSIS AND CLASSIFICATION

Natural Language Processing (NLP) and Machine Learning (ML) are two subfields of Artificial Intelligence that are often used together in text analysis and classification. NLP algorithms make sense of human language, while ML algorithms self-optimize through experience (Riesener et al., 2022).

Recent research has explored the potential of integrating NLP and ML in information systems within organisations. Riesener et al. (2022) suggest combining NLP and ML to automate manual data quality management activities. These AI tools can analyse natural language, identify patterns, and use the findings to predict new information. This can help to streamline data quality management processes, improve accuracy, and lower the risk of mistakes caused by human bias or oversight. As a result, the

implementation of AI tools such as NLP and ML has the potential to enhance decision-making processes and drive business outcomes.

Other studies have explored applying NLP and ML tools in text analysis and classification scenarios. For instance, Shehadeh et al. (2021) investigated utilising an NLP tool in software development to automatically classify functional and non-functional requirements based on user requirements analysis. Typically, user requirements are expressed in natural language and require significant human effort to classify accurately. This tool can reduce the time and cost involved in the manual classification task during software development (Shehadeh et al., 2021).

Similarly, Miliano et al. (2020) evaluated the effectiveness of a Machine Learning-based helpdesk ticketing application that categorises problems by analysing the titles of the tickets logged into the system. Using Machine Learning to classify helpdesk tickets can reduce processing time and prevent errors that may result from human oversight (Miliano et al., 2020).

2.5.2 INTEGRATION OF ARTIFICIAL INTELLIGENCE INTO ORGANISATIONAL FUNCTIONS

Integrating Artificial Intelligence into different organisational functions has been a topic of interest in recent research. According to Jeong et al. (2020), Machine Learning, a subset of Artificial Intelligence, statistical analysis, algorithms, probability, and mathematical optimisation techniques are used to create models using data. Jeong et al. (2020) researched the integration of Machine Learning technology in process lead time prediction models, which can be used to estimate production time in master data systems.

In Customer Relationship Management (CRM), AI-enabled technologies have also been increasingly adopted to enhance customer experience. Li and Xu (2022) conducted a study that evaluated using AI-enabled technologies such as virtual assistants, email marketing, visual detection, voice recognition, and product recommendations in CRM. The study found that the use of these technologies had a significant impact on customer experience and improved customer loyalty.

2.5.3 ENHANCEMENTS TO MASTER DATA QUALITY MANAGEMENT THROUGH THE USE OF ARTIFICIAL INTELLIGENCE

Research has shown that manual maintenance processes and the ever-increasing volumes of data are significant challenges for data quality management in organisations (Schäffer & Leyh, 2017). To overcome these challenges, there is a considerable need to automate manual data maintenance and processing tasks, such as data quality management activities, the creation of new data records, and measuring data quality tasks (Schäffer & Leyh, 2017).

Artificial Intelligence tools, such as Machine Learning algorithms, can significantly improve the identification of duplicate master data records (Zong et al., 2017). For instance, Zong et al. (2017) applied a probabilistic neural network to establish the resemblances between master data records in an ERP system, significantly improving the identification of duplicate master data records.

Moreover, Riesener et al. (2022) investigated using Artificial Intelligence tools such as NLP and ML to automate Master Data Management. This suggested approach could free up highly skilled organisational resources by automating appropriate data maintenance activities and allowing these resources to focus on more essential tasks. Organisations implementing AI in their MDM system gain a competitive advantage because of data maintenance's increased speed and scale and subsequent reliability of business processes (Riesener et al., 2022).

2.6 THEORETICAL FRAMEWORK

The theoretical framework employed in this research is the Diffusion of Innovation (DOI) theory, popularised by Rogers (2010). The DOI Theory is a valuable guideline for research about adopting innovative technologies, especially technologies with a broad scope of potential uses within different contexts (Rogers, 2010).

Rogers (2010) states people generally possess different levels of willingness to adopt innovations and that the adoption rate of innovation is positively influenced by aspects such as relative advantage, compatibility, trialability, and visibility and negatively influenced by complexity. Innovation can refer to a new practice, idea, philosophy or product. As so-called *Innovators* and *Early Adopters* embrace and endorse a particular innovation, the pinnacle of adoption is eventually reached, called the Critical Mass, which can be seen in the peak of the chart below.

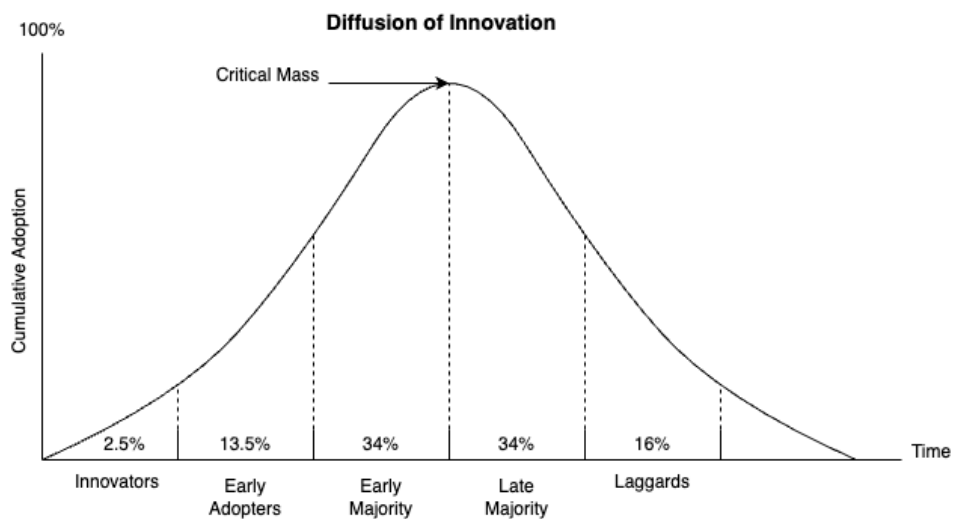


Figure 4: Diffusion of Innovations Theory, adopted from Rogers (2010).

The characteristics that influence innovation's success are identified by Rogers (2010) as "observability, relative advantage, compatibility, trialability and complexity". The DOI theory has been used in numerous research papers focusing on the impact of innovations and the likelihood that people will adopt them. It would, therefore, be suitable to employ this theory in the present study to examine the effect that factors

such as complexity, relative advantage, and compatibility may have on adopting Artificial Intelligence for enhanced MDM systems.

Using the DOI theory, the research can be positioned so that the result sheds light on the factors that make applying AI in Data Management worthwhile. In the current study, the innovation that will be examined is the usage of AI tools in Master Data Management functions within organisations that use ERP systems.

The study aims to explore the AI adoption curve within the context of MDM and larger ERP systems, and thus, it is essential to base this research on the DOI theory. The adoption rate may be positive when employing AI tools in MDM functions, which is simple and provides a relative advantage to the organisation. By using the DOI theory to identify the factors that positively and negatively influence the usage of AI tools for different MDM functions, the research can provide a guideline of which MDM functionality organisations should consider enhancing via integration with AI tools that offer data management capabilities.

2.7 CONCLUSION

In conclusion, MDM is crucial for organisations to maintain high-quality data in their ERP systems. However, various challenges still exist in maintaining good master data quality due to manual maintenance processes, increasing data volumes, and the complexity of the data. Fortunately, researchers have started investigating the use of AI tools such as ML and NLP to enhance master data quality management in organisations.

Data maintenance and processing tasks automated using AI can significantly improve business processes' speed, accuracy, and reliability, leading to a competitive advantage for organisations that implement AI in their MDM systems. Therefore, future research in this area should continue to focus on developing new AI-based approaches to improving master data quality management in ERP systems.

3 RESEARCH METHODOLOGY

The figure below shows the layout and position of Chapter 3 in relation to the rest of the research paper.

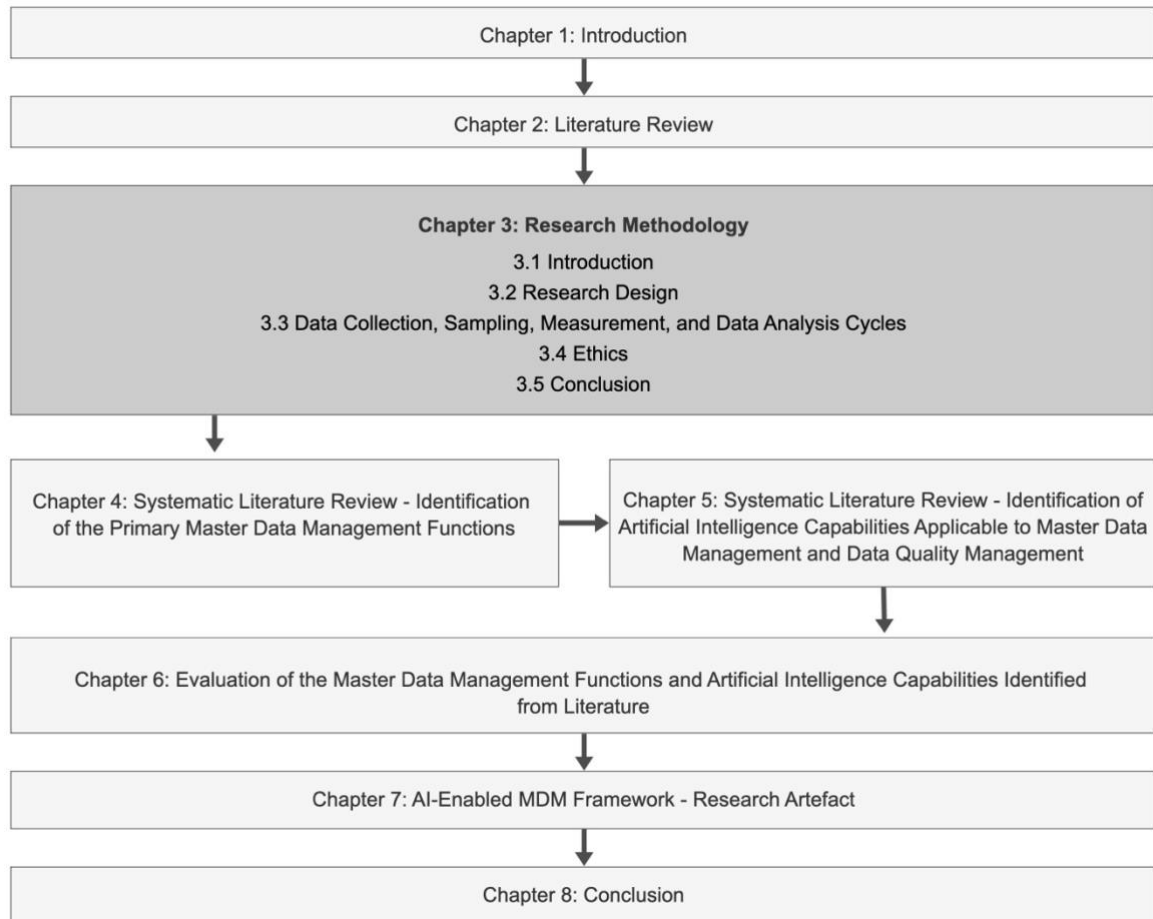


Figure 5: Structure of Research Paper - Position of Chapter 3

3.1 INTRODUCTION

This research follows a Design Science Research (DSR) approach (Hevner et al., 2004). The necessary data is collected using a systematic literature review. The research findings are then evaluated by a survey targeting data management professionals. The following section describes these methods, the data analysis approach, and ethical considerations for the study in more detail.

3.2 RESEARCH DESIGN

As mentioned, a Design Science Research (DSR) strategy was chosen for this research paper. This strategy was first defined by Hevner et al. (2004) and has since been established as a viable approach in information systems research by other scholars (Goldkuhl, 2012). The research design is based on a pragmatist research philosophy, described in more detail in the following section.

3.2.1 RESEARCH PHILOSOPHY

This study adopted a pragmatist research paradigm appropriate for design science research (Goldkuhl, 2012; Hevner et al., 2004). Goldkuhl (2012) stated that there are different types of pragmatism in design science research: methodological pragmatism, general function pragmatism, referential pragmatism, and local functional pragmatism. This study employed the local functional pragmatism approach, which focuses on creating an artefact that addresses a specific need and can be applied to the context for which it was designed (Goldkuhl, 2012).

According to the pragmatist paradigm, theory and practical artefacts are necessary for design science research, and the effectiveness of a created artefact should be measured by its implications in practice (Hevner et al., 2004). Goldkuhl (2012) describes the pragmatist attitude as seeking to build a better world by searching for possibilities within the social and technical spheres. Pragmatism combines prospective knowledge with normative knowledge. The focus on normative knowledge in pragmatist research ensures that the research outcome considers the desirable result, not only the possible outcome.

Prospective knowledge, also referred to as usefulness, and normative concern, also called moral rightness, are equally important in pragmatist research, as stated in Goldkuhl (2012). This makes the pragmatist paradigm ideal for the research topic at hand, where the application of artificial intelligence was researched to ensure that the research outcome was equally desirable and possible.

3.2.2 RESEARCH STRATEGY – DESIGN SCIENCE RESEARCH

This study employed a DSR approach, as devised in the seminal paper by Hevner et al. (2004). DSR was a suitable research methodology for investigating the integration of Artificial Intelligence (AI) with Master Data Management (MDM) systems, as this approach produced a useful research artefact. Academic literature previously lacked an appropriate framework for integrating AI capabilities into MDM systems. Therefore, the primary objective of this research was to create this framework and present it as a DSR artefact.

Due to the complex nature of the artefact that was to be developed for the research topic, adopting a DSR strategy that leverages design processes alongside iterative development and evaluation cycles ensured that the developed artefact addressed the research problem. The artefact produced by this research assists data management practitioners in enhancing specific MDM functions using AI capabilities.

As seen in the diagram below, an iterative design approach was applied to design an accurate model of the different MDM functions within an Enterprise Resource Planning (ERP) system and the AI capabilities which can enhance these MDM functions. The research strategy diagram is adapted from the original by Vaishnavi and Kuechler (2004) visualising this study's research approach.

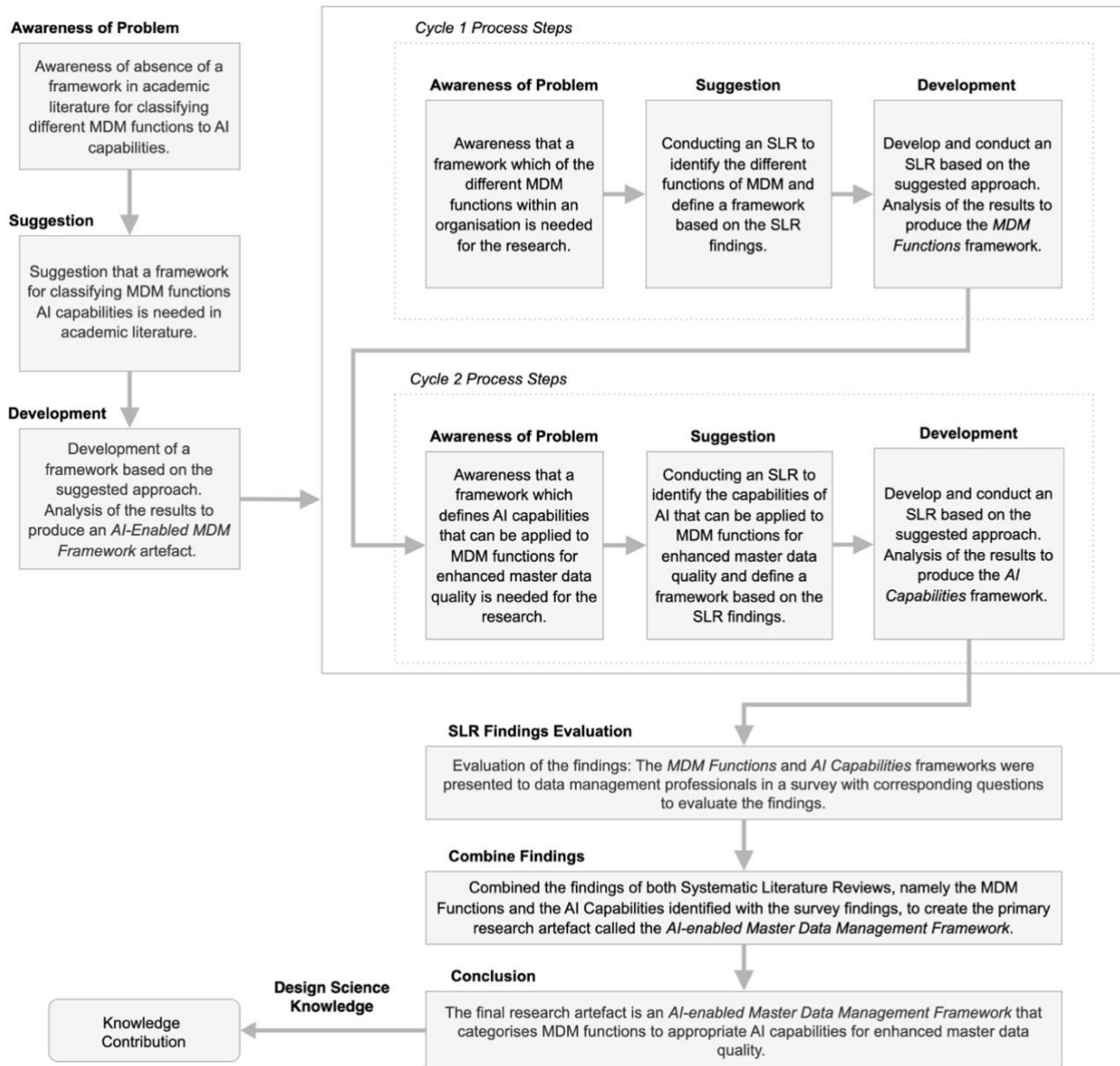


Figure 6: Design Science Research Process Cycle, based on Vaishnavi and Kuechler (2004)

The DSR Process Model consists of two cycles, each consisting of process steps for the *Awareness of the Problem*, *Suggestion*, and *Development*. The SLR findings were evaluated by presenting data management professionals with a survey based on both SLR findings. After that, the findings and survey results were combined to produce the final research artefact. The knowledge contribution to design

science knowledge is an *AI-Enabled MDM Framework* research artefact delivered in the *Conclusion* process step.

The DSR artefact was therefore developed via two Systematic Literature Reviews (SLR) cycles. The two distinct cycles of data collection cycles, sampling, measurement, and data analysis, are discussed in more detail in the following section.

3.3 DATA COLLECTION, SAMPLING, MEASUREMENT, AND DATA ANALYSIS CYCLES

Systematic Literature Reviews (SLR) were employed as this research's primary data collection strategy, and a survey was conducted with data management professionals to measure the quality and usefulness of the SLR findings. The following sections provide an overview of each cycle of data collection, sampling, measurement, and analysis that was undertaken to produce the study's design science research contribution.

3.3.1 SURVEY SAMPLE POPULATION

A survey was conducted and sent to data management professionals to assess the findings of the data collected during the Systematic Literature Reviews conducted in two cycles, which are further described in the following sections. The process involved selecting industry professionals who work with master data in an ERP context and presenting them with the research findings and corresponding survey questions.

The researcher accessed a group of 10-20 industry professionals who work with Master Data Management (MDM) and who primarily have experience working in SAP data consulting roles. The survey sample population was thus limited to 10-20 participants. This sample group all work for the same organisation, and the researcher gained permission to conduct a survey within this company targeting these specific industry professionals. The research participants' years of experience working with MDM systems varied, and they also had differing levels of seniority within the organisation.

The following sections describe the two cycles of Systematic Literature Reviews that were conducted and describe the survey questions that were posed to the MDM professionals relating to the SLR findings.

3.3.2 CYCLE 1 – DATA COLLECTION, SAMPLING, MEASUREMENT AND DATA ANALYSIS

Cycle 1 – Overview

As the *Cycle 1 Process Steps* in the diagram below indicate, an SLR data collection strategy was used to establish a framework of different Master Data Management (MDM) functions within an organisation. These MDM functions extracted from the literature were evaluated after the second SLR was conducted, which extracted the AI capabilities applicable to the MDM functions extracted from the first SLR. The findings from both SLRs were only evaluated in the SLR Findings Evaluation step.

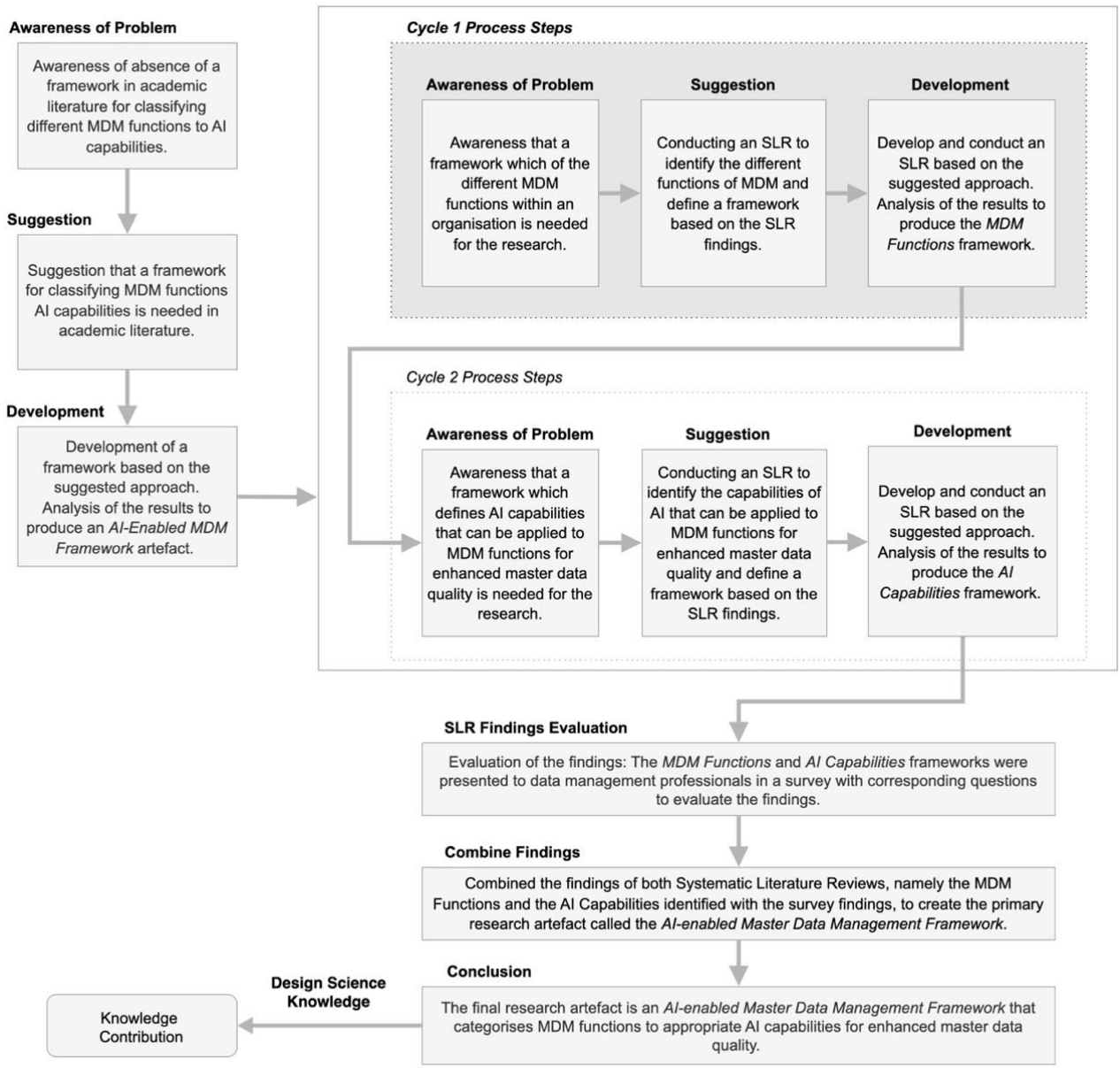


Figure 7: Research Methodology - DSR Cycle 1 Process Steps

The data collection, sampling, measurement and data analysis steps for this SLR will be discussed in detail in the following sections.

Cycle 1 – Data Collection: Systematic Literature Review to Identify Master Data Management Functions

The first step in the data collection strategy was to conduct a Systematic Literature Review (SLR) on the most common Master Data Management (MDM) functions in organisations that use Enterprise Resource Planning (ERP) software. This step was necessary to create a framework of MDM Functions extracted from the current literature.

Identifying the most common MDM functions across different organisations with ERP landscapes was crucial to ensuring that the research artefact created in the final step of the research was accurate and generalisable. Therefore, the MDM functions that are the most prominent in literature were considered the predominant categories and served as the basis of the final research artefact. The sampling details are discussed in more detail in the next section.

Cycle 1 – Sampling: Systematic Literature Review

The first cycle of research utilised an SLR approach for data sampling. The SLR methodology was designed to reduce bias in the selection, extraction, and publication of data and aims to find all appropriate studies to address the research question (Nightingale, 2009).

The following academic databases were searched for relevant data during the SLR:

- IEEE Xplore Digital Library,
- ACM Digital Library,
- ScienceDirect, and
- Scopus.

The search string used to search the databases listed above was: (*“Master Data” OR “Master Data Management”*) AND (*“Enterprise Resource Planning” OR ERP*).

The findings of the SLR were analysed and used to create a framework that categorises the primary Master Data Management functions in an organisation. Thematic analysis was selected as the study’s data analysis strategy, which is a descriptive analysis method for finding themes within data in qualitative research (Castleberry & Nolen, 2018). The framework’s usability was determined through a survey conducted with data management industry professionals, as described in detail in the next section.

Cycle 1 – Measurement: Survey for Evaluating MDM Functions

A survey was conducted and sent to data management professionals to assess the findings of the data collected during the SLR that identified the primary MDM functions in organisations using ERP software. The process involved selecting industry professionals who work with master data in an ERP context and presenting them with the framework of identified primary MDM functions with the corresponding survey questions. The questions on the survey that applied to the identified MDM functions are the following:

- Which Master Data Management Functions have the most significant impact on master data quality?
- Which Master Data Management Functions need the most human involvement?
- Which Master Data Management Functions are the most time-consuming?

The survey response options comprised the MDM functions that were extracted from the literature.

Cycle 1 – Data Analysis

After the survey was presented to data management professionals, the data gathered was analysed to refine the framework created during the first SLR process. The purpose was to determine which MDM functions identified as the most important for maintaining high-quality data in an MDM system.

3.3.3 CYCLE 2 – DATA COLLECTION, SAMPLING, DATA ANALYSIS, AND MEASUREMENT

Cycle 2 – Overview

As the *Cycle 2 Process Steps* in the diagram below indicated, a second SLR was conducted in the *Suggestion* process step as part of the data collection strategy. This SLR was directed toward creating a framework that identifies AI capabilities that can be applied to data management functions.

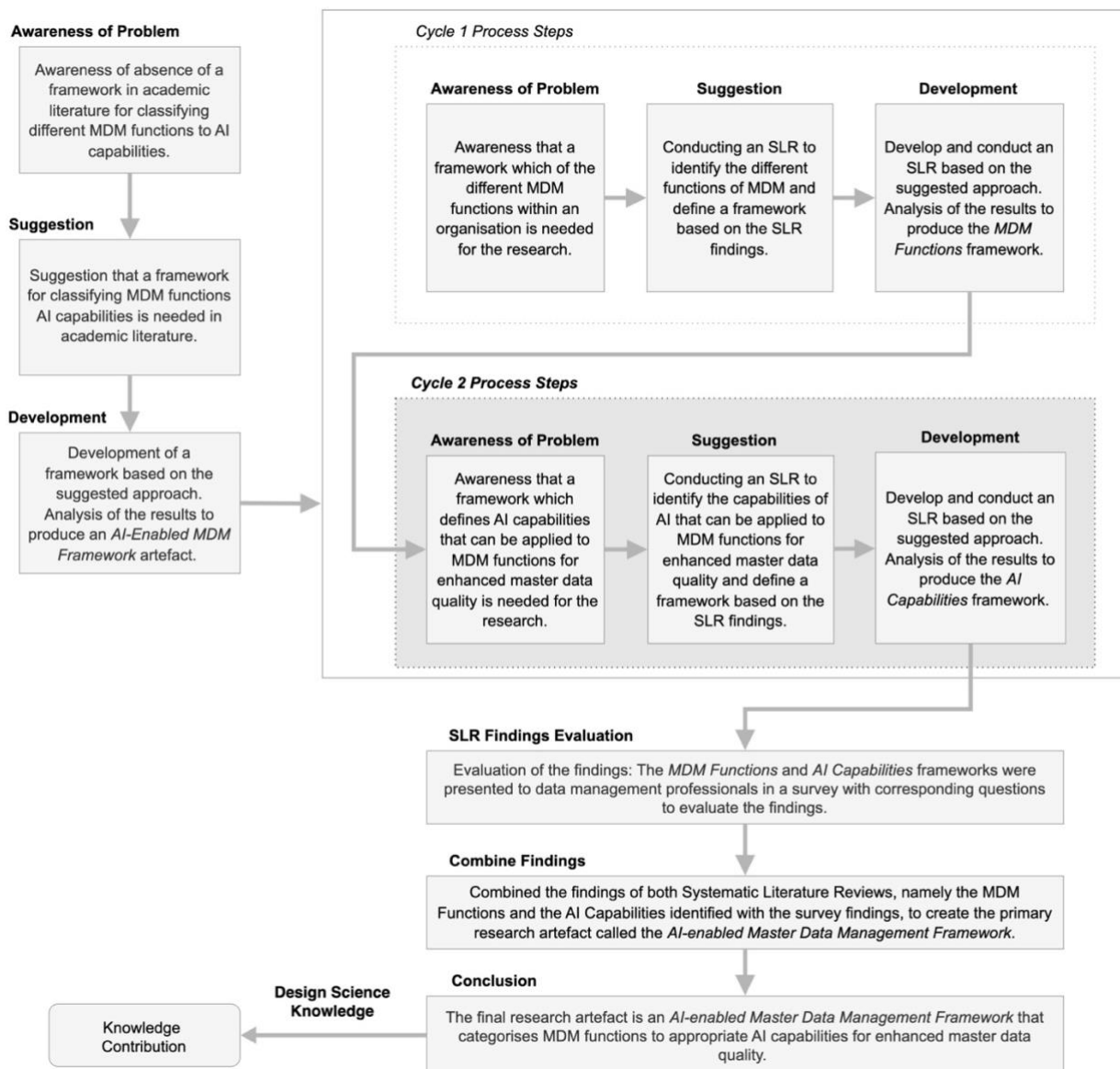


Figure 8: Research Methodology - DSR Cycle 2 Process Steps

The final research artefact is later created by integrating the findings of the previous data collection steps, namely the *MDM Functions Framework* developed by the first SLR and the AI data management capabilities identified in the second SLR. After the final *Development* step, the result is an *AI-Enabled MDM Framework* that categorises MDM functions and tasks to appropriate AI capabilities. The final evaluation activity takes place during the *Final Evaluation* process step. At this step in the process, the *AI-Enabled MDM Framework* is presented to data management professionals in a research survey with questions applicable to the framework. The steps for data collection, sampling, and data analysis are discussed in detail in the following sections.

Cycle 2 – Data Collection: Systematic Literature Review to Identify Data Management Capabilities of Artificial Intelligence Tools

The second SLR was conducted to identify the data management capabilities of Artificial Intelligence (AI) tools. Its findings are integrated with those from the first SLR, which identified the organisational Master Data Management functions. These findings were combined to produce a framework that maps specific AI data management capabilities to an organisation's various MDM functions. The framework created in this SLR is not the primary research artefact but serves as the basis for the development of the primary DSR artefact in the Combine Findings step of the process flow.

Cycle 2 – Sampling: Systematic Literature Review

The sampling for the second research cycle was carried out using a SLR method. As mentioned, the SLR methodology was designed to find all appropriate academic research to address this study's primary research question (Nightingale, 2009).

The following academic databases were searched:

- IEEE Xplore Digital Library,
- ACM Digital Library,
- ScienceDirect,
- Emerald Insight and
- Google Scholar.

The search string that was used to search the databases for the second SLR was the following: ("*Master Data Management*" OR "*Data Quality Management*") AND ("*Artificial Intelligence*" OR AI OR "*Machine Learning*" OR ML OR "*Natural Language Processing*" OR NLP).

Cycle 2 – Measurement: Survey for Evaluating AI Capabilities

A survey was conducted and sent to data management professionals to assess the findings of the data collected during the SLR that identified the AI capabilities applicable to Master Data Management (MDM) and Data Quality Management (DQM). The process involved selecting industry professionals who work with master data in an ERP context and presenting them with the framework of identified AI capabilities

with the corresponding survey questions. The survey questions applicable to the AI capabilities identified were the following:

- Which AI Capabilities do you think are the most important for organisations to adopt to improve Master Data quality in the ERP system?
- Which AI capabilities would be most easily implemented into a Master Data Management system?
- Would an AI-Enabled Master Data Management reference framework be applicable to industry professionals as guide for implementing AI tools into MDM systems?

The survey response options comprised the AI capabilities applicable to MDM functions that were extracted from the literature.

Cycle 2 – Data Analysis

The second round of data analysis involved integrating the findings from both SLRs to create an integrated framework that maps specific AI data management capabilities to different MDM functions. Thematic analysis was also used as the data analysis strategy for the second SLR. As mentioned earlier, it is an analysis method for discovering themes within data in qualitative research (Castleberry & Nolen, 2018). The result of the data analysis serves as the basis of the primary research artefact that was designed in the *Combine Findings* step in the process flow. The following section describes the evaluation of the SLR findings and the combination of the findings to create the primary research artefact, called the *AI-Enabled Master Data Management Framework*.

3.3.4 EVALUATION OF FINDINGS

This describes the evaluation of the previous two SLR findings and how the research findings combined to create the primary research artefact, namely the *AI-Enabled Master Data Management (MDM) Framework*.

The diagram below outlines the process followed to create the DSR research artefact. The *SLR Findings Evaluation* step involved evaluating the findings of the MDM Functions and AI Capabilities frameworks that were presented to data management professionals in a survey with corresponding questions to evaluate the SLR findings. During the *Combine Findings* step, the findings of both SLRs were combined to produce the primary research artefact called the *AI-Enabled MDM Framework*. This artefact was developed using the findings from the previous SLRs, namely the *MDM Functions* and the *AI Capabilities* identified from research.

The process flow concludes with the final research artefact, the *AI-Enabled MDM Framework* that categorises the MDM functions according to the appropriate AI capabilities and tools identified from literature for enhanced master data quality in ERP systems. This *AI-Enabled MDM Framework* serves as the study's primary Design Science Knowledge Contribution.

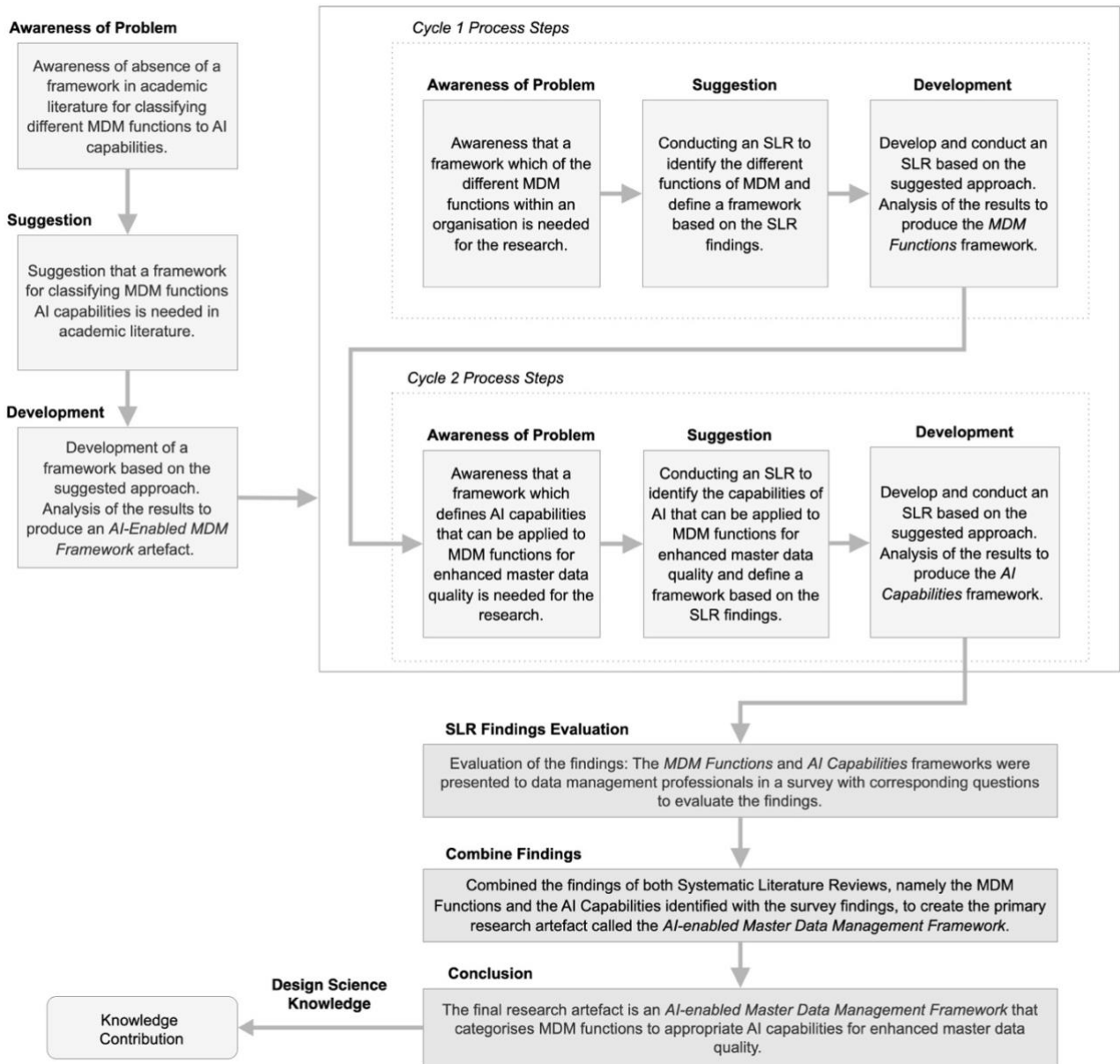


Figure 9: Research Methodology - Evaluation of Findings

The following section outlines the ethical considerations applicable to the research methodology of the study at hand.

3.4 ETHICS

The research strategy used systematic literature reviews and a survey data collection method. Special attention was paid to ensuring that the survey was conducted with participants who freely consented to participate in the study. Participants' privacy was respected, and as Cacciattolo (2015) advised, care was taken to ensure that the individuals were not coerced into participating but freely decided to do so.

Additionally, the researcher ensured that the participants did not disclose any private information which might contain sensitive organisational or project-specific data.

3.5 CONCLUSION

This chapter has defined the research methodology and the steps of the research process to ensure the research strategy is clear and feasible. A Design Science Research strategy and a pragmatist research philosophy were chosen (Goldkuhl, 2012; Hevner et al., 2004). Two Systematic Literature Reviews (SLR) are undertaken in the following chapters as the study's primary data collection method, and the findings are analysed and integrated to create an appropriate DSR artefact. The SLR findings were evaluated by conducting a research survey intended for data management professionals. After that, the final research artefact was produced and discussed.

4 SYSTEMATIC LITERATURE REVIEW – IDENTIFICATION OF THE PRIMARY MASTER DATA MANAGEMENT FUNCTIONS

The figure below shows the layout and position of Chapter 4 in relation to the rest of the research paper.

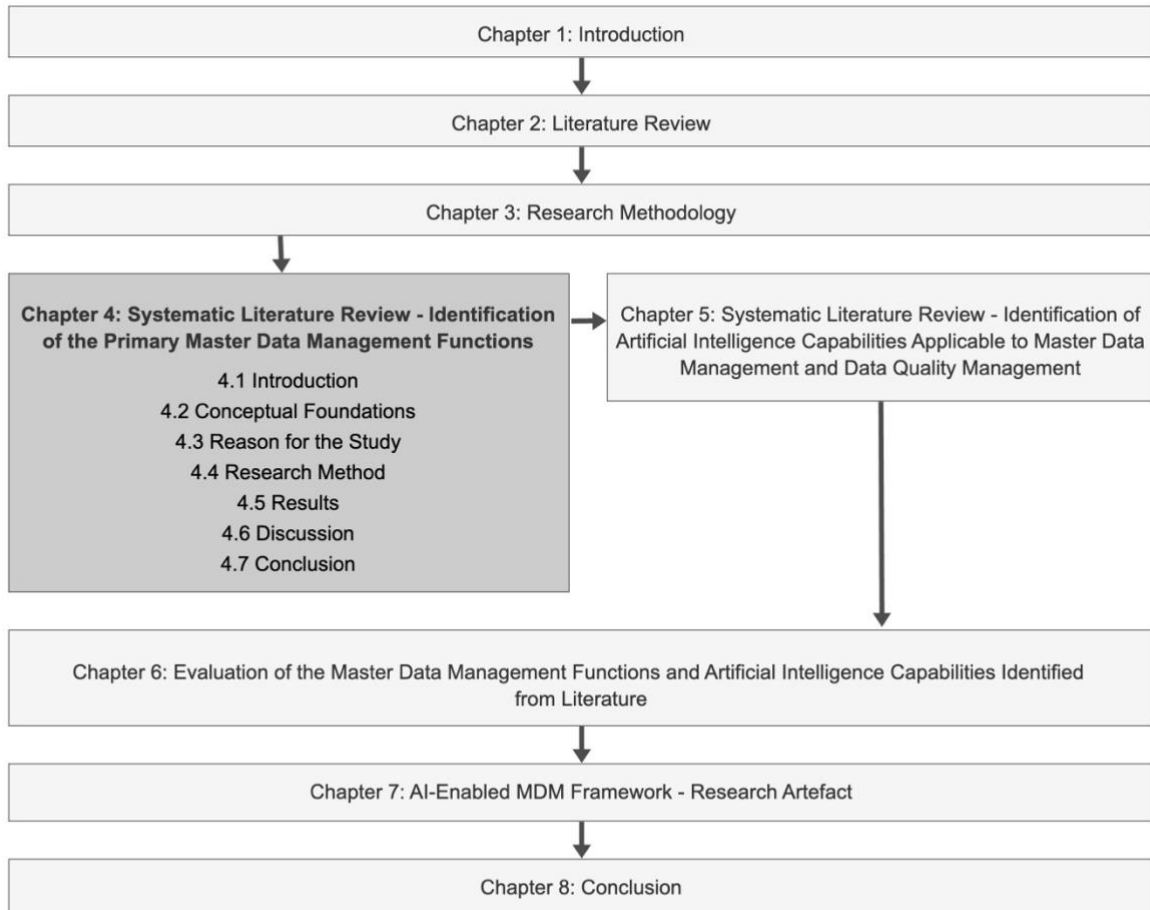


Figure 10: Structure of Research Paper - Position of Chapter 4

The following section describes the Systematic Literature Review (SLR) conducted to identify the most prominent Master Data Management (MDM) functions from current academic literature. It describes the study's conceptual foundations, objectives, research questions, and research method and concludes with the results and a discussion of the findings.

This is the first of two Systematic Literature Reviews that were undertaken to accomplish the primary research outcomes of the study, namely, producing a framework that identifies how Artificial Intelligence capabilities can be applied to Master Data Management functionality to enhance the quality of master data in an Enterprise Resource Planning system.

4.1 INTRODUCTION

The efficiency of business processes in Enterprise Resource Planning (ERP) systems relies on the quality of the master data in the ERP system's Master Data Management (MDM) tool (Malihi et al., 2016; Schäffer & Leyh, 2017). Without a high level of master data quality, an organisation faces immense difficulty in successfully executing day-to-day transactions as part of the broader business processes, and this can ultimately determine the level of business success (Schäffer & Leyh, 2017). Thus, managing the quality of master data records in an ERP system is essential for establishing effective business processes.

Organisations implement MDM tools to gain a competitive advantage by maintaining quality master data in their systems. This can save financial resources in business operations, reduce the risk of legal problems, and remove the risk of lost revenue (Szívós & Orosz, 2014). Ensuring a high level of master data quality is crucial as it defines the key business entities in the organisation and is shared across different tools (Harrison et al., 2019).

Growing volumes of organisational data drive companies to investigate the latest approaches for managing their information, such as automating data processing and maintenance activities (Schäffer & Leyh, 2017). In addition to increasing data volumes, manual maintenance activities' inefficiency is a powerful driving force for automating master data maintenance tasks (O'Farrell et al., 2013).

As mentioned, this study uses a Systematic Literature Review (SLR) approach to determine the primary functions of MDM systems in the context of ERP implementations. The following section provides the conceptual foundations of the study.

4.2 CONCEPTUAL FOUNDATIONS

This section defines the conceptual foundations referred to throughout the Systematic Literature Review. The terms Master Data Management Systems, Functions of Master Data Systems, and Enterprise Resource Planning Systems are defined from the literature to establish the fundamental terms and definitions used in the study.

4.2.1 MASTER DATA MANAGEMENT SYSTEMS

This section focuses on defining Master Data Management systems. Firstly, the definitions of Master Data and Master Data Management are discussed, and the definition of Master Data Management systems is then outlined.

Master Data

Master Data is data that defines an organisation's primary business objects. It describes the attributes of those objects and serves as the foundation for quality decision-making and improving business processes (Schäffer & Leyh, 2017). The most prominent master data objects from the literature are Material Master

Data, Product Master Data, Customer Master Data, Supplier Master Data, Financial Master Data, and Human Resources Master Data (Maedche, 2010; Schäffer & Leyh, 2017; Zhang & Ye, 2011).

Master Data Management

Master Data Management (MDM) is defined by Siren et al. (2015) as a general term encompassing all the activities that take place in an organisation to create and maintain high-quality, reliable master data that is used throughout the enterprise. Tambo (2018) states that MDM also involves aligning organisational strategy with data management strategies by developing and implementing appropriate policies. The solutions used to drive the MDM maintenance activities are technical (Siren et al., 2015; Tambo, 2018), which leads to the following section that describes the technologies and tools used for MDM in the organisation.

Master Data Management Systems

A Master Data Management (MDM) system is a distinct part of Master Data Management. It refers to the specific MDM system that maintains the organisation's master data (Grangel-González et al., 2023). Some MDM systems utilised in large enterprises are SAP Master Data Management, MDM Server for Product Information Management, IBM development, and Informatica MDM (Prokhorov & Kolesnik, 2018).

4.2.2 FUNCTIONS OF MASTER DATA MANAGEMENT SYSTEMS

Functions of Master Data Management systems centre around the functional capabilities of an MDM System (Vilminko-Heikkinen & Pekkola, 2019). These functions focus on tasks, activities, and functionality carried out on an MDM system by the users responsible for managing the organisation's master data (Vilminko-Heikkinen & Pekkola, 2019; Weber et al., 2009).

The tools mentioned in the previous section offer various functionalities for master data quality management, such as providing customisable data maintenance workflows, customising data models, cleansing, consolidating, and synchronising with data warehouses and other applications in the organisation (Prokhorov & Kolesnik, 2018). These functions are the primary research focus for this systematic literature review.

4.2.3 ENTERPRISE RESOURCE PLANNING SYSTEMS

An Enterprise Resource Planning (ERP) system is business software that allows an organisation to manage and process business information comprehensively and equips the organisation with a structured and streamlined approach to managing resources such as finance, materials, and human resources (Malihi et al., 2016).

Master data is used in the ERP business processes and is vital to the overall efficiency of business processes and decision-making (Schäffer & Leyh, 2017). In this study, Master Data Management systems are investigated in the context of Enterprise Resource Planning systems.

4.3 REASON FOR THE STUDY

4.3.1 OBJECTIVE

This SLR aims to identify which functions of Master Data Management (MDM) are most prominent in the academic literature. The study seeks to address the gap in the current scholarly literature on identifying the primary functions of an MDM system within the context of Enterprise Resource Planning (ERP) systems. It aims to advance academic literature by producing the MDM Functions Framework. This artefact will assist future studies, such as research that focuses on identifying how new and upcoming innovations in the field of data management will affect the MDM functions identified in this SLR. This will provide researchers and industry professionals with a reference framework on which to base further research in the field of MDM.

4.3.2 RESEARCH QUESTION

What are the primary functions of a Master Data Management system in an organisation utilising Enterprise Resource Planning Systems?

4.4 RESEARCH METHOD

The research method used for identifying the primary functions of a Master Data Management (MDM) system is a Systematic Literature Review (SLR), as defined by Boland et al. (2017). It is described as a structured and systematic method of discovering, evaluating, and combining literature related to the research objective. The primary goal of the SLR is to produce verifiable information that serves as evidence about the research objective.

4.4.1 SEARCH STRATEGY

Sources relevant to the study at hand have been extracted by using the search string defined below to select journal articles that can help in answering the SLR research question: (*“Master Data” OR “Master Data Management”*) AND (*“Enterprise Resource Planning” OR ERP*).

The search string defined above was used to search for research material to use in this study for the following databases:

- IEEE Xplore Digital Library,
- ACM Digital Library,

- ScienceDirect, and
- Scopus.

4.4.2 SELECTION CRITERIA

Predefined selection criteria have been applied to the journal articles to extract the most relevant sources for use in the SLR. The inclusion and exclusion criteria specified to refine the selection are outlined below.

Inclusion Criteria

1. The article is accessible and written in English.
2. The document type is a journal, review, research, or conference paper.
3. The article discusses Master Data Management (MDM) in Enterprise Resource Planning (ERP) systems.

Exclusion Criteria

1. Duplicate research articles are excluded.
2. Research that does not discuss Master Data Management within the context of Enterprise Resource Planning (ERP) Systems.
3. Sources that do not define any specific Master Data Management components or functions.

4.4.3 SCREENING AND SELECTION

As part of the SLR screening and selection step, a filtering process based on the inclusion and exclusion criteria was applied to ensure high-quality academic journal articles were selected. The results of the initial database search with the defined search string are displayed in the table below. These results are limited to articles written in English and the specific document types mentioned in the inclusion criteria above.

Database	Results
IEEE Xplore Digital Library	37
ACM Digital Library	46
ScienceDirect	265
Scopus	71
Total	419

Table 1: SLR Databases

After identifying the initial articles using the defined search string, the filtering removed the duplicate articles. The abstracts of the remaining articles were then screened for adequacy, and irrelevant articles were excluded. After that, full-text articles were assessed against the defined inclusion and exclusion

criteria. All studies deemed adequate for answering the research question were included in the literature review.

4.4.4 PRISMA FLOWCHART

A Prisma Flowchart graphically depicts the record identification, screening, assessment for eligibility, and inclusion of all the journal articles identified. The results of each of the filtration steps are displayed as described previously.

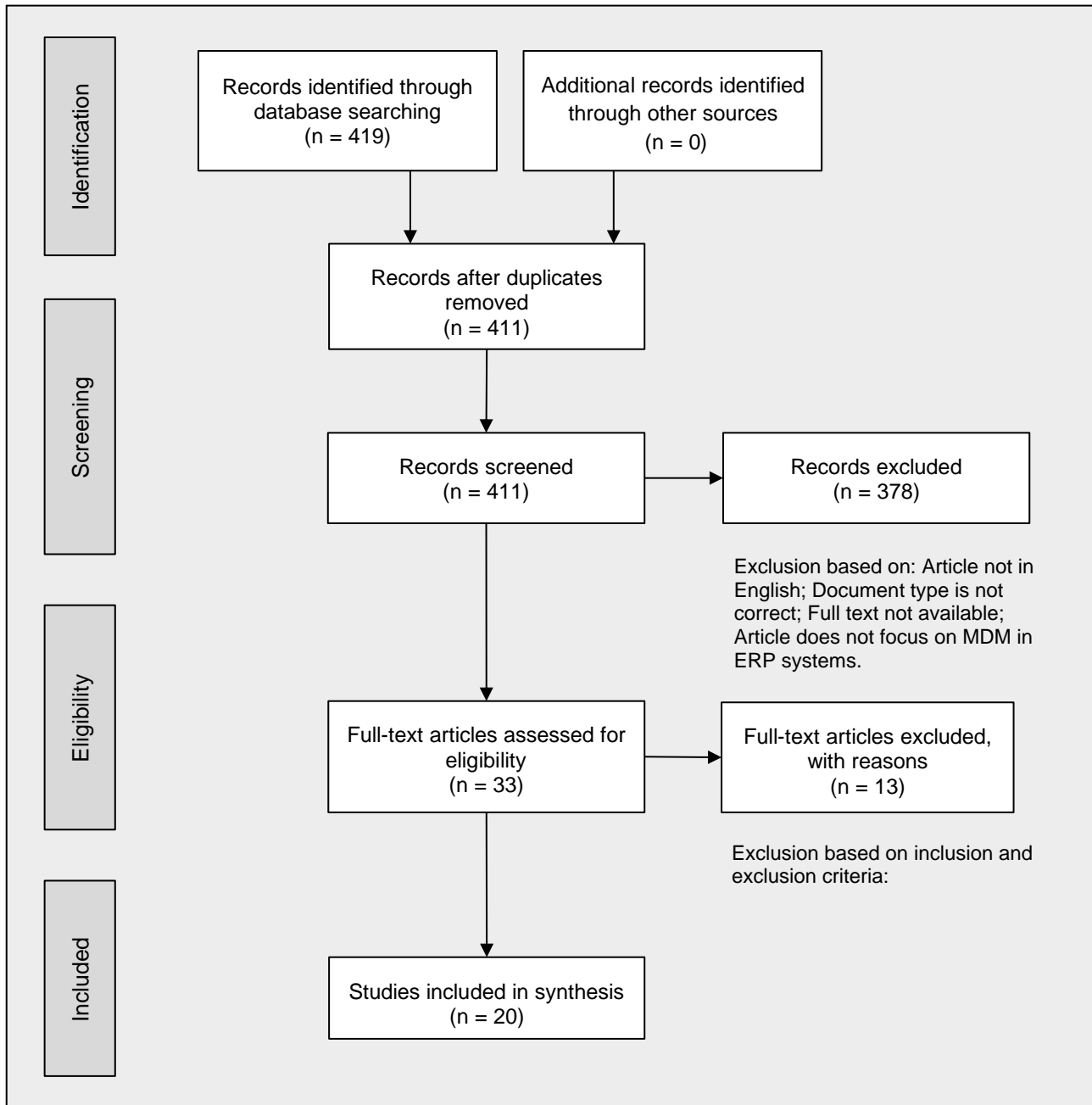


Figure 11: PRISMA Flowchart – Adapted from Moher et al. (2009).

4.4.5 QUALITY ASSESSMENT

A set of predefined quality assessment criteria has been defined to assess the academic quality of the journal articles included in this literature review. Quality assessment questions have been applied to each source to determine its quality and relevance to answering the research question. The following list contains the quality assessment questions:

- Question 1: Is the article primarily focused on Master Data Management?
- Question 2: Does the article adequately define any Master Data Management Functions?
- Question 3: Is Master Data Management described within the context of Enterprise Resource Planning systems?

The following table contains the answer code and description for each respective quality assessment question as defined:

Question	Answer Code	Description
1	Y (Yes)	The article is primarily focused on Master Data Management.
	P (Partially)	The article does not primarily focus on Master Data Management but does mention it within the broader topic.
	N (No)	The article does not mention Master Data Management.
2	Y (Yes)	The article adequately defines multiple Master Data Management Functions.
	P (Partially)	The article indirectly defines some Master Data Management Functions.
	N (No)	The article does not define any Master Data Management Functions.
3	Y (Yes)	Master Data Management is described within the context of Enterprise Resource Planning systems.
	P (Partially)	Master Data Management is described in Enterprise Resource Planning and other information systems.
	N (No)	Master Data Management is not described within the context of Enterprise Resource Planning systems.

Table 2: Quality Assessment Codes

The scoring method for the quality assessment is carried out as per Kitchenham et al. (2009), whereby the following quality scores are applied for the relevant answer codes per source research article:

Y (Yes) = 1, P (Partial) = 0.5, and N (No) = 0, or Unknown (unspecified).

The following tables outline the list of sources used in the study and the evaluation scores of the quality assessment criteria as applied to each of the sources included in the study. The first table contains the list of sources in the SLR, with information about the author, publication year, title, and the journal in which the source was published. The second table contains the quality evaluation scores for each source.

ID	Author	Title	Journal Publication	Database	Year
S1	(Dittrich et al., 2009)	ERP Customization as Software Engineering: Knowledge Sharing and Cooperation	IEEE Software	IEEE	2009
S2	(Zhang & Ye, 2011)	Data Maintenance and Control Strategy of Group Management	2011 Second International Conference on Innovations in Bio-inspired Computing and Applications	IEEE	2011
S3	(Knolmayer & Röthlin, 2006)	Quality of material master data and its effect on the usefulness of distributed ERP systems	Lecture Notes in Computer Science	Scopus	2006
S4	(Maedche, 2010)	An ERP-centric master data management approach	16th Americas Conference on Information Systems 2010, AMCIS 2010	Scopus	2010
S5	(Knapp & Hasibether, 2011)	Material master data quality	2011 17th International Conference on Concurrent Enterprising, ICE 2011 - Conference Proceedings	Scopus	2011
S6	(O'Farrell et al., 2013)	Ameliorating ERP workflow using a sociomaterial lens	19th Americas Conference on Information Systems, AMCIS 2013 - Hyperconnected World: Anything, Anywhere, Anytime	Scopus	2013
S7	(Szívós & Orosz, 2014)	The role of data authentication and security in the audit of financial statements	Acta Polytechnica Hungarica	Scopus	2014
S8	(Siren et al., 2015)	The requirements and needs of global data usage in product lifecycle management	Lecture Notes in Computer Science	Scopus	2015
S9	(Malihi et al., 2016)	Customisation of master data elements in ERP systems aligned with business goals	International Journal of Information Systems and Change Management	Scopus	2016
S10	(Schäffer & Leyh, 2017)	Master data quality in the era of digitization - toward inter-organizational master data quality in value networks: A problem identification	Lecture Notes in Business Information Processing	Scopus	2017
S11	(Zong et al., 2017)	Identification of approximately duplicate material records in ERP systems	Enterprise Information Systems	Scopus	2017
S12	(Tambo, 2018)	PLM or ERP, the chicken or the EGG: Towards an enterprise level master data management approach for improving innovation and supply chain collaboration	Towards Sustainable Technologies and Innovation - Proceedings of the 27th Annual Conference of the International Association for Management of Technology, IAMOT 2018	Scopus	2018
S13	(Hannila et al., 2019)	Product and supply chain related data, processes and information systems for product portfolio management	International Journal of Product Lifecycle Management	Scopus	2019
S14	(Ender Hancioglu, 2020)	A logic to handle, define and process item characteristics in ERP	CEUR Workshop Proceedings	Scopus	2020
S15	(Grangel-González et al., 2023)	LIS: A Knowledge Graph-Based Line Information System	Lecture Notes in Computer Science	Scopus	2023
S16	(Prokhorov & Kolesnik, 2018)	Development of a master data consolidation system model (on the example of the banking sector)	Procedia Computer Science	Science Direct	2018
S17	(Vilminko-Heikkinen & Pekkola, 2019)	Changes in roles, responsibilities and ownership in organizing master data management	International Journal of Information Management	Science Direct	2019
S18	(Weber et al., 2009)	One Size Does Not Fit All---A Contingency Approach to Data Governance	Data and Information Quality	ACM	2009
S19	(Harrison et al., 2019)	The Data Firehose and AI in Government: Why Data Management is a Key to Value and Ethics	Proceedings of the 20th Annual International Conference on Digital Government Research	ACM	2019
S20	(Tong et al., 2020)	The Realization for Automated Warehouse Based on the Integration of ERP and WMS	Proceedings of the 2020 the 7th International Conference on Automation and Logistics (ICAL)	ACM	2020

Table 3: Sources Included in SLR

The following table consists of the quality assessment scores for each literature paper in the Systematic Literature Review.

ID	Q1	Q2	Q3	Total Score
S1	N	P	Y	1.5
S2	P	Y	Y	2.5
S3	P	Y	Y	2.5
S4	Y	Y	Y	3
S5	P	Y	Y	2.5
S6	P	Y	Y	2.5
S7	P	Y	Y	2.5
S8	Y	P	Y	2.5
S9	N	P	N	0.5
S10	P	P	P	1.5
S11	N	P	N	0.5
S12	P	N	Y	1.5
S13	N	N	P	0.5
S14	P	P	N	1
S15	N	P	Y	1.5
S16	Y	Y	Y	3
S17	Y	P	Y	2.5
S18	Y	P	Y	2.5
S19	N	P	Y	1.5
S20	N	N	P	0.5

Table 4: Quality Evaluation Scores

4.4.6 DATA EXTRACTION

The data extraction process consists of identifying the most prominent functions of Master Data Management (MDM) mentioned in the papers included in the SLR. Individual sections from the text of each research paper in the literature reviewed were examined to identify the key codes and themes and subsequently derive the MDM functions from each paper. Additional dimensions extracted from each paper were the industries specified in the text and the master data objects mentioned in each paper. This was carried out to provide further insight into the type of master data being managed by the MDM system and the industries in which the organisations using MDM systems are situated.

4.4.7 DATA ANALYSIS

Thematic analysis is the data analysis method employed in this SLR, as defined by Braun and Clarke (2006). Following the thematic analysis approach, the key themes and codes within the literature texts included in the study were identified to produce a final set of MDM functions extracted from the literature

to answer the research question. The key themes and codes extracted from the text were MDM functional tasks and activities, along with the industry and the master data objects mentioned in the paper. The MDM functional tasks and activities extracted varied from master data maintenance and modification to data quality improvements such as master data duplicate detection and validation processes.

To determine which MDM functions and tasks should be grouped together, the most prominent wording was selected as the functional classification. The MDM functions from different papers that had the same meaning but were described in different words were grouped together. The original and categorised list of MDM functions are shown in the table in section 4.5.2.

4.5 RESULTS

The results from the data analysis described in the previous section consist of a list of crucial Master Data Management (MDM) functions identified from each source. The study was conducted using the carefully screened and selected literature, as described in the research methodology in section 4.4.

In this section, the MDM research trends are reviewed. The initial MDM functions are identified in the literature, and appropriate categories are listed with corresponding sources linked to each MDM function. The categories are based on descriptions of MDM functions from literature with similar meanings. The categorised framework called the MDM Functions Framework, is presented in section 4.5.3.

This section concludes with an overview of the industries using MDM systems and the primary master data objects found in the literature.

4.5.1 MASTER DATA MANAGEMENT RESEARCH TRENDS

The graph below depicts the Master Data Management (MDM) research trends over the past eighteen years. Though the graph consists only of the papers in this SLR, more research on MDM topics is needed.

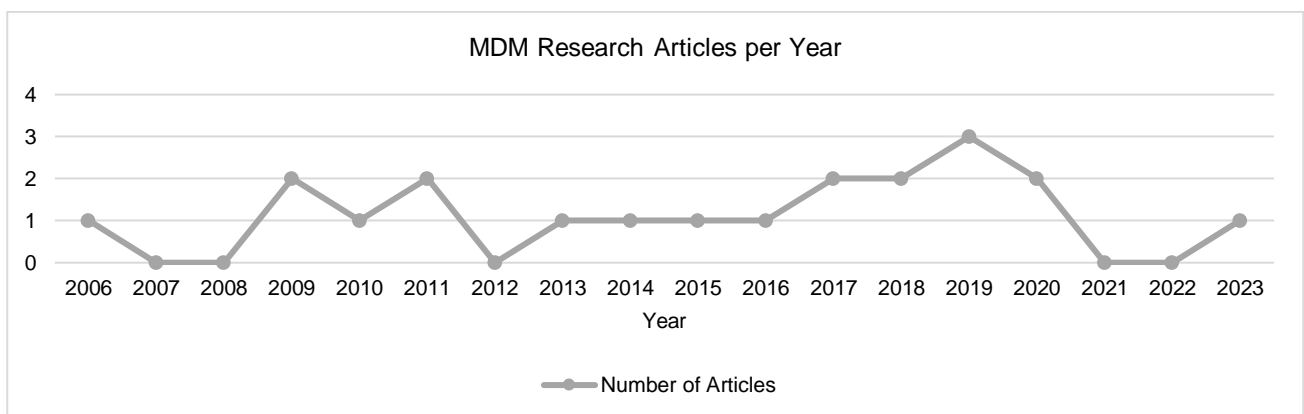


Figure 12: MDM Research Trends

4.5.2 MASTER DATA MANAGEMENT FUNCTIONS

The following table comprises the Master Data Management (MDM) functions and function categories identified from the literature. The corresponding sources for each function identified are also listed for full transparency. Some of the terminology used in the literature to describe the MDM functions has a similar meaning to other MDM functions in the list. For this reason, these MDM functions with the same meaning are grouped to categorise the MDM functions framework.

The first column in the table below contains the category assigned for each MDM function in the second column. For example, the MDM Functions '*Master Data Workflow Processes*', '*Master Data Quality Workflow*', '*Master Data Workflows*', '*Master Data Change Process*', '*Master Data Management Processes*', and '*Collaborative and Customisable Workflows*' extracted from the literature was categorised as Master Data Workflows. In the same way, the other MDM functions identified from the literature have been categorised based on the similarity in meaning of the MDM functions.

MDM Function Category	MDM Functions	References
Master Data Model and Workflow Customisation	Master Data Workflow Processes	(Zhang & Ye, 2011), (O'Farrell et al., 2013)
	Master Data Quality Workflow	(Maedche, 2010)
	Master Data Workflows	(O'Farrell et al., 2013), (Vilminko-Heikkinen & Pekkola, 2019)
	Master Data Change Process	(Szívós & Orosz, 2014), (Prokhorov & Kolesnik, 2018)
	Master Data Management Processes	(Weber et al., 2009)
	Collaborative and Customisable Workflows	(Prokhorov & Kolesnik, 2018)
	Master Data Model Customisation	(Dittrich et al., 2009), (Prokhorov & Kolesnik, 2018)
Master Data Import and Distribution	Master Data Distribution	(Knolmayer & Röthlin, 2006), (Maedche, 2010), (Prokhorov & Kolesnik, 2018), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019)
	Master Data Sharing	(Grangel-González et al., 2023), (Siren et al., 2015), (Maedche, 2010),
	Master Data Publication	(Prokhorov & Kolesnik, 2018)
	Master Data Exchange	(Schäffer & Leyh, 2017)
	Master Data Replication	(Maedche, 2010), (Siren et al., 2015)

MDM Function Category	MDM Functions	References
	Master Data Synchronisation	(Prokhorov & Kolesnik, 2018)
	Master Data Collection	(Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019)
	Master Data Import	(Dittrich et al., 2009), (Zhang & Ye, 2011)
Master Data Creation and Maintenance	Master Data Maintenance	(Knolmayer & Röthlin, 2006), (Maedche, 2010), (Zhang & Ye, 2011), (O'Farrell et al., 2013), (Szívós & Orosz, 2014), (Malihi et al., 2016), (Schäffer & Leyh, 2017), (Tambo, 2018), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009)
	Master Data Processing	(Maedche, 2010), (Schäffer & Leyh, 2017)
	Master Data Creation	(Maedche, 2010), (Knapp & Hasibether, 2011), (Zhang & Ye, 2011), (Knolmayer & Röthlin, 2006), (O'Farrell et al., 2013), (Szívós & Orosz, 2014), (Schäffer & Leyh, 2017), (Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009)
	Master Data Deletion	(Zhang & Ye, 2011), (Szívós & Orosz, 2014), (Weber et al., 2009)
	Master Data Modifications	(Zhang & Ye, 2011), (Knapp & Hasibether, 2011), (Vilminko-Heikkinen & Pekkola, 2019)
	Master Data Updates	(Knolmayer & Röthlin, 2006), (Szívós & Orosz, 2014), (Tambo, 2018)
	Master Data Mass Maintenance	(O'Farrell et al., 2013)
	Master Data Mass Change Processes	(O'Farrell et al., 2013)
	Master Data Search	(Szívós & Orosz, 2014)
	Master Data Analysis and Reporting	Master Data Analysis and Reporting
Master Data Duplicate Detection and De-duplication	Master Data Duplication Detection	(Knolmayer & Röthlin, 2006), (Knapp & Hasibether, 2011), (Prokhorov & Kolesnik, 2018)
	Master Data De-duplication	(Maedche, 2010), (Zong et al., 2017), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019), (Harrison et al., 2019)

MDM Function Category	MDM Functions	References
	Eliminating Duplicate Master Data Records	(Prokhorov & Kolesnik, 2018)
	Master Data Deduplication and Merging	(Szívós & Orosz, 2014)
	Removing Redundant Master Data	(Siren et al., 2015)
Master Data Validation and Business Rules	Master Data Validation and Checks	(Knolmayer & Röthlin, 2006), (Knapp & Hasibether, 2011), (Szívós & Orosz, 2014), (Schäffer & Leyh, 2017), (Grangel-González et al., 2023), (Prokhorov & Kolesnik, 2018)
	Business Rules for MDM	(Malihi et al., 2016)
Master Data Enrichment and Cleansing	Master Data Enrichment	(Maedche, 2010), (Prokhorov & Kolesnik, 2018)
	Master Data Cleansing	(Zhang & Ye, 2011), (Maedche, 2010), (Knapp & Hasibether, 2011), (Prokhorov & Kolesnik, 2018)
	Master Data Corrections	(Knapp & Hasibether, 2011)
	Master Data Transformation	(Maedche, 2010)
	Master Data Harmonisation	(Maedche, 2010), (Knapp & Hasibether, 2011), (Grangel-González et al., 2023), (Prokhorov & Kolesnik, 2018)
Master Data Classification and Taxonomy Management	Material Master Data Classification	(Knapp & Hasibether, 2011), (Zong et al., 2017)
	Product Master Data Classification	(Ender Hancioglu, 2020)
	Data Classification	(Prokhorov & Kolesnik, 2018)
	Master Data Taxonomy Management	(Prokhorov & Kolesnik, 2018)
Master Data Consolidation	Master Data Consolidation	(Maedche, 2010), (Szívós & Orosz, 2014), (Siren et al., 2015), (Prokhorov & Kolesnik, 2018), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019)
	Master Data Standardisation	(Prokhorov & Kolesnik, 2018)
	Master Data Access Management	(Prokhorov & Kolesnik, 2018), (Weber et al., 2009) (Maedche, 2010), (Siren et al.,

MDM Function Category	MDM Functions	References
Master Data Roles and Authorisation Management		2015), (Vilminko-Heikkinen & Pekkola, 2019)
	Master Data Roles Management	(Dittrich et al., 2009), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009),
	Prohibiting Local Changes of Global Master Data Objects	(Maedche, 2010)
	Master Data Authorisation Management	(Szívós & Orosz, 2014), (Siren et al., 2015), (Vilminko-Heikkinen & Pekkola, 2019)
	Master Data Security	(Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009)

Table 5: MDM Functions Identified from Literature

4.5.3 MASTER DATA MANAGEMENT FUNCTIONS FRAMEWORK

The MDM functions identified and classified in the previous section have been combined to produce the primary findings of this SLR in the form of the MDM Functions Framework. The framework contains the MDM Function, the definition of each function from the literature, and the references to the literature papers from which the functions have been identified and described.

MDM Function	MDM Function Definition	References
Master Data Model and Workflow Customising	Master data workflow customising refers to collaborative and customisable change process workflow procedures for creating and maintaining master data according to evolving business rules and organisational requirements. Master data model customising refers to making master data model adjustments and changes according to the business requirements.	(Zhang & Ye, 2011), (O'Farrell et al., 2013), (Maedche, 2010), (Vilminko-Heikkinen & Pekkola, 2019), (Szívós & Orosz, 2014), (Prokhorov & Kolesnik, 2018), (Weber et al., 2009), (Dittrich et al., 2009)
Master Data Import and Distribution	Master data importing refers to collecting and inserting master data into the MDM system. Master data sharing and distribution refers to the distribution of master data to other target IT systems in the organisation, such as operational and analytical applications, to ensure data integrity. It is also called an information usage control mechanism that allows the master data to be reused by sharing it between applications.	(Knolmayer & Röthlin, 2006), (Maedche, 2010), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019) (Grangel-González et al., 2023), (Siren et al., 2015), (Schäffer & Leyh, 2017), (Dittrich et al., 2009), (Zhang & Ye, 2011)
Master Data Creation and Maintenance	Master data creation and maintenance refers to the creation, modification, search, and deletion functions within a Master Data Management system. Master Data Creation defines a new master data record in the MDM application. Master data modifications refer to modifications or updates to existing master data in the MDM system. Master data search refers to searching for master data records in the MDM system. Master data deletion refers to removing master data from the MDM system. Master data single and mass maintenance refers to maintaining or making changes to one or multiple numbers of master data records in an MDM system.	(Maedche, 2010), (Knapp & Hasibether, 2011), (Zhang & Ye, 2011), (Knolmayer & Röthlin, 2006), (O'Farrell et al., 2013), (Szívós & Orosz, 2014), (Schäffer & Leyh, 2017), (Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009), (Tambo, 2018), (Malihi et al., 2016), (Prokhorov & Kolesnik, 2018)
Master Data Analysis and Reporting	Master data analysis and reporting refers to the functionality of the MDM system to support analysis and reporting on master data objects, such as supplier and customer master data.	(Prokhorov & Kolesnik, 2018)
Master Data Duplicate Detection and De-duplication	Master data duplicate detection and de-duplication refers to the functionality of the MDM system to eliminate duplicate records by searching for duplicate master data that has been created and removing those records from the system.	(Knolmayer & Röthlin, 2006), (Knapp & Hasibether, 2011), (Prokhorov & Kolesnik, 2018), (Maedche, 2010), (Zong et al., 2017), (Vilminko-Heikkinen & Pekkola, 2019), (Harrison et al., 2019), (Szívós & Orosz, 2014), (Siren et al., 2015)
Master Data Validation and Business Rules	Master data validation and business rules refer to the data input validation processes based on the organisation's business rules for master data creation in the MDM system. These are validation rules such as blocking users from entering invalid values, setting specific data fields as mandatory and implementing customisable data validation checks.	(Knolmayer & Röthlin, 2006), (Knapp & Hasibether, 2011), (Szívós & Orosz, 2014), (Schäffer & Leyh, 2017), (Grangel-González et al., 2023), (Prokhorov & Kolesnik, 2018), (Malihi et al., 2016)
Master Data Enrichment and Cleansing	Master data enrichment refers to enriching the data with necessary information, and cleansing refers to cleaning and harmonising the master data in the system according to organisational standards.	(Maedche, 2010), (Prokhorov & Kolesnik, 2018), (Zhang & Ye, 2011), (Knapp & Hasibether, 2011), (Grangel-González et al., 2023)
Master Data Classification and Taxonomy Management	Master data classification and taxonomy management refers to the classification of master data in line with specific criteria and organisational taxonomies. Classification plays an important role in managing product master data. Classifying similar parts allows the user to identify similarities in master data characteristic values easily, and it simplifies managing product descriptions for multiple languages in the system.	(Knapp & Hasibether, 2011), (Zong et al., 2017), (Ender Hancioglu, 2020), (Prokhorov & Kolesnik, 2018)
Master Data Consolidation	Master data consolidation refers to combining different versions of the same data element from various information systems into a single version in a central system, referred to as the	(Maedche, 2010), (Szívós & Orosz, 2014), (Siren et al., 2015), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019)

MDM Function	MDM Function Definition	References
	"single version of the truth", also called a "golden record". The golden record is then synchronised to the appropriate applications in the organisation.	
Master Data Roles and Authorisation Management	Master data roles and authorisation management refers to the definition, configuration, and assignment of access rights in the MDM system according to user roles, which ensures that authorised users gain only the necessary access to data in the system according to their role specifications.	(Prokhorov & Kolesnik, 2018), (Weber et al., 2009), (Maedche, 2010), (Siren et al., 2015), (Vilminko-Heikkinen & Pekkola, 2019), (Dittrich et al., 2009), (Szívós & Orosz, 2014)

Table 6: MDM Functions Framework

4.5.4 MASTER DATA MANAGEMENT FUNCTIONS LITERATURE TRENDS

The following graph depicts how many of the research papers in each type of Master Data Management function were mentioned.

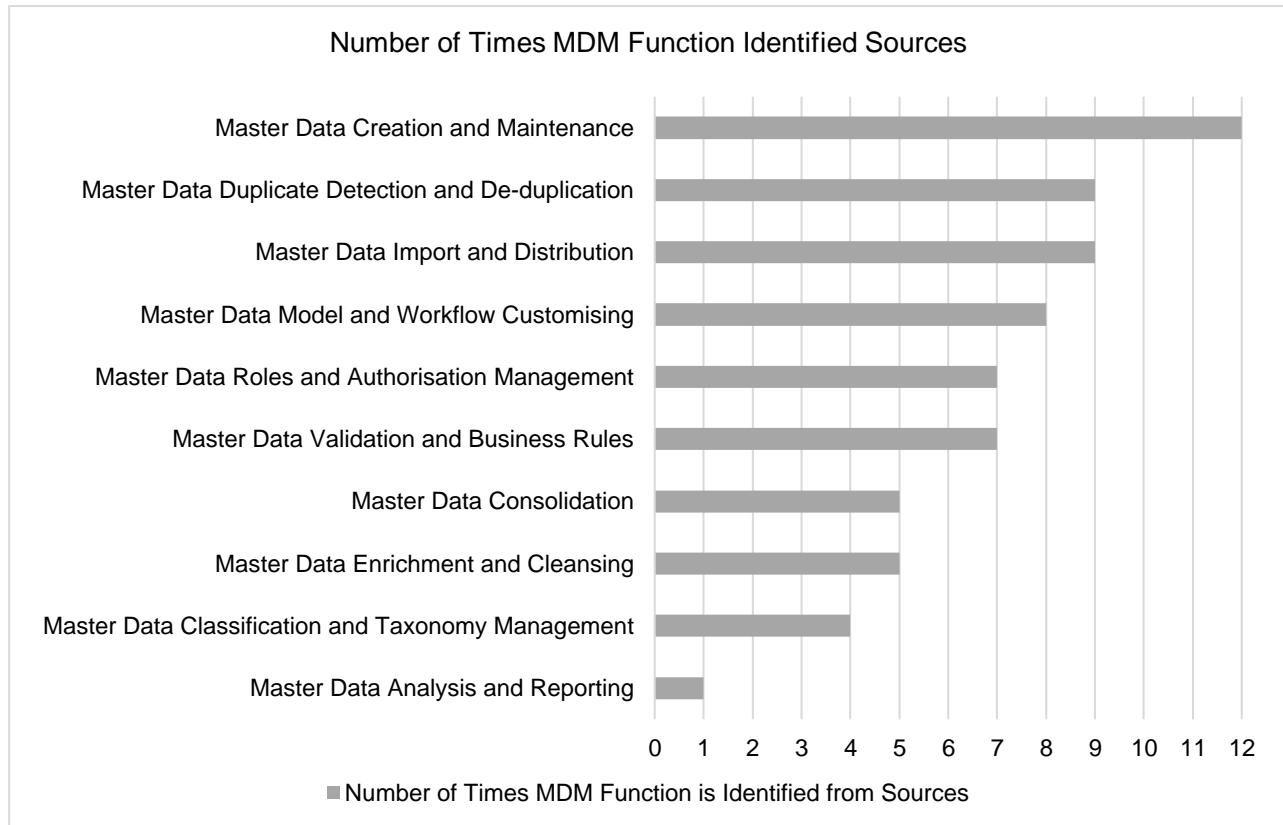


Figure 13: MDM Functions Literature Trends

This provides a visualisation of the MDM functions that are most prominent in the literature, namely Master Data Creation and Maintenance, Master Data Duplication Detection and De-duplication and Master Data Import and Distribution. The less prominent MDM functions were not excluded from the final MDM Functions Framework to ensure that the framework is comprehensive and includes all the functions identified in the literature.

4.5.5 INDUSTRIES

Master Data Management tools can be found in ERP implementations for various industries. The primary industries utilising MDM systems have been extracted from the literature and are depicted in the graph below.

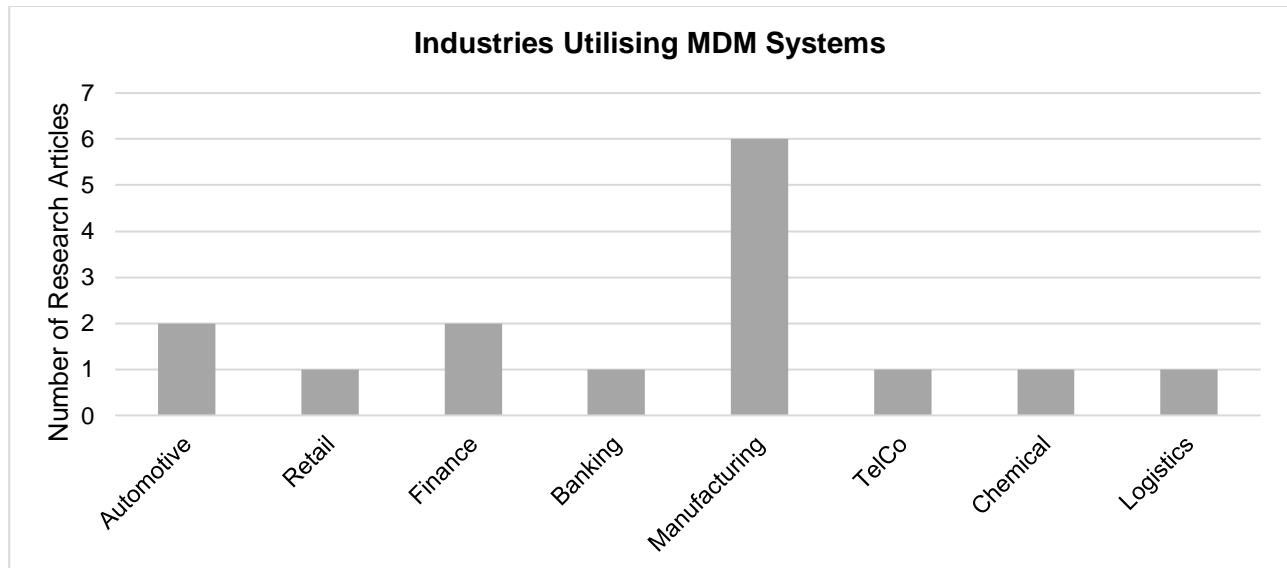


Figure 14: Industries Utilising MDM Systems

From the literature in the review, industries such as Manufacturing, Automotive, and Finance stood out as some of the primary organisational industries using MDM systems in the context of ERP software. These results are based on the selected literature to answer the research question and cannot be used as definite evidence for industry trends. The industries were extracted to provide additional insight into the type of master data managed by the MDM systems.

4.5.6 MASTER DATA OBJECTS

The final dimension extracted from the literature review was the Master Data Objects mentioned in each paper. The following table lists the Master Data Objects with the corresponding references to the literature papers in which they were mentioned.

Master Data Objects	References
Material Master Data	(Zhang & Ye, 2011), (Knapp & Hasibether, 2011), (O'Farrell et al., 2013), (Malihi et al., 2016), (Schäffer & Leyh, 2017), (Zong et al., 2017), (Prokhorov & Kolesnik, 2018)
Product Master Data	(Zhang & Ye, 2011), (Maedche, 2010), (Knapp & Hasibether, 2011), (Malihi et al., 2016), (Schäffer & Leyh, 2017), (Hannila et al., 2019), (Ender Hancioglu, 2020), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009)
Financial Master Data	(Zhang & Ye, 2011), (Szívós & Orosz, 2014)

Master Data Objects	References
Ledger Accounts Data	(Szívós & Orosz, 2014)
Accounting Master Data	(Maedche, 2010)
Customer Master Data	(Maedche, 2010), (Knapp & Hasibether, 2011), (O'Farrell et al., 2013), (Szívós & Orosz, 2014), (Malihi et al., 2016), (Schäffer & Leyh, 2017), (Hannila et al., 2019), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009), (Harrison et al., 2019)
Supplier Master Data	(Maedche, 2010), (Malihi et al., 2016), (Prokhorov & Kolesnik, 2018)
Vendor Master Data	(Knapp & Hasibether, 2011), (O'Farrell et al., 2013), (Szívós & Orosz, 2014), (Schäffer & Leyh, 2017), (Hannila et al., 2019), (Weber et al., 2009)
Business Partner	(Harrison et al., 2019)
Human Resources Master Data	(Zhang & Ye, 2011)
Employee Master Data	(Malihi et al., 2016), (Vilminko-Heikkinen & Pekkola, 2019), (Harrison et al., 2019)
Location Master Data	(Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019)

Table 7: Master Data Objects

To further analyse the Master Data Objects referenced in each paper, the data from the above table is visualised as a graph below.

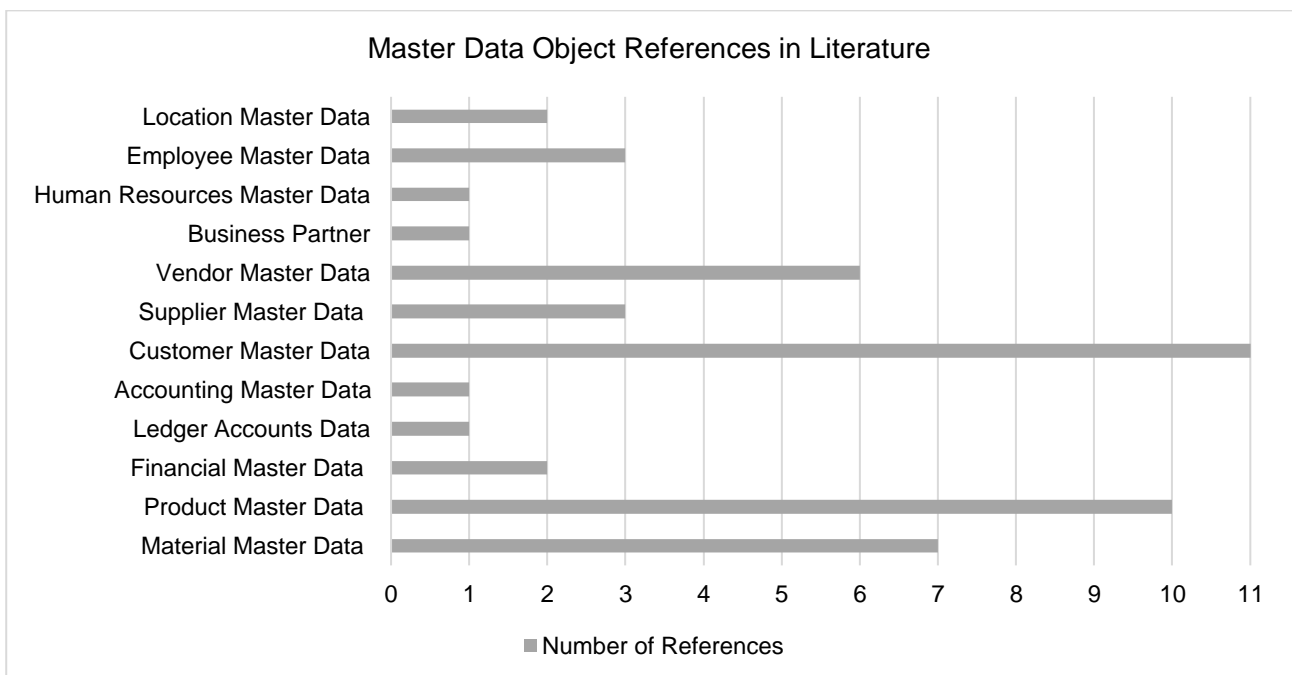


Figure 15: Master Data Objects

The results from the graph suggest that the management of more prominent master data objects like material, product, customer, and vendor master data is of primary importance in organisations that have implemented and are making use of an MDM system.

4.6 DISCUSSION

This systematic literature review focuses on identifying an extensive list of MDM functions from which to create the MDM Functions Framework. The MDM functions found in the literature were categorised according to similar functionality, and the table in section 4.5.3 contains the final results of the data analysis.

Numerous MDM functions were identified from the research, such as *Master Data Model and Workflow Customising*, *Master Data Creation and Maintenance*, *Master Data Duplicate Detection and De-duplication*, *Master Data Classification and Taxonomy Management*, and more. From the literature, some of these MDM functions are mentioned to be particularly problematic for organisations in managing their master data. One such problem area is a lack of adaptability in the MDM function *Master Data Workflow Customising*, which was brought to attention by O'Farrell et al. (2013). Inefficiencies in *Master Data Creation* and *Master Data Maintenance* were also highlighted by O'Farrell et al. (2013) as being troublesome.

Furthermore, *Master Data Roles and Authorisation Management* can lead to poor master data quality when user roles and responsibilities are inadequately managed for this MDM function (Vilminko-Heikkinen & Pekkola, 2019). In the paper by Vilminko-Heikkinen and Pekkola (2019), duplicate data was identified as one of the biggest factors leading to poor data quality. Thus, great attention must be paid to the *Master Data Duplicate Detection and De-duplication* MDM function. Another important MDM function identified from the literature is the *Master Data Classification and Taxonomy Management*. This function is mostly applicable for managing material and product master data, where the taxonomy and classification play a major role in ensuring that manufacturing parts are correctly classified (Ender Hancioglu, 2020; Zong et al., 2017). A lack of common master data taxonomy can lead to inconsistent material master data classification. It can cause multiple data deficiencies, thus requiring appropriate governance (Knapp & Hasibether, 2011).

Szívós and Orosz (2014) mention that to improve the quality of master data, ways of automating the validation process must be investigated to reduce the so-called "human factor". Harrison et al. (2019) state that adequate MDM systems are a fundamental building block for enabling the usage of Artificial Intelligence in an organisation and that investments must be made for improved organisational data governance.

4.7 CONCLUSION

This study has focused on identifying the primary Master Data Management (MDM) functions for an organisation utilising Enterprise Resource Planning (ERP) systems. Each identified MDM function was listed, and appropriate categorisation was applied to groups of similar MDM functions. Furthermore, these

functions were defined according to the descriptions from the literature from which they were identified and extracted.

The findings from the literature used in the study assist in identifying these MDM functions and also highlight some of the MDM functions that were particularly problematic. The MDM functions *Master Data Workflow Customising*, *Master Data Creation and Maintenance*, *Master Data Roles and Authorisation Management*, *Master Data Duplicate Detection and De-duplication*, *Master Data Classification* and *Taxonomy Management* were identified as some of the problematic areas.

The study also extracted the industries that use MDM systems and the master data objects that are being managed with these tools. The industries mentioned in the literature that use MDM functions are the automotive, retail, finance, banking, manufacturing, telco, chemical, and logistics industries. The master data objects identified from the studies included in the review were Material, Product, Financial, Ledger Accounts, Accounting, Customer, Supplier, Vendor, Business Partner, Human Resources, Employee, and Location Master Data.

The study has achieved its primary objective of identifying the MDM functions for an organisation utilising ERP systems. Additional research is necessary to establish which MDM functions can be enhanced by implementing tools such as Artificial Intelligence capabilities to help improve the master data quality in these MDM systems and the subsequent ERP systems.

5 SYSTEMATIC LITERATURE REVIEW – IDENTIFICATION OF ARTIFICIAL INTELLIGENCE CAPABILITIES APPLICABLE TO MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT

The figure below shows the layout and position of Chapter 5 in relation to the rest of the research paper.

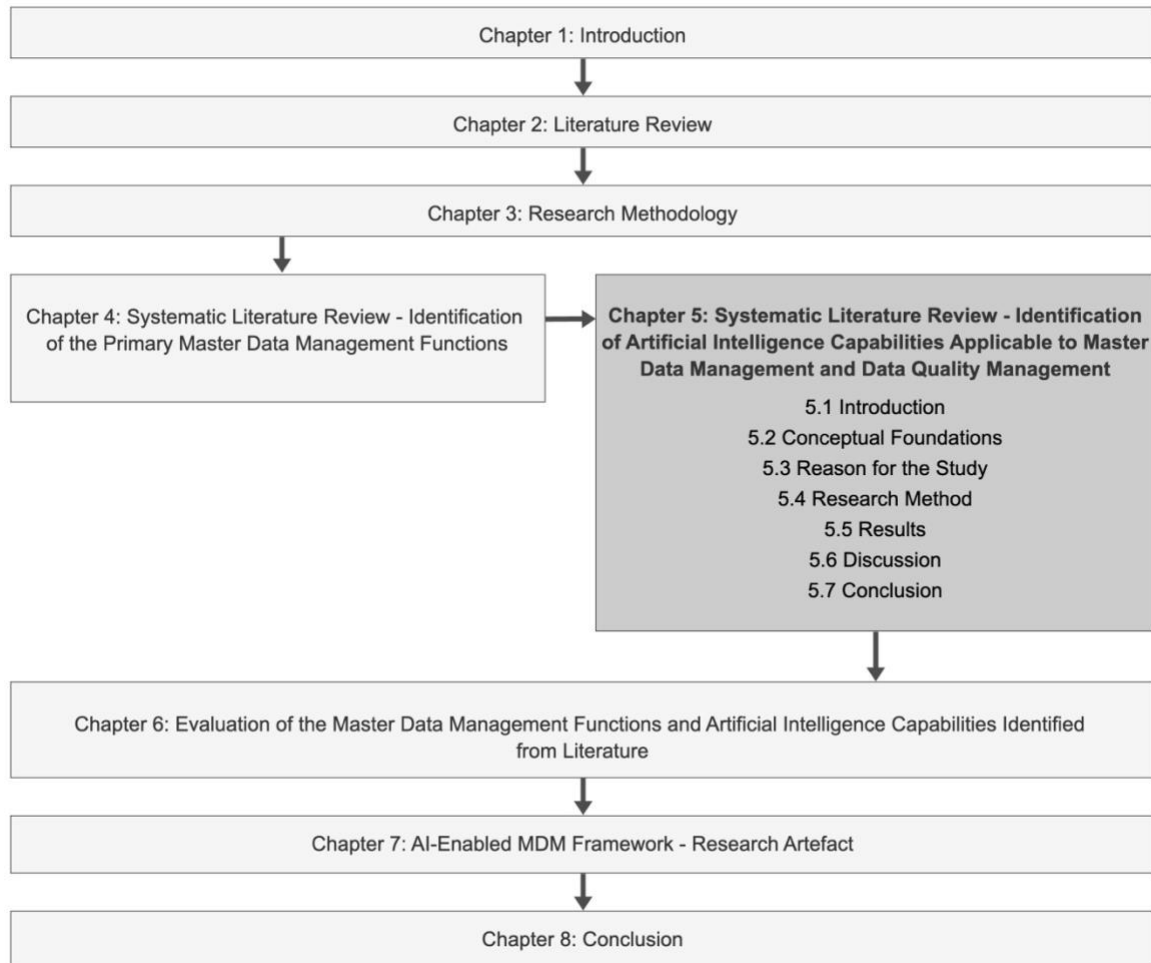


Figure 16: Structure of Research Paper - Position of Chapter 5

The following section describes the Systematic Literature Review (SLR) conducted to identify the capabilities of Artificial Intelligence (AI) that apply to Master Data Management (MDM) and Data Quality Management (DQM) functions, which aim to increase the overall data quality of master data residing in organisations' information systems, such as Enterprise Resource Planning (ERP) systems.

This is the second of two literature reviews that were conducted to accomplish the research outcome of the main study. The findings from this SLR are later combined with those of the first SLR in chapter 4, to

produce the key research artefact, namely an *AI-Enabled Master Data Management Framework* that maps specific AI capabilities to the various MDM functions of an organisation. This SLR is fundamental to constructing the final artefact because it identifies an exhaustive list of AI capabilities that can be applied to MDM and DQM, which will build on the findings from the previous study that identified the primary MDM functions for an organisation utilising ERP systems.

The SLR approach ensures that the study follows a strategic and organised process of identifying the ideal papers from relevant databases, without overlooking important literature and information. The conceptual foundations of the SLR, the research objective, the research question, and the research method is laid out in the following sections, and the SLR concludes with the results and discussion of the findings.

5.1 INTRODUCTION

As the field of data management progresses towards utilising Artificial Intelligence (AI) tools for managing data and automating tasks that were previously performed manually, there is a significant need for researchers to investigate the impact and application of these tools. Data quality is a crucial aspect of the adoption of AI in industrial enterprises utilising tools such as Enterprise Resource Planning (ERP), Product Lifecycle Management (PLM) and Manufacturing Execution Systems (MES) (Gröger, 2021).

As mentioned, this study used a Systematic Literature Review (SLR) approach to determine which Artificial Intelligence capabilities can be applied to Master Data Management functions for increased data quality. Numerous research papers highlight the potential of applying AI solutions to data processing and maintenance to enhance those processes and aid in data analytics for business decision-making (Kulkov, 2021; Riesener et al., 2022). Natural Language Processing (NLP) and Machine Learning (ML) are among the AI tools that have the most powerful impact on Master Data Management (MDM) functions in an organisation (Riesener et al., 2022). Thus, NLP and ML were included in the study's search strategy to ensure all capabilities related to AI tools and methods applicable to MDM have been extracted from the literature.

This SLR aims to identify the primary capabilities of AI that can be applied to MDM and DQM. There is a need to address the lack in academic literature that identifies a comprehensive list of AI capabilities that can be applied to MDM and DQM for an increase in the overall master data quality. The results of this study will advise future researchers and academics about the state of AI and automation in the field of data management. It will assist research that aims to enhance the different aspects of MDM and DQM. The following section provides the conceptual foundations of the study.

5.2 CONCEPTUAL FOUNDATIONS

This section defines the conceptual foundations referred to throughout the Systematic Literature Review. Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Data Quality Management and Master Data Management are defined from the literature to establish the fundamental terminology and definitions used in the study.

5.2.1 ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI) is defined by Aldoseri et al. (2023) as the capability of machines to execute functions that normally require human reasoning and judgement. This is accomplished by imitating human intelligence to learn, make decisions and solve problems, and it includes a large subset of tools such as Machine Learning and Natural Language Processing (Aldoseri et al., 2023; Riesener et al., 2022).

5.2.2 MACHINE LEARNING

Machine Learning (ML) is defined in Pontello et al. (2021) as “the field of study that gives computers the ability to learn without explicit programming”. It is a powerful tool that can automatically learn and make data attribute predictions from analysing intricate patterns in business data (Tverdal et al., 2024).

Riesener et al. (2022) and Aldoseri et al. (2023) support this definition by stating that Machine Learning is a subset of Artificial Intelligence that involves training computer algorithms to make predictions and decisions by analysing and learning the patterns in data. The paper by Leary (2014) also states that ML approaches produce business insights by analysing relationships among data from different data sources.

5.2.3 NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is a type of Artificial Intelligence (AI) tool that can recognise data patterns in large datasets (Riesener et al., 2022). It refers to the capability of a computer to recognise and discern human language through text and verbal expression and subsequently interpret and produce human language (Aldoseri et al., 2023).

Researchers such as Spreitzenbarth et al. (2024) state that NLP can assist in data quality enhancement activities such as the de-duplication of master data and correcting data entry errors. As an analytical technique, NLP can discover patterns in data quality (Rangineni et al., 2023), and it is an important tool for automating tasks in Master Data Management systems (Roth et al., 2022).

5.2.4 DATA QUALITY MANAGEMENT

Data Quality Management (DQM) comprises data quality-enhancing methods such as removing erroneous data, quality monitoring and detection, and fixing incorrect data (Tverdal et al., 2024). According to the paper Tverdal et al. (2024), DQM includes techniques such as Data Profiling, Data Cleansing, and Data Repair to improve the quality of data across systems. The paper Jonathan and Frank (2023) states that the assessment and monitoring of data quality forms part of DQM.

5.2.5 MASTER DATA MANAGEMENT

Master Data Management is defined in section 4.2.1 as a general term used to define the tasks and activities that organisations undertake to manage high-quality master data in the Enterprise Resource Planning (ERP) systems (Siren et al., 2015). The purpose of MDM is to supply valid and high-quality data to the value chain, enabling smooth business processes (Riesener et al., 2022).

5.3 REASON FOR THE STUDY

5.3.1 OBJECTIVE

This SLR aims to identify the primary capabilities of Artificial intelligence (AI) that can be applied to functions of a Master Data Management (MDM) system from current academic literature. The study aims to address the gap in AI and MDM literature identifying the capabilities of AI that can be applied to MDM systems for increased master data quality in the overall Enterprise Resource Planning (ERP) system. It aims to advance academic literature by producing a list of Artificial Intelligence capabilities that can be mapped to specific Master Data Management functions.

The results will assist future research, such as studies focusing on identifying how to integrate AI functionality into an organisation's MDM systems. This will provide researchers and industry professionals with a reference framework for further research in the field of AI and MDM.

5.3.2 RESEARCH QUESTION

What are the primary capabilities of Artificial Intelligence that can be applied to Master Data Management functions and Data Quality Management functions?

5.4 RESEARCH METHOD

The Systematic Literature Review (SLR) research method was used in this study to identify the Artificial Intelligence capabilities applicable to Master Data Management and Data Quality Management functions, as defined by Boland et al. (2017). The Systematic Literature Review method is a structured and

methodical method of discovering, evaluating, and combining literature that relates to the research objective to produce research evidence information.

5.4.1 SEARCH STRATEGY

Sources relevant to the SLR have been extracted by using the following search string for identifying journal articles to assist in answering the research question:

("Master Data Management" OR "Data Quality Management") AND ("Artificial Intelligence" OR AI OR "Machine Learning" OR ML OR "Natural Language Processing" OR NLP)

The following databases were searched using the defined search string above:

- IEEE Xplore Digital Library,
- ACM Digital Library,
- ScienceDirect,
- Emerald Insight, and
- Google Scholar.

5.4.2 SELECTION CRITERIA

Predefined selection criteria was applied to the journal articles to extract the most relevant sources for use in the SLR. The inclusion and exclusion criteria that was defined to refine the selection of journal articles for inclusion in the SLR are outlined below.

Inclusion Criteria

1. The article is accessible and written in English.
2. The document type is a journal, review, research, or conference paper.
3. The article discusses the application of Artificial Intelligence (AI), Machine Learning (ML), or Natural Language Processing (NLP) to Master Data Management (MDM) or Data Quality Management (DQM).

Exclusion Criteria

1. Duplicate research articles are excluded.
2. Research that does not discuss Artificial Intelligence, Machine Learning, or Natural Language Processing.
3. Sources that do not demonstrate how Artificial Intelligence, Machine Learning, or Natural Language Processing can be applied to Master Data Management or Data Quality Management.

5.4.3 SCREENING AND SELECTION

The research articles found using the search string were filtered, as part of the screening and selection step, based on the inclusion and exclusion criteria so as to ensure that the academic articles selected for analysis in the SLR are suitable for informing the research question. The defined search string was applied to each of the databases as set out in the search strategy, and the table below displays the initial database search results. The results in the table of SLR databases are limited to articles written in English and the document types specified in the inclusion criteria.

Database	Results
IEEE Xplore Digital Library	65
ACM Digital Library	117
ScienceDirect	209
Emerald Insight	108
Total	499

Table 8: SLR Databases

In addition to the databases mentioned above, Google Scholar was also searched to find relevant articles. The Google Scholar search resulted in 882 documents that were included in the total documents that were used in the screening and selection. This changed the total results to 1381 documents included in the analysis.

Following the collection of initial articles, duplicates were removed as part of the filtration process and the abstracts of the remaining articles were then screened for appropriacy. Thereafter, irrelevant articles were excluded. Full-text articles were subsequently assessed against the defined inclusion and exclusion criteria, and all studies considered appropriate for the primary topic of the study were included in the literature review.

5.4.4 PRISMA FLOWCHART

The Prisma Flowchart is presented below to graphically show the record identification, screening, assessment for eligibility, and inclusion of all the journal articles identified for the study. The results of each step are displayed as described in the screening and selection section.

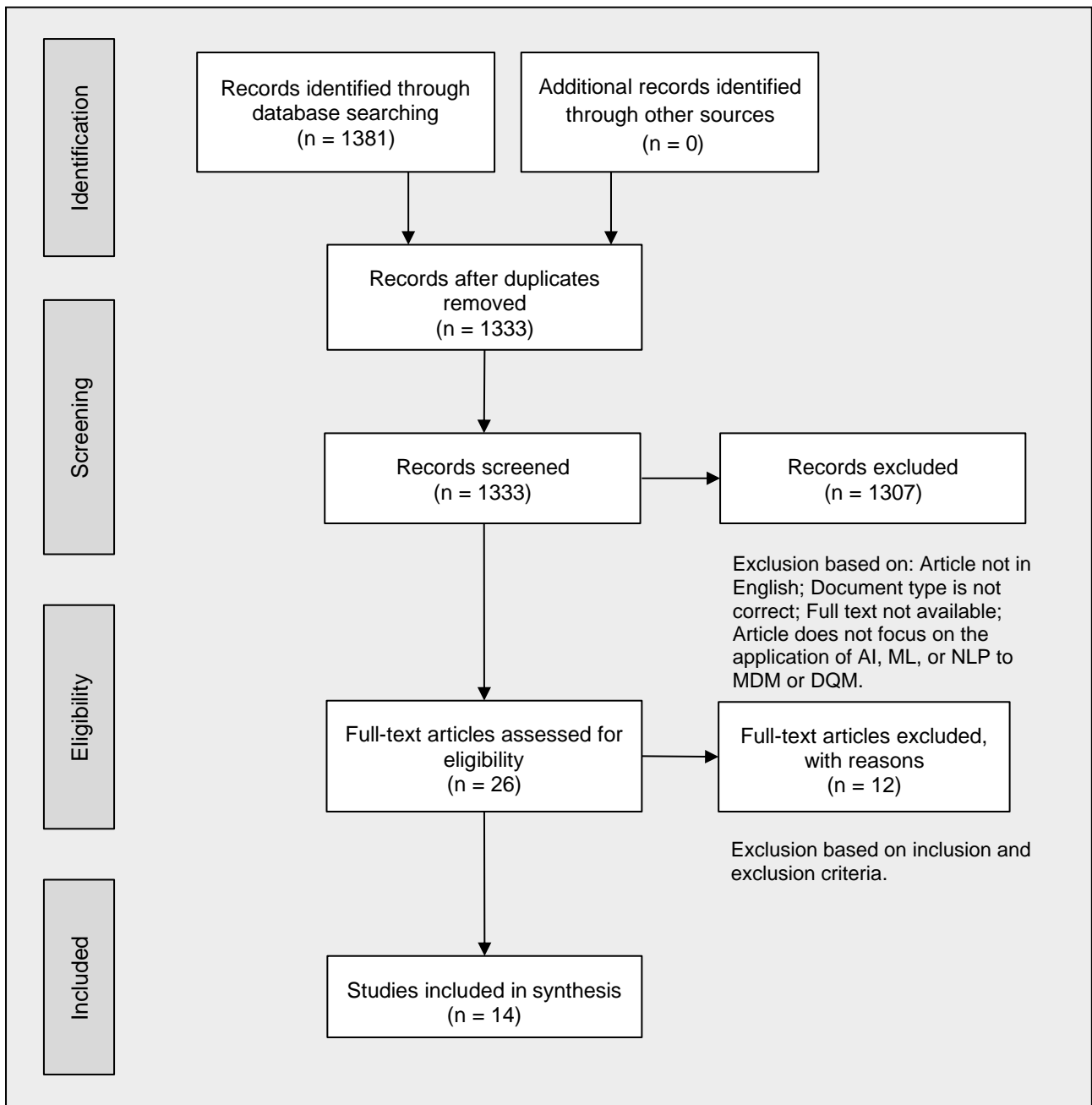


Figure 17: PRISMA Flowchart – Adapted from Moher et al. (2009).

5.4.5 QUALITY ASSESSMENT

Predefined journal article quality assessment criteria have been defined to assess the academic quality of each of the journal articles included in this literature review. The quality assessment questions defined below have been applied to each source to determine its quality and relevance to the research topic:

- Question 1: Is the article primarily focused on Artificial Intelligence, Machine Learning or Natural Language Processing?
- Question 2: Does the article adequately define any applications of Artificial Intelligence, Machine Learning or Natural Language Processing to Master Data Management or Data Quality Management?
- Question 3: Is Master Data Management or Data Quality Management described within the context of Enterprise Resource Planning systems?

The following table contains the answer code and description for each quality assessment question:

Question	Answer Code	Description
1	Y (Yes)	The article is primarily focused on Artificial Intelligence, Machine Learning or Natural Language Processing.
	P (Partially)	The article does not primarily focus on Artificial Intelligence, Machine Learning or Natural Language Processing but does mention it within the broader topic.
	N (No)	The article does not mention on Artificial Intelligence, Machine Learning or Natural Language Processing.
2	Y (Yes)	The article adequately defines one or more applications of Artificial Intelligence, Machine Learning or Natural Language Processing to Master Data Management or Data Quality Management.
	P (Partially)	The article indirectly defines one or more applications of Artificial Intelligence, Machine Learning or Natural Language Processing to Master Data Management or Data Quality Management.
	N (No)	The article does not define any applications of Artificial Intelligence, Machine Learning or Natural Language Processing to Master Data Management or Data Quality Management.
3	Y (Yes)	Master Data Management or Data Quality Management is described within the context of Enterprise Resource Planning systems.
	P (Partially)	Master Data Management or Data Quality Management is not described specifically within the context of Enterprise Resource Planning systems but within the broader context of enterprise information systems.
	N (No)	Master Data Management and Data Quality Management is not described within the context of Enterprise Resource Planning systems or within the context of enterprise information systems.

Table 9: Quality Assessment Codes

As per Kitchenham et al. (2009), scoring of the quality assessment is carried out by applying the following quality scores to each of the relevant answer codes per the research article:

Y (Yes) = 1, P (Partial) = 0.5, and N (No) = 0, or Unknown (unspecified).

The tables on the following pages contain the source list of research articles included in the study and the evaluation scores assigned to each, according to the quality assessment. The first table contains the list of sources used in the SLR, with information about the author, title, journal publication, database and year in which the source was published. Following the first table, the second table presents the quality evaluation scores for each research article.

The following table contains the source list of research articles included in the Systematic Literature Review. The Author, Title, Database, and Year is included.

ID	Author	Title	Database	Year
S1	(Riesener et al., 2022)	Methodology for Automated Master Data Management using Artificial Intelligence	IEEE	2022
S2	(Pontello et al., 2021)	Meta-learning approach for implementation of AI methods in the context of CRISP-DM with case studies from master data management	IEEE	2021
S3	(Leary, 2014)	Embedding AI and Crowdsourcing in the Big Data Lake	IEEE	2014
S4	(Zhang et al., 2021)	Leveraging Approximate Constraints for Localized Data Error Detection	ACM	2021
S5	(Tverdal et al., 2024)	Edge-based Data Profiling and Repair as a Service for IoT	ACM	2024
S6	(Guo et al., 2018; Xu et al., 2021)	Research on Key Problems of Data Quality in Large Industrial Data Environment	ACM	2018
S7	(Spreitzenbarth et al., 2024)	Artificial intelligence and machine learning in purchasing and supply management: A mixed-methods review of the state-of-the-art in literature and practice	Science Direct	2024
S8	(Pakkala et al., 2024)	Improving efficiency and quality of operational industrial production assets information management in customer – vendor interaction	Science Direct	2024
S9	(Kulkov, 2021)	The role of artificial intelligence in business transformation: A case of pharmaceutical companies	Science Direct	2021
S10	(Liu & Anderson, 2024)	A Hybrid Framework for Master Data Management: Integrating Machine Learning and Traditional Approaches	Google Scholar	2024
S11	(Rangineni et al., 2023)	A Review on enhancing data quality for optimal data analytics performance	Google Scholar	2023
S12	(Jonathan & Frank, 2023)	Real-Time Data Monitoring and Anomaly Detection with AI: A Comprehensive Overview	Google Scholar	2023
S13	(Jesmeen et al., 2018)	A survey on cleaning dirty data using machine learning paradigm for big data analytics	Google Scholar	2018
S14	(Roth et al., 2022)	Towards Augmented MDM: Overview of Design and Function Areas–A Literature Review	Google Scholar	2022

Table 10: Sources Included in SLR

The following table consists of the quality assessment scores given to each literature paper included in the Systematic Literature Review.

ID	Q1	Q2	Q3	Total Score
S1	1	1	0	2
S2	1	1	0.5	2.5
S3	1	1	1	3
S4	1	1	0	2
S5	0	1	0	1
S6	0.5	0.5	0	1
S7	1	1	0.5	2.5
S8	0.5	0.5	1	2
S9	1	0.5	0	1.5
S10	0.5	1	0	1.5
S11	0.5	1	0	1.5
S12	1	1	0	2
S13	1	1	0	2
S14	1	0.5	1	2.5

Table 11: Quality Evaluation Scores

5.4.6 DATA EXTRACTION

The data extraction process consisted of identifying specific dimensions in each of the papers included in the SLR. Individual sections from the text of each research paper in the literature reviewed were examined to identify the key codes and themes in the text that relate to the research topic.

For each paper, the AI capabilities that relate to either Master Data Management or Data Quality Management were extracted, as well as the type of technology mentioned in the paper, either Artificial Intelligence (AI), Machine Learning (ML) and/or Natural Language Processing (NLP). Additionally, whether the paper made reference to Master Data Management (MDM) or Data Quality Management (DQM) was also extracted as a dimension for each literature paper.

5.4.7 DATA ANALYSIS

To identify the themes and codes within the literature papers included in the study, thematic analysis was applied to the research, as set out by Braun and Clarke (2006). These codes and themes were identified and classified by extracting the Artificial Intelligence (AI) capabilities from each literature paper. The type of AI technology mentioned in each paper was captured and whether each AI applicability is applicable to Master Data Management (MDM) or Data Quality Management (DQM) was also captured. The primary technologies mentioned in the papers were AI, Machine Learning (ML) and Natural Language Processing (NLP).

The themes and codes extracted differed per literature paper. AI capabilities, such as master data pattern recognition, automated master data harmonisation, and more, were identified from the literature. The most prominent wording from the literature was used as the themes and codes to determine the AI capabilities that should be grouped together. The final list of AI capabilities for Master Data Management and Data Quality Management was produced using these themes and codes. The findings have been grouped in the following section, and references to the relevant papers are included to inform the research question.

5.5 RESULTS

5.5.1 RESEARCH TRENDS

There is a limited amount of research on Artificial Intelligence (AI) capabilities relating to Master Data Management (MDM) and Data Quality Management (DQM) in the AI and Data Management subject area. Most papers on this topic have been written in recent years, the oldest paper included in the study having been published in 2014. The following graph displays the research trends based on the literature papers included in the study, an upward curve in literature papers is evident.

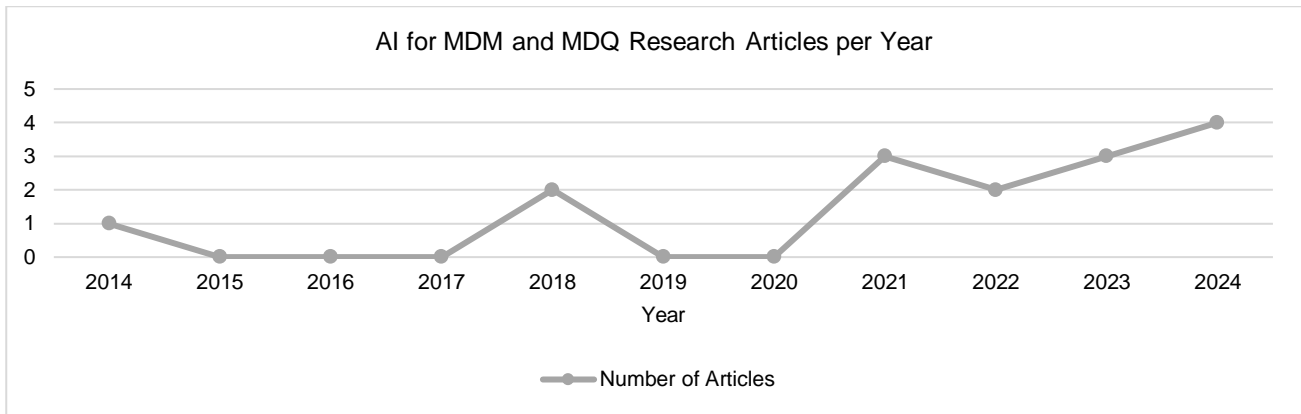


Figure 18: AI for MDM and DQM Research Trends

5.5.2 ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT – INITIAL FINDINGS

The following table consists of the Artificial Intelligence (AI) capabilities that are applicable to Master Data Management and Data Quality Management, as identified from the literature. The dimensions extracted from each literature paper are the AI capabilities identified, the technology mentioned in the study, and whether the AI capabilities identified in the literature are applicable to Master Data Management (MDM) or Data Quality Management (DQM).

Reference	AI Capabilities Identified	Technology	Applicable to MDM or DQM
(Riesener et al., 2022)	Mass Data Maintenance; Master Data Maintenance Automation; Duplicate Record Identification; Data Attributes Prediction	AI, ML, NLP	MDM
(Pontello et al., 2021)	Master Data Pattern Recognition and Clustering	AI, ML	MDM
(Leary, 2014)	Master Data De-Duplication	AI, ML	MDM
(Zhang et al., 2021)	Master Data Error Detection; Master Data De-Duplication	AI, ML	DQM
(Tverdal et al., 2024)	Detection and Correction of Missing or Erroneous Master Data	ML	DQM
(Guo et al., 2018)	Master Data Duplicate Detection	AI, ML	DQM
(Spreitzenbarth et al., 2024)	Master Data De-Duplication	AI, ML, NLP	MDM
(Pakkala et al., 2024)	Identification and Predictions for Missing Master Data Fields; Automatic Master Data Harmonisation; Master Data Enrichment	AI, ML	MDM

Reference	AI Capabilities Identified	Technology	Applicable to MDM or DQM
(Kulkov, 2021)	Master Data Analysis	AI, ML	MDM
(Liu & Anderson, 2024)	Master Data Pattern Recognition; Master Data Anomaly Detection; Master Data Matching and De-duplication; Identification of Inconsistencies and Hidden Relationships in Master Data	AI, ML	MDM
(Rangineni et al., 2023)	Master Data Anomaly and Inconsistency Detection; Identification and Prediction of Erroneous and Inconsistent Master Data; Automated Master Data Quality Assessment and Checks; Detection of Master Data Quality Patterns	AI, ML, NLP	DQM
(Jonathan & Frank, 2023)	Master Data Anomaly Detection; Master Data Monitoring; Automated Data Cataloguing; Master Data Ingestion and Integration (as part of Automated Data Cataloguing); Master Data Search and Discovery (as part of Automated Data Cataloguing); Master Data Lineage and Relationships Management (as part of Automated Data Cataloguing); Master Data Quality Assessment	AI, ML, NLP	DQM
(Jesmeen et al., 2018)	Automated Master Data Cleansing	ML	DQM
(Roth et al., 2022)	Master Data Integration	AI, ML, NLP	MDM

Table 12: AI Capabilities Identified from Literature

To identify the AI capabilities identified from the literature included in this study, the findings from each paper were classified according to AI capabilities in the following section. Thereafter, these AI capabilities were reviewed according to the literature trends to determine which AI capabilities are the most prominent from the results of this study.

5.5.3 ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT – CLASSIFICATION

The following table consists of the Artificial Intelligence (AI) capabilities that apply to Master Data Management and Data Quality Management. The AI capabilities are classified according to similarities in the descriptions of each identified capability from each literature paper, and the papers used to identify each AI capability are grouped together.

AI Data Management Capabilities Identified	Reference
Automated Master Data Maintenance	(Riesener et al., 2022)
Mass Data Maintenance	(Riesener et al., 2022)
Data Attributes Prediction	(Riesener et al., 2022), (Pakkala et al., 2024), (Rangineni et al., 2023)
Master Data Duplicate Detection and De-Duplication	(Riesener et al., 2022), (Leary, 2014), (Zhang et al., 2021), (Guo et al., 2018), (Spreitzenbarth et al., 2024), (Liu & Anderson, 2024)
Master Data Error Detection and Correction	(Zhang et al., 2021), (Tverdal et al., 2024), (Rangineni et al., 2023)
Identification of Inconsistent and Missing Master Data Fields	(Pakkala et al., 2024), (Jonathan & Frank, 2023)
Automated Master Data Harmonising, Enrichment and Cleansing	(Pakkala et al., 2024), (Jesmeen et al., 2018)
Automated Master Data Quality Assessment and Checks	(Rangineni et al., 2023)
Master Data Analysis and Monitoring	(Kulkov, 2021), (Jonathan & Frank, 2023)
Master Data Pattern and Anomaly Detection	(Liu & Anderson, 2024), (Rangineni et al., 2023), (Jonathan & Frank, 2023)
Master Data Cataloguing	(Jonathan & Frank, 2023), (Roth et al., 2022)

Table 13: AI Capabilities for MDM and DQM - Classification

Master Data Anomaly Detection (Jonathan & Frank, 2023; Liu & Anderson, 2024), *Master Data Pattern Recognition* (Liu & Anderson, 2024), and *Master Data Inconsistency Detection* (Rangineni et al., 2023) were grouped together under the classification *Master Data Pattern and Anomaly Detection*. Terms relating to Data De-duplication such as *Duplicate Record Identification* (Riesener et al., 2022), *Master Data Matching and De-duplication* (Liu & Anderson, 2024) and *Master Data De-Duplication* (Spreitzenbarth et al., 2024), were grouped together under the category called *Master Data Duplicate Detection and De-Duplication*. References to erroneous and inconsistent Master Data, such as *Identification and Prediction of Erroneous and Inconsistent Master Data* extracted from Rangineni et al. (2023) and *Detection and Correction of Missing or Erroneous Master Data* extracted from Tverdal et al. (2024) were grouped together under *Master Data Error Detection*.

The references to prediction of master data such as *Identification and Predictions for Missing Master Data Fields* (Pakkala et al., 2024) were grouped together under the category *Data Attributes Prediction*. Capabilities relating to the management of master data quality, such as *Detection of Master Data Quality Patterns* (Rangineni et al., 2023) and *Master Data Quality Assessment* (Jonathan & Frank, 2023) were grouped together under the category *Master Data Quality Management*. The AI capabilities *Master Data Ingestion and Integration*, *Master Data Search and Discovery* and *Master Data Lineage and*

Relationships Management were grouped under *Data Cataloguing* as per the paper by Jonathan and Frank (2023).

5.5.4 ARTIFICIAL INTELLIGENCE CAPABILITIES LITERATURE TRENDS

To determine which AI capabilities are the most prominent from the results of this study, the results from the data categorisation step are displayed in the following graph which depicts how many research papers mentions each Artificial Intelligence (AI) capability applicable to Master Data Management (MDM) and Data Quality Management (DQM).

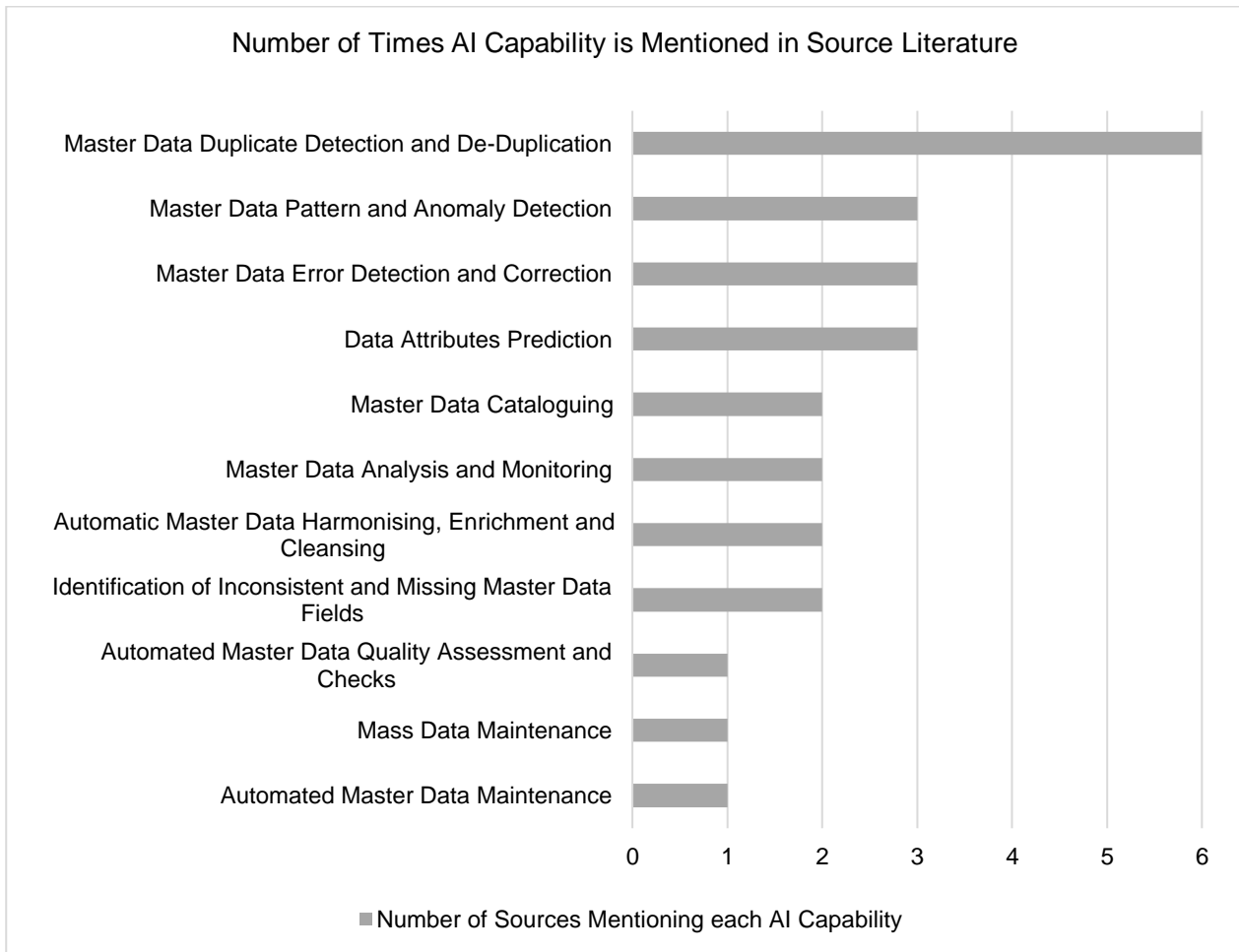


Figure 19: MDM Functions Literature Trends

This provides a visualisation of the AI capabilities applicable to MDM and DQM which are most prominent in the literature. It is clear that *Master Data Duplicate Detection and De-Duplication* is the most popular capability of AI that applies to MDM and DQM. Thereafter, *Master Data Pattern and Anomaly Detection*, *Master Data Error Detection and Correction*, and *Master Data Attributes Prediction* are the most noteworthy AI capabilities extracted from the literature papers.

5.6 DISCUSSION

This systematic literature review identified the primary capabilities of Artificial Intelligence (AI) that are applicable to Master Data Management (MDM) and Data Quality Management (DQM) functions. The AI capabilities identified in the literature were appropriately classified to produce the final list of AI capabilities and the most popular findings in literature were discovered.

The most popular application of AI to MDM and DQM extracted from literature is *Master Data Duplicate Detection and De-Duplication*. Leary (2014) states that AI can be used to remove duplicate records from databases, and Spreitzenbarth et al. (2024) mentions the application of Natural Language Processing (NLP) to assist in removing duplicate supplier master data.

Furthermore, *Master Data Pattern and Anomaly Detection*, *Master Data Error Detection and Correction*, and *Master Data Attributes Prediction* are mentioned multiple times in the literature. Liu and Anderson (2024) recommend a hybrid framework for MDM using traditional approaches combined with Machine Learning (ML) techniques to advance master data quality management and they claim that patterns and anomalies in master data can be identified by supervised learning algorithms.

Using tools such as ML, AI and NLP can contribute to master data error detection and the prediction of master data attributes (Pakkala et al., 2024; Rangineni et al., 2023; Riesener et al., 2022; Zhang et al., 2021). Rangineni et al. (2023) mentions that automated cleansing algorithms that utilise Machine Learning technologies are essential for detecting anomalies and highlighting inconsistent master data.

The AI capabilities that are mentioned less frequently in literature are *Identification of Inconsistent and Missing Master Data Fields* (Jonathan & Frank, 2023; Pakkala et al., 2024), *Automatic Master Data Harmonising, Enrichment and Cleansing* (Jesmeen et al., 2018; Pakkala et al., 2024), *Master Data Analysis and Monitoring* (Jonathan & Frank, 2023; Kulkov, 2021), and *Master Data Cataloguing* (Jonathan & Frank, 2023; Roth et al., 2022).

Jonathan and Frank (2023) claim that master data ingestion and integration can be automated by utilising AI-powered data cataloguing, which possess automation capabilities for extracting and loading data in a streamlined process. They also state that metadata enrichment can be promoted by means of AI-powered data cataloguing which makes use of automated metadata extraction techniques, which helps ensure users have a thorough understanding of their data (Jonathan & Frank, 2023). Pakkala et al. (2024) identify master data enrichment as an aspect of master data harmonisation functionality that automatically identifies missing master data, and Jesmeen et al. (2018) recognise the ability of ML techniques to clean master data. Pontello et al. (2021) state that AI and ML methods can be used for pattern recognition in master data so that relationships between master data can be automatically identified. Liu and Anderson (2024) support this by stating that ML techniques such as supervised learning algorithms can enable enhanced data matching by training these algorithms to recognise master data patterns.

The less prominent AI capabilities identified in the literature are *Mass Data Maintenance*, *Automated Master Data Maintenance* (Riesener et al., 2022), and *Automated Master Data Quality Assessment and*

Checks (Rangineni et al., 2023). The literature studies included in the SLR provide a comprehensive list of AI capabilities that are applicable to MDM and DQM. The findings of this study highlight the promising potential of automating and streamlining MDM and DQM processes by employing AI, ML and NLP technologies for enhanced data quality.

5.7 CONCLUSION

This systematic literature review identifies the primary capabilities of Artificial Intelligence (AI) that can be applied to Master Data Management (MDM) and Data Quality Management (DQM). The study addresses the gap in current literature by providing a comprehensive list of the types of AI capabilities that can be applied to MDM and DQM for an increase in the overall quality of master data.

The results of this study informs future research on AI and automation in the field of data management. It specifically assists research aimed at enhancing the different functions of Master Data Management and Data Quality Management. The findings also equip data management industry professionals with a reference of which AI capabilities to consider in future as AI tools advance and start to permeate the different functions of data management solutions.

The study has achieved its objective of identifying the AI capabilities that are applicable to MDM and DQM. Additional research is necessary to establish which AI capabilities identified in this study could be most effective in increasing master data quality in the broader context of Master Data Management systems for organisations utilising Enterprise Resource Planning (ERP) tools.

6 EVALUATION OF THE MASTER DATA MANAGEMENT FUNCTIONS AND ARTIFICIAL INTELLIGENCE CAPABILITIES IDENTIFIED FROM LITERATURE

The figure below shows the layout and position of Chapter 6 in relation to the rest of the research paper.

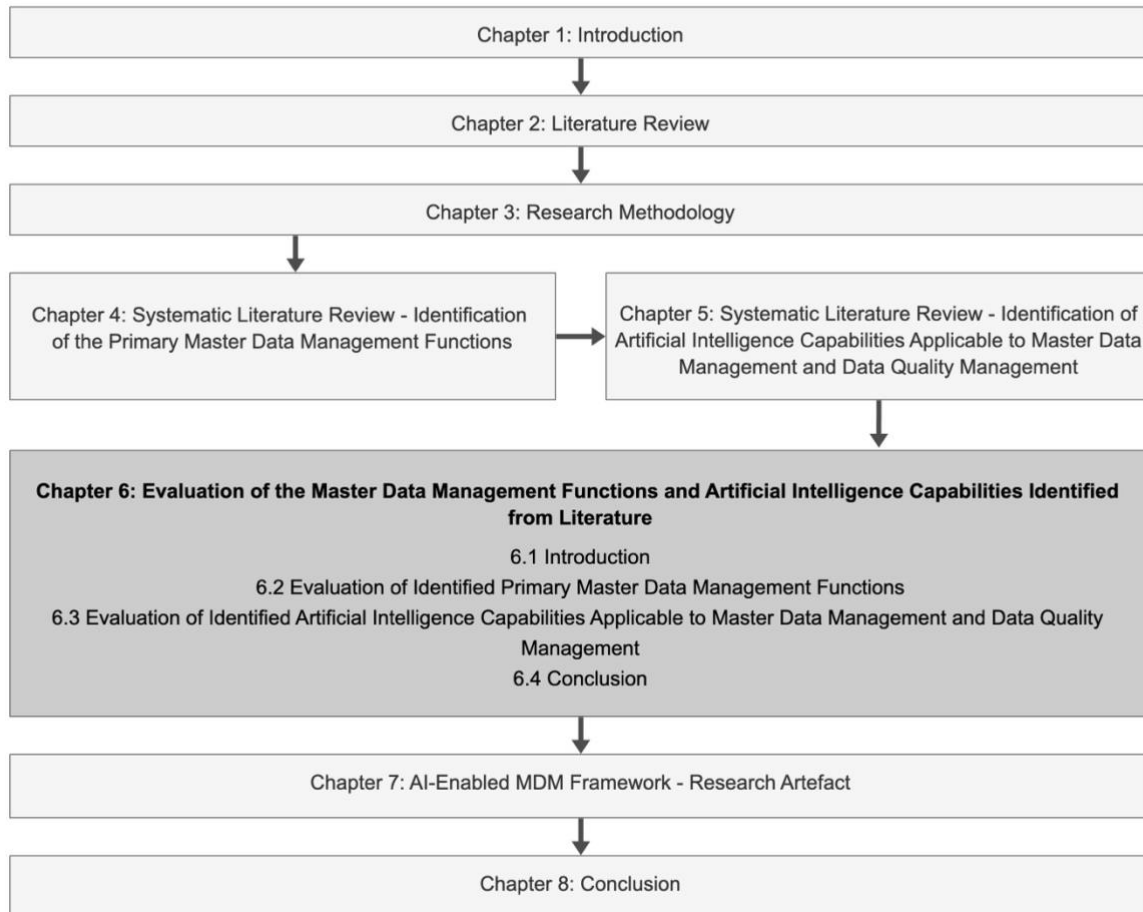


Figure 20: Structure of Research Paper - Position of Chapter 6

6.1 INTRODUCTION

This section serves as the evaluation phase of the research for assessing the findings of the two respective Systematic Literature Reviews (SLR) that were carried out in Chapter 4 and Chapter 5. The *Master Data Management (MDM) Functions* that were identified from literature and the *AI Capabilities Applicable to Master Data Management (MDM) and Data Quality Management (DQM)* identified from literature were presented to data management professionals with corresponding questions in the form of a survey.

The purpose of the survey is to determine the impact of each MDM function on master data quality, to determine which of the MDM functions require the most human involvement and which are the most time-

consuming. The survey also evaluates which AI capabilities are the most important for organisations to adopt and which could be most easily implemented into an MDM system.

The survey questions were designed with the diffusion of innovations (DOI) theory in mind to determine the perceived complexity and relative advantage of implementing AI tools in MDM Functions. The survey responses of the data management professionals helped determine the MDM functions they perceive as being most time-consuming, which need the most manual involvement, and which have the most significant impact on master data quality. The perceived complexity of integrating AI capabilities for specific MDM functions was determined by asking which AI capabilities extracted from literature could be most easily implemented into an MDM system.

The DOI innovation referred to in this research study is the idea of enhancing MDM functionality with AI tools and other AI capabilities. The survey findings clearly show that the adoption of AI capabilities in Master Data Management systems is perceived as a complex task that requires much research.

6.2 EVALUATION OF IDENTIFIED PRIMARY MASTER DATA MANAGEMENT FUNCTIONS

The findings from the first Systematic Literature Review (SLR), conducted in Chapter 4 to identify the primary functions of a Master Data Management (MDM) system in an organisation utilising Enterprise Resource Planning systems, were presented to a group of experts in the field of MDM to evaluate the quality of the SLR findings. The survey questions were based on the findings from the SLR in the form of the *MDM Functions Framework*, which can be seen in section 4.5.3. They were designed with the DOI theory in mind to establish the perceived difficulty and complexity of managing master data.

As mentioned, a survey data collection method was utilised, and a Google Forms survey was created to conduct the research. The research participants were 18 data consultants who have expertise in Master Data Management.

6.2.1 SURVEY OUTLINE – MASTER DATA MANAGEMENT FUNCTIONS

The survey questions applicable to the identified MDM functions are the following:

- Which Master Data Management Functions have the biggest impact on master data quality?
- Which Master Data Management Functions need the most human involvement?
- Which Master Data Management Functions are the most time-consuming?

The response options for each of the above questions were the same options that can be viewed below and the research participants were able to select all options from this list that were applicable:

- Master Data Model and Workflow Customising
- Master Data Import and Distribution
- Master Data Creation and Maintenance
- Master Data Analysis and Reporting

- Master Data Duplicate Detection and De-duplication
- Master Data Validation and Business Rules
- Master Data Enrichment and Cleansing
- Master Data Classification and Taxonomy Management
- Master Data Consolidation
- Master Data Roles and Authorisation Management

The next section displays the survey responses captured for each of the questions.

6.2.2 SURVEY RESPONSES – MASTER DATA MANAGEMENT FUNCTIONS

Survey Responses: Master Data Management Functions – Impact on Master Data Quality

The first question related to Master Data Management (MDM) Functions on the survey was:

- Which Master Data Management Functions have the most significant impact on master data quality?

The result from this research survey question was extracted from the Google Form and is displayed in the figure below.

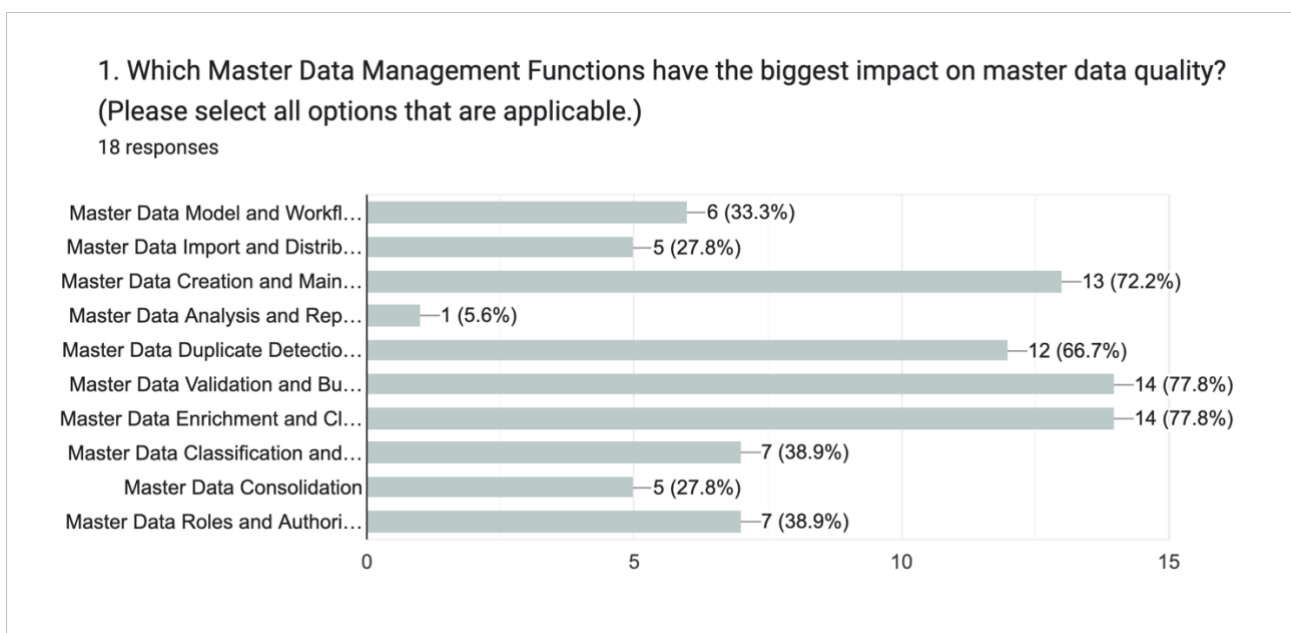


Figure 21: Survey Responses: MDM Functions – Impact on Master Data Quality

Survey Responses: Master Data Management Functions – Requires Most Human Involvement

The second question related to Master Data Management (MDM) Functions on the survey was:

- Which Master Data Management Functions need the most human involvement?

The result from this research survey question was extracted from the Google Form and is displayed in the figure below.

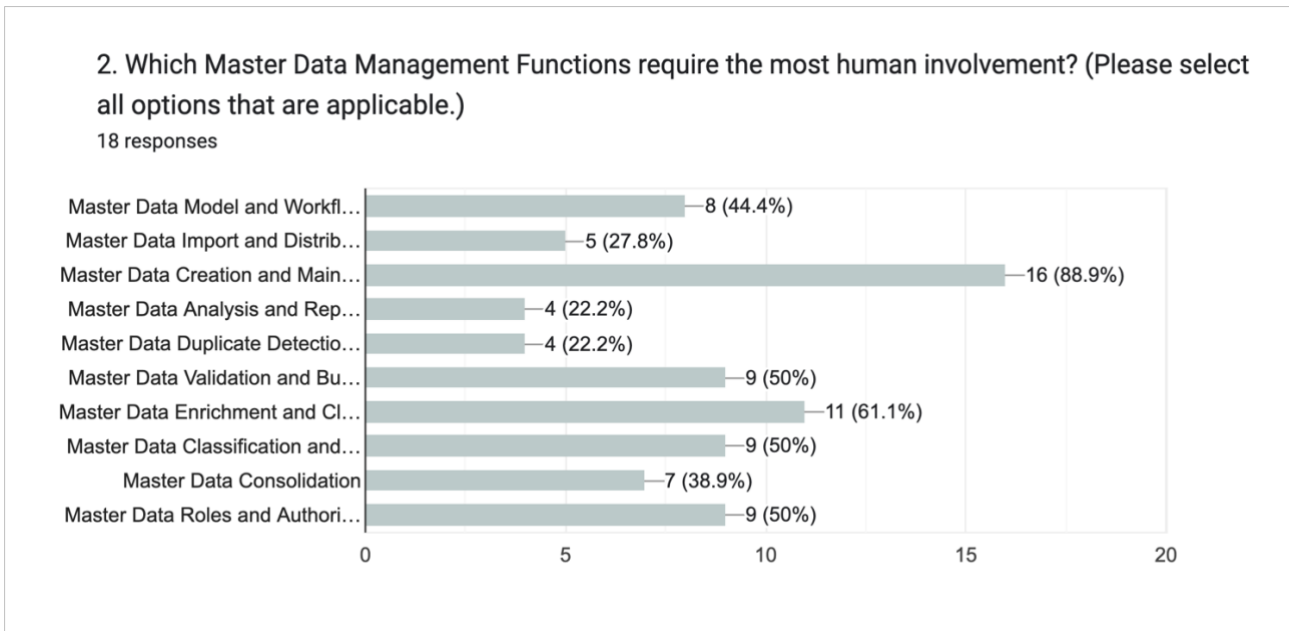


Figure 22: MDM Functions – Most Human Involvement

Survey Responses: Master Data Management Functions – Most Time-Consuming

The third question related to Master Data Management (MDM) functions on the survey was:

- Which Master Data Management Functions are the most time-consuming?

The result from this research survey question was extracted from the Google Form and is displayed in the figure below.

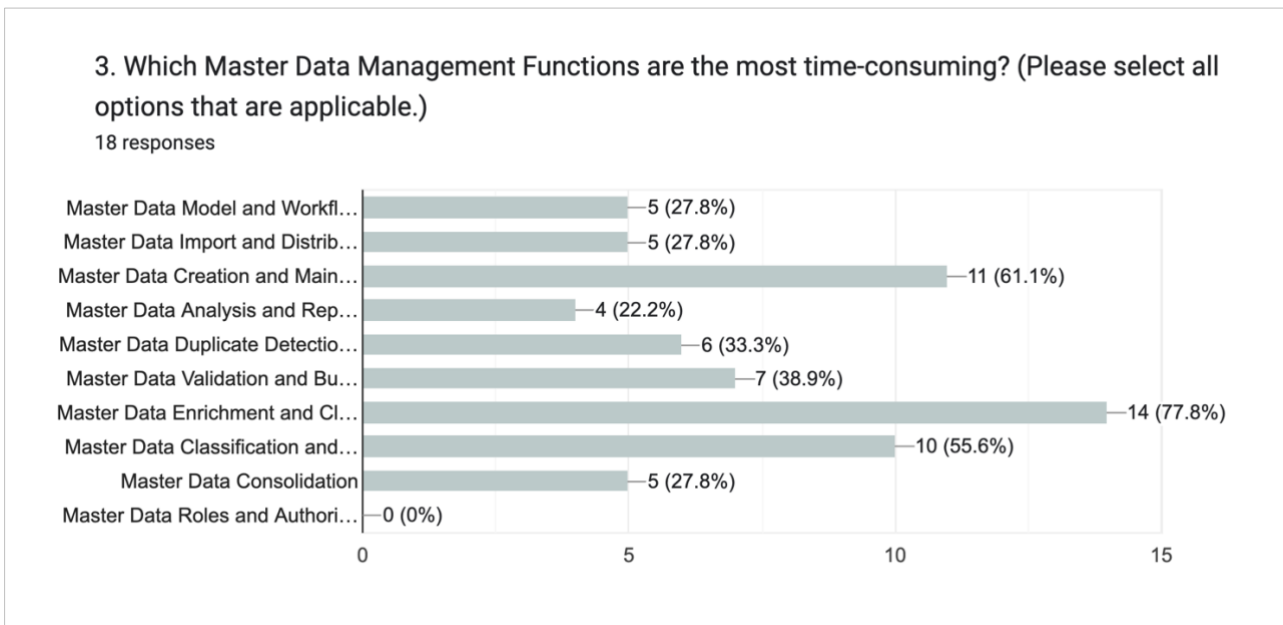


Figure 23: MDM Functions – Most Time-Consuming

6.2.3 ANALYSIS OF SURVEY FINDINGS – MASTER DATA MANAGEMENT FUNCTIONS

In this section, the survey questions related to the Master Data Management (MDM) functions are analysed and displayed in graphs to highlight each question's most popular response options. These findings were compared with the trends in literature for MDM functions seen in section 4.5.4 and combined with the survey findings and literature trends of the identified Artificial Intelligence (AI) Capabilities to create the final research artefact in Chapter 7.

Analysis of Survey Findings: Master Data Management Functions – Impact on Master Data Quality

The following graph provides a visualisation of the MDM functions that have the biggest impact on master data quality according to the data management professionals who took part in the research survey.

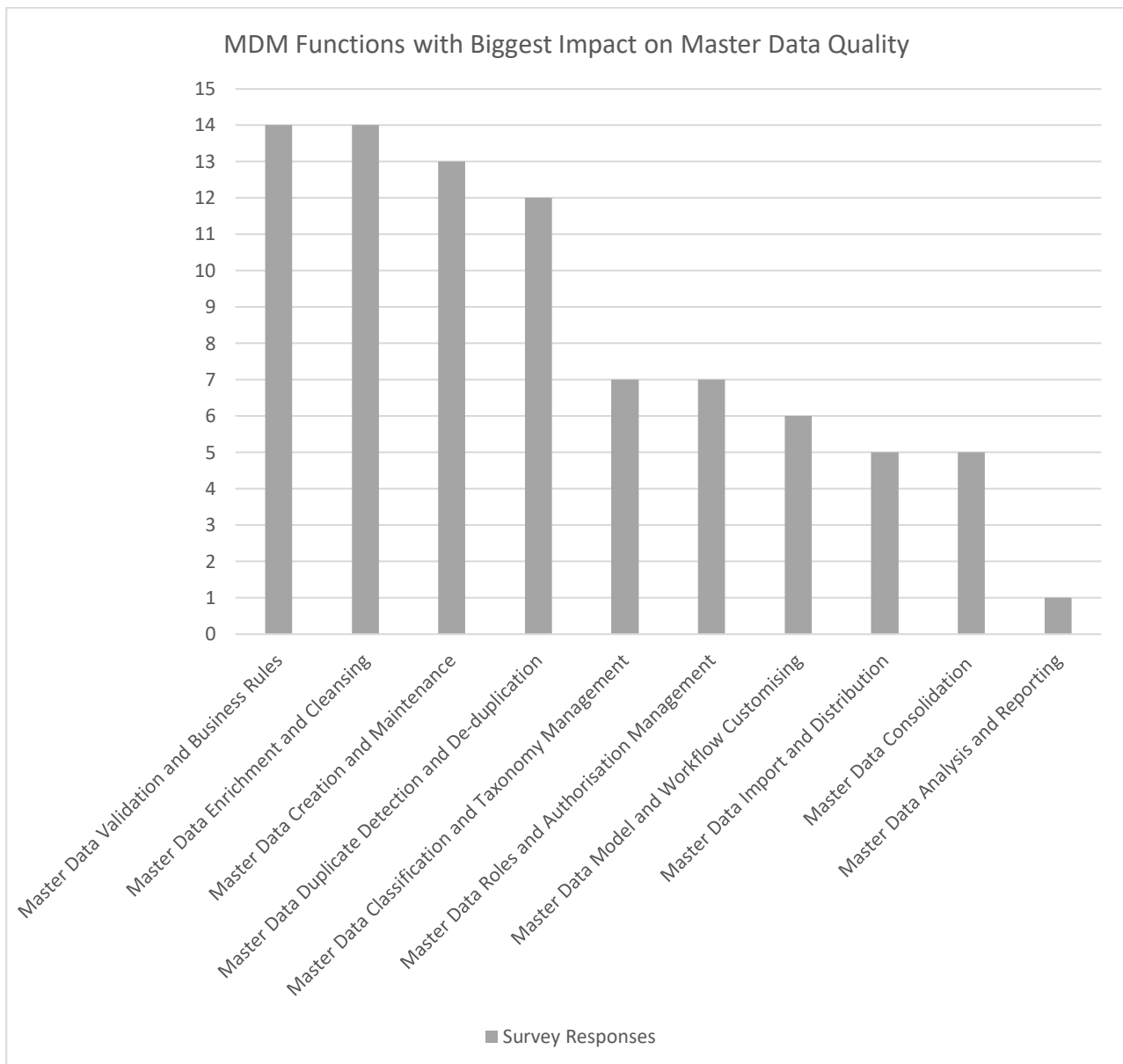


Figure 24: Analysis of Survey Findings: MDM Functions – Impact on Master Data Quality

From this, we can infer that the MDM functions which have the most significant effect on master data quality are *Master Data Validation and Business Rules*, *Master Data Enrichment and Cleansing*, *Master Data Creation and Maintenance*, and *Master Data Duplicate Detection and De-duplication*.

Analysis of Survey Findings: Master Data Management Functions – Requires Most Human Involvement

The next graph provides a visualisation of the MDM functions that require the most human involvement according to the data management professionals who took part in the research survey.

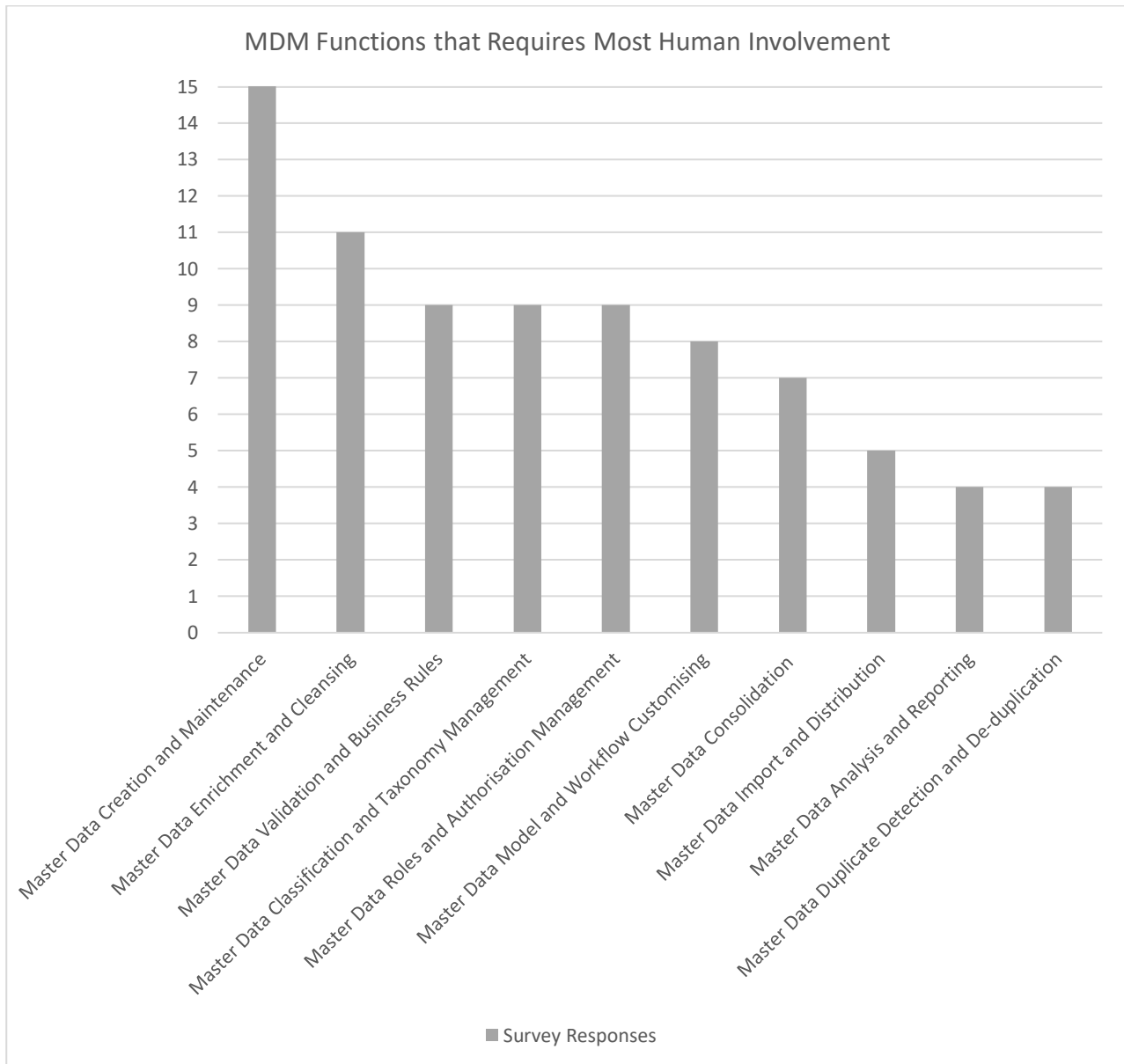


Figure 25: Analysis of Survey Findings: MDM Functions – Requires Most Human Involvement

From this we can conclude that the MDM functions which require the most human involvement are *Master Data Creation and Maintenance*, and *Master Data Enrichment and Cleansing*.

Analysis of Survey Findings: Master Data Management Functions – Most Time-Consuming

The graph provides a visualisation of the most time-consuming MDM functions according to the data management professionals who took part in the research survey.

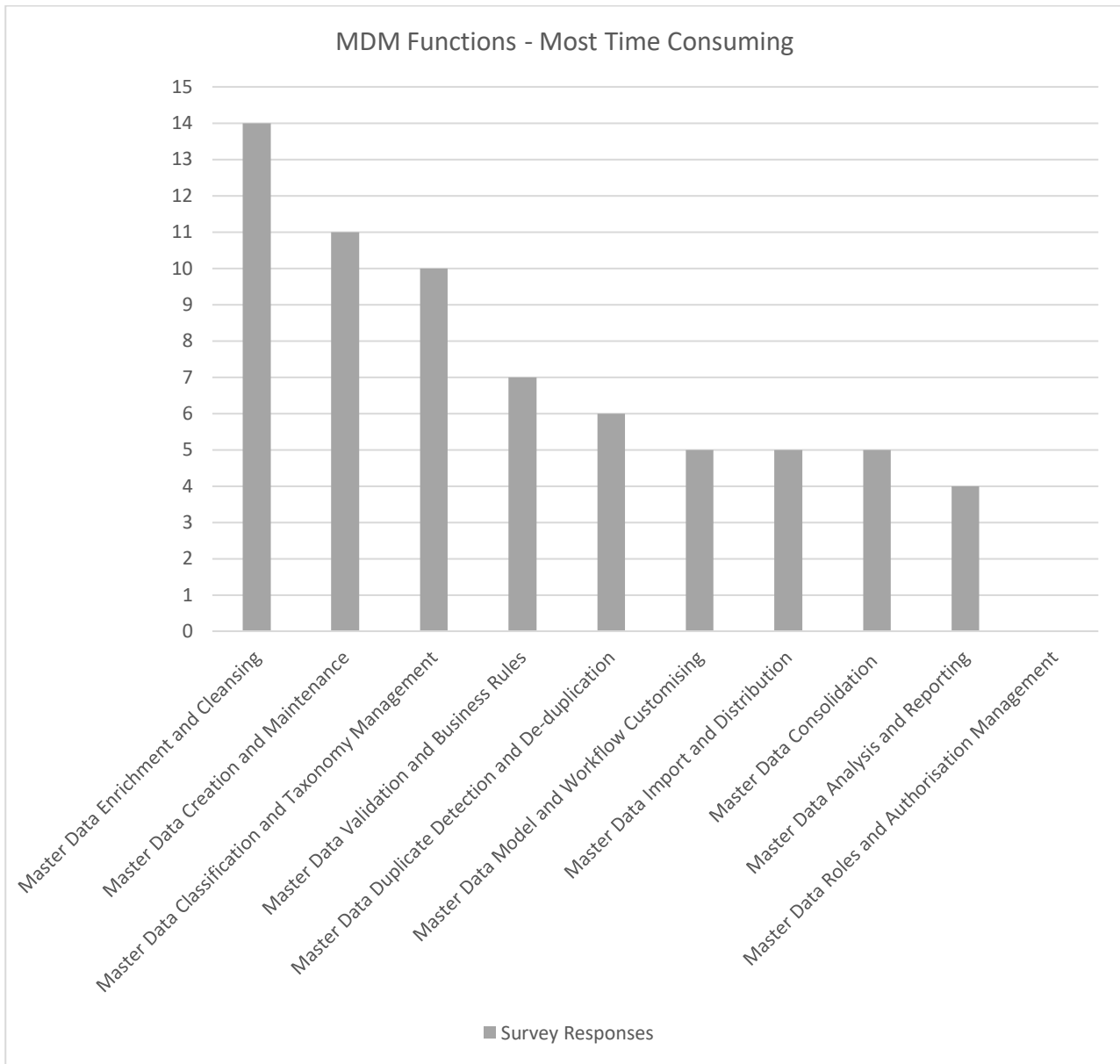


Figure 26: Analysis of Survey Findings: MDM Functions – Most Time-Consuming

From the graph we can conclude that the MDM functions which are the most time-consuming are *Master Data Enrichment and Cleansing*, *Master Data Creation and Maintenance*, and *Master Data Classification and Taxonomy Management*. It is also worth noting that none of the research participants deemed the *Master Data Roles and Authorisation Management* as an MDM function that is most time-consuming.

Analysis of Survey Findings: Master Data Management Functions – Combined Findings

The following graph summarises the survey findings for each of the Master Data Management (MDM) functions. Each of the previous dimensions of the survey findings, namely Impact on Master Data Quality, Most Human Involvement Required, and Most Time-Consuming, are presented below for each MDM function. The results are grouped according to the highest survey responses for the combination of all three questions relating to MDM functions.

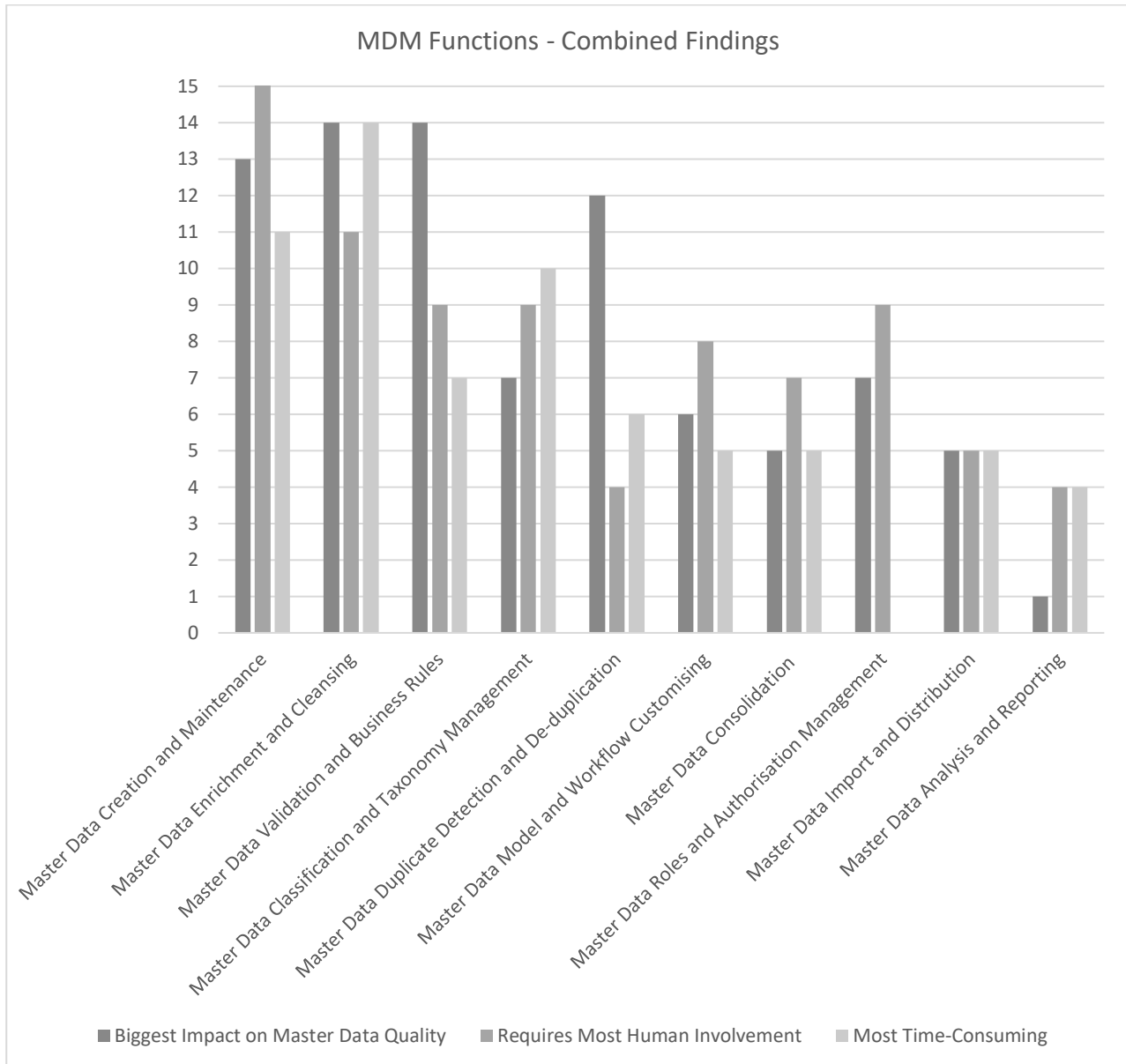


Figure 27: Analysis of Survey Findings: MDM Functions – Combined Findings

From the diagram based on the survey response, we can see that though the MDM functions *Master Data Enrichment and Cleansing* and *Master Data Validation and Business Rules* have the biggest impact on master data quality, *Master Data Creation and Maintenance* requires the most human involvement, making it the most prominent MDM Function extracted from the literature based on the survey findings.

6.3 EVALUATION OF IDENTIFIED ARTIFICIAL INTELLIGENCE CAPABILITIES APPLICABLE TO MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT

The findings from the second Systematic Literature Review (SLR), conducted in Chapter 5 to identify the primary capabilities of Artificial Intelligence (AI) that can be applied to Master Data Management (MDM) and Data Quality Management (DQM), were presented to a group of experts in MDM to evaluate the quality of the SLR findings. The survey questions were based on the conclusions of this SLR in the form of the AI Capabilities for MDM and DQM Framework, which can be seen in section 5.5.3.

The survey questions were designed with the DOI theory to establish the perceived difficulty and complexity of adopting AI capabilities in an MDM system. The perceived advantage and usefulness of adopting AI into MDM systems were also established by the responses. All the respondents indicated that an AI-Enabled MDM reference framework would be applicable to industry professionals as a guide for implementing AI tools into MDM systems.

A survey data collection method was followed, and a Google Forms survey was created to conduct the research. The research participants were 18 data consultants who have expertise in Master Data Management.

6.3.1 SURVEY OUTLINE – ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT

The survey questions applicable to the AI capabilities identified were the following:

- Which AI Capabilities do you think are the most important for organisations to adopt to improve Master Data quality in the ERP system?
- Which AI capabilities would be most easily implemented into a Master Data Management system?
- Would an AI-Enabled Master Data Management reference framework be applicable to industry professionals as guide for implementing AI tools into MDM systems?

The response options for each of the above questions were the same options that can be viewed below, and the research participants were able to select all options from this list that were applicable:

- Automated Master Data Maintenance
- Mass Data Maintenance
- Data Attributes Prediction
- Master Data Duplicate Detection and De-Duplication
- Master Data Error Detection and Correction
- Identification of Inconsistent and Missing Master Data Fields
- Automated Master Data Harmonising, Enrichment and Cleansing
- Automated Master Data Quality Assessment and Checks
- Master Data Analysis and Monitoring
- Master Data Pattern and Anomaly Detection

- Master Data Cataloguing

The following section displays the survey responses captured for each of the questions.

6.3.2 SURVEY RESPONSES – ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT

Survey Responses: Artificial Intelligence Capabilities – Most Important for Organisations to Adopt

The first question related to the AI Capabilities applicable to MDM and DQM on the survey was:

- Which AI Capabilities do you think are the most important for organisations to adopt to improve Master Data quality in the ERP system?

The result from this research survey question was extracted from the Google Form and is displayed in the below figure.

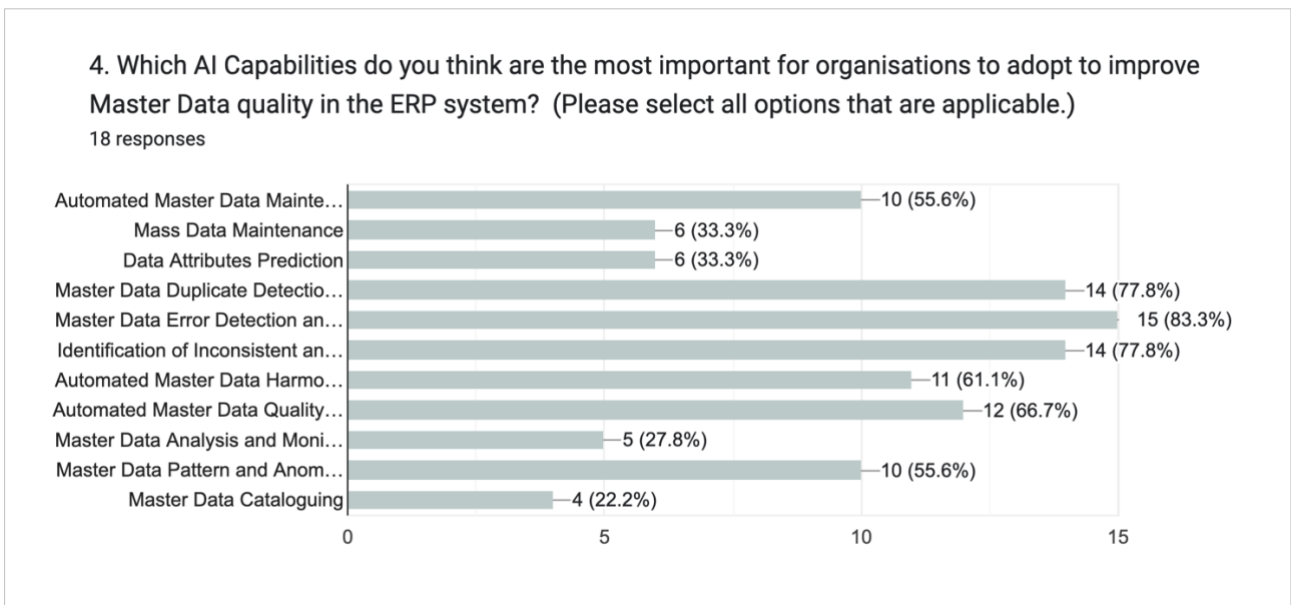


Figure 28: AI Capabilities – Most Important for Organisations to Adopt

Survey Responses: Artificial Intelligence Capabilities – Most Easily Implemented into a Master Data Management System

The second question related to the AI Capabilities applicable to MDM and DQM on the survey was:

- Which AI capabilities would be most easily implemented into a Master Data Management system?

The result from this research survey question was extracted from the Google Form and is displayed in the below figure.

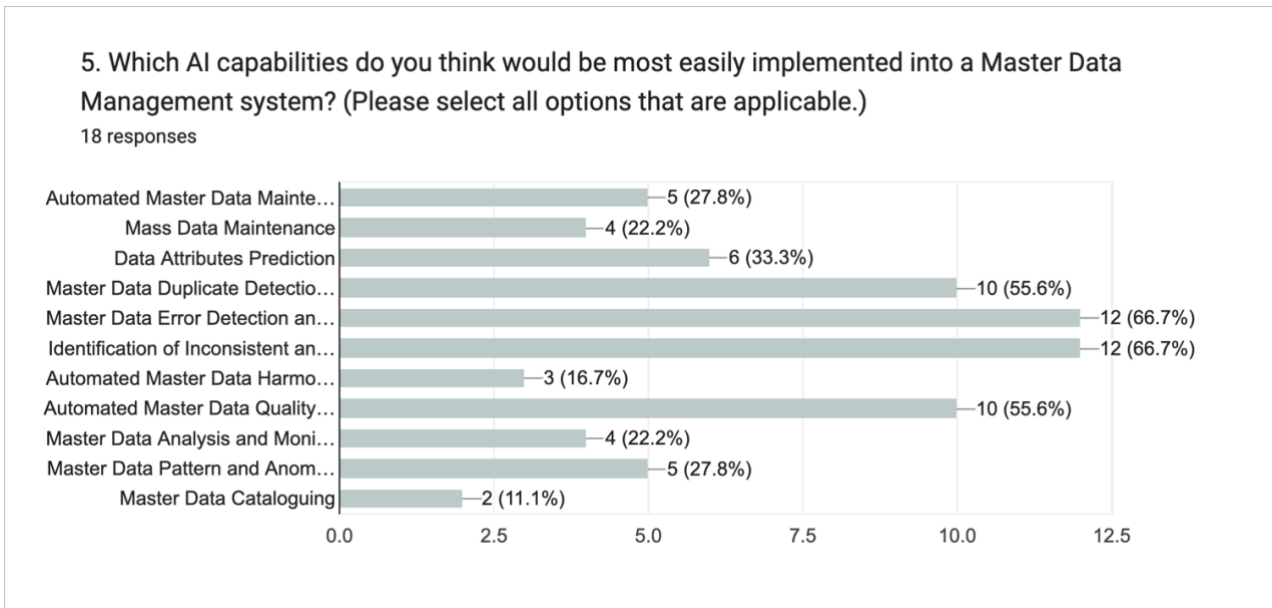


Figure 29: AI Capabilities – Most Easily Implemented into an MDM System

Survey Responses: Artificial Intelligence Capabilities – Usefulness of an AI-Enabled MDM Reference Framework

The third question related to the AI Capabilities applicable to MDM and DQM on the survey was:

- Would an AI-Enabled Master Data Management reference framework apply to industry professionals as guide for implementing AI tools into MDM systems?

The result from this research survey question was extracted from the Google Form and is displayed in the below figure.

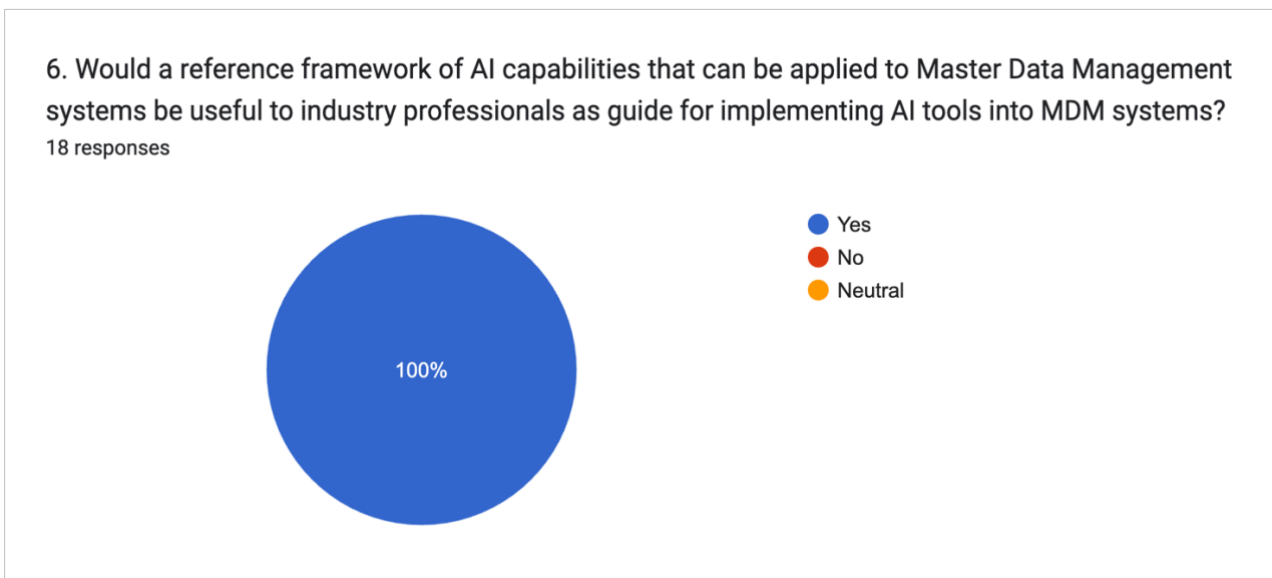


Figure 30: AI Capabilities – Usefulness of an AI-Enabled MDM Reference Framework

6.3.3 ANALYSIS OF SURVEY FINDINGS – ARTIFICIAL INTELLIGENCE CAPABILITIES FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT

In this section, the survey questions related to Artificial Intelligence (AI) Capabilities are analysed and displayed in graphs to highlight each question's most popular response options. These findings will be compared with the trends in literature for AI Capabilities that can be seen in section 5.5.4 and combined with the survey findings and literature trends of the identified Master Data Management (MDM) functions in the previous section to create the final research artefact in Chapter 7.

Analysis of Survey Findings: Artificial Intelligence Capabilities – Most Important for Organisations to Adopt

The following graph visualises the AI capabilities identified from the literature as the most important for organisations to adopt, according to the data management professionals who participated in the research survey.

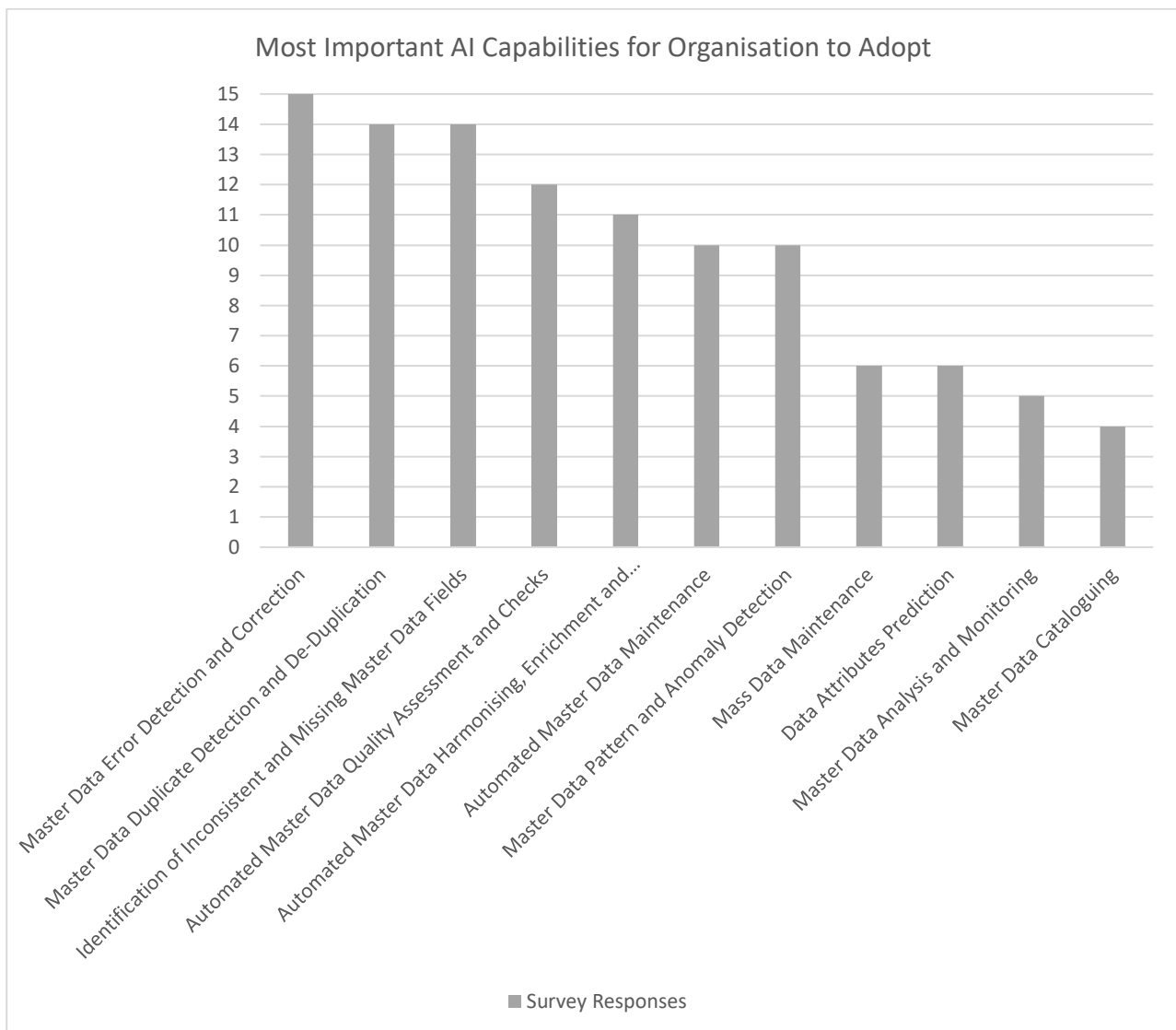


Figure 31: Analysis of Survey Findings: AI Capabilities – Most Important for Organisations to Adopt

From this graph, we can see that the AI capabilities that are most important for organisations to adopt are *Master Data Error Detection and Correction*, *Master Data Duplicate Detection and De-duplication*, and *Identification of Inconsistent and Missing Master Data Fields*.

Analysis of Survey Findings: Artificial Intelligence Capabilities – Easily Implemented into an MDM System

The following graph provides a visualisation of the AI capabilities identified from the literature that are most easily implemented into an MDM system, according to the data management professionals who participated in the research survey.

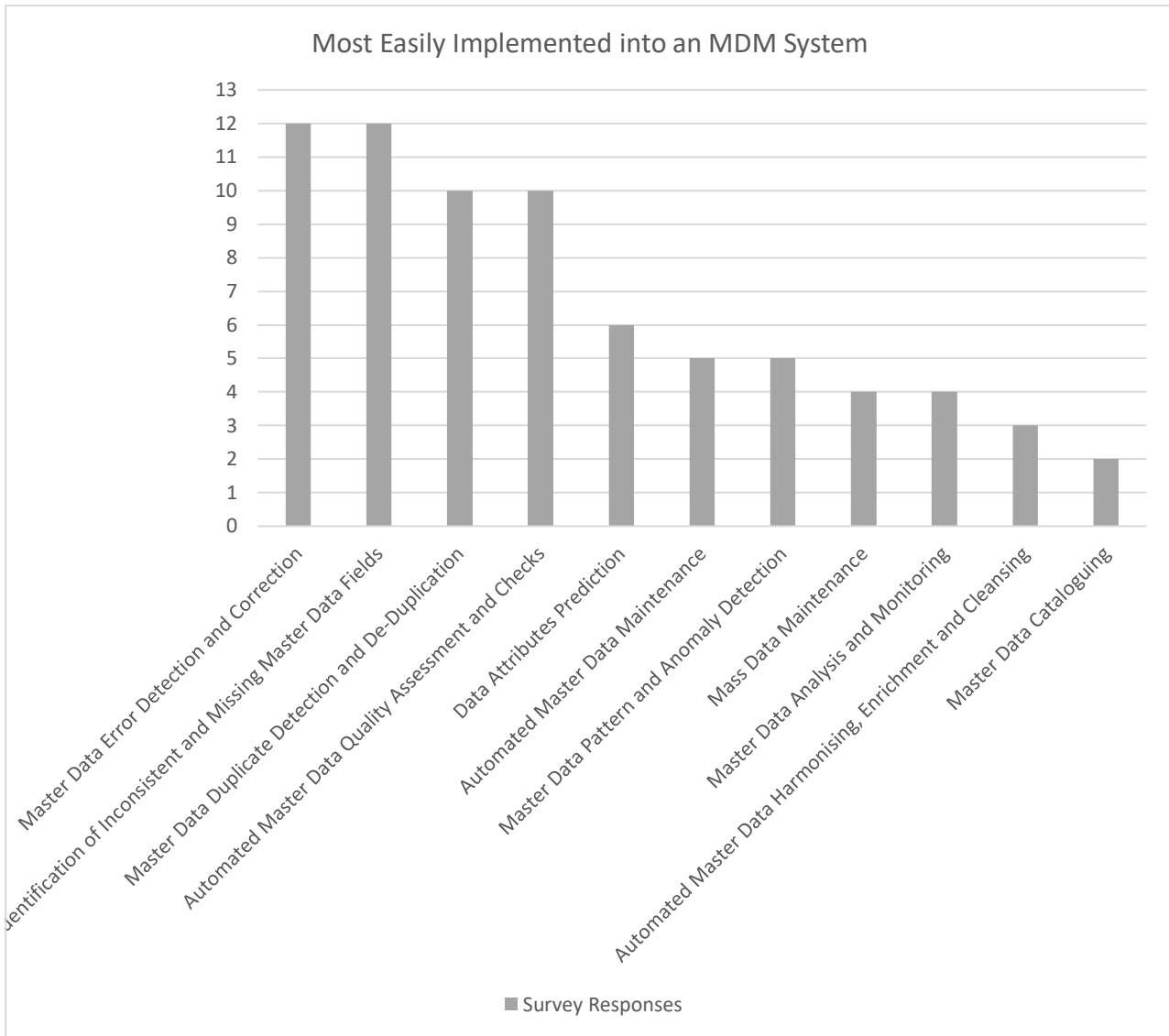


Figure 32: Analysis of Survey Findings: AI Capabilities – Easily Implemented into an MDM System

From this graph, we can see that the AI capabilities that are easiest for organisations to implement are *Master Data Error Detection and Correction*, *Identification of Inconsistent and Missing Master Data Fields*, and *Master Data Duplicate Detection and De-Duplication*.

Analysis of Survey Findings: Artificial Intelligence Capabilities – Combined Findings

The graph below combines the graphs in the previous sections and displays the most important AI capabilities for organisations to adopt. It is integrated with the graph that displays the AI capabilities that are most easily implemented into an MDM system. The results are grouped according to the highest survey responses for the combination of the questions relating to AI capabilities.

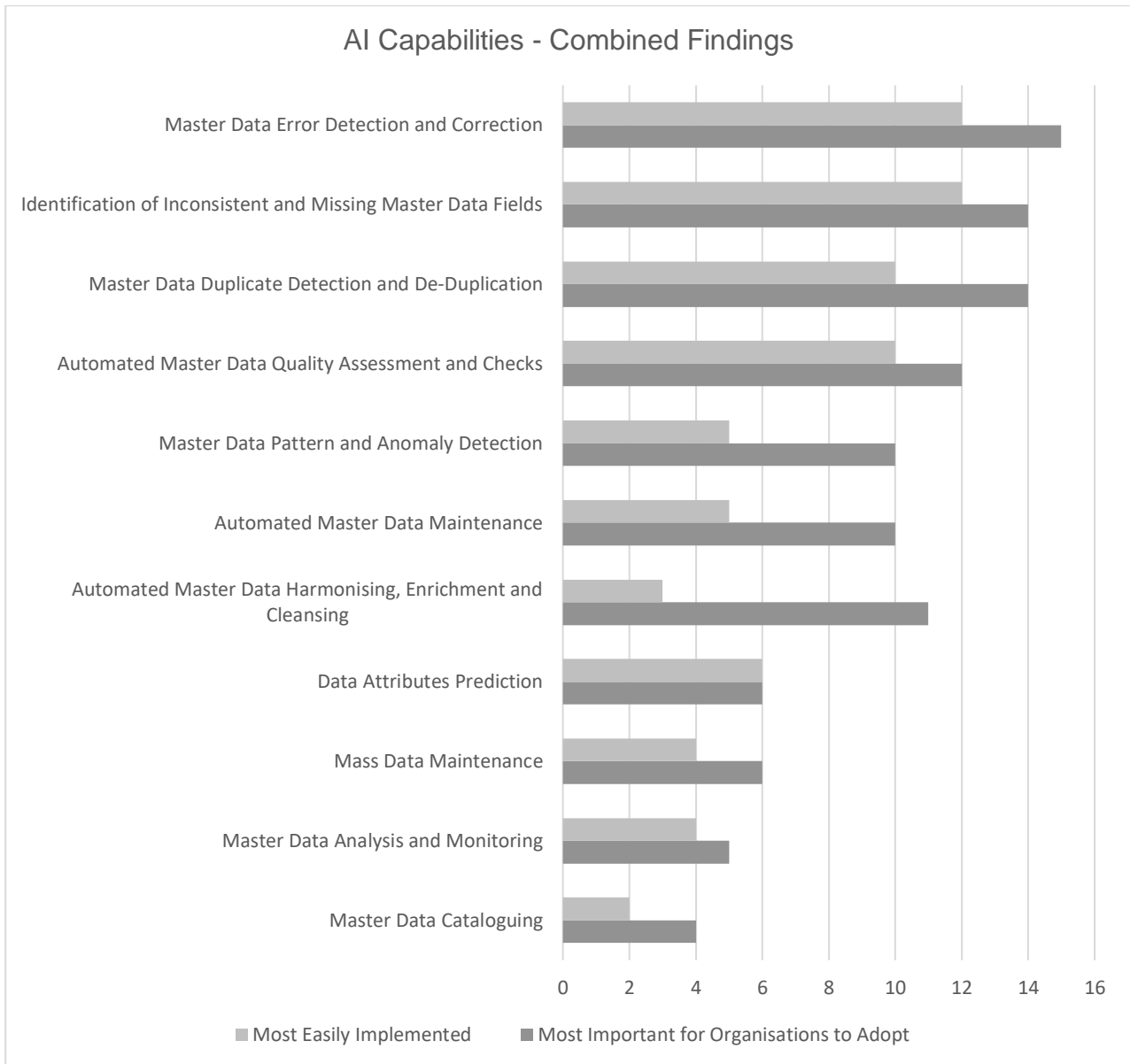


Figure 33: Analysis of Survey Findings: AI Capabilities – Combined Findings

From the diagram based on the survey response, we can see that the top 3 AI capabilities that are most easily implemented and most important for organisations to adopt, according to data management professionals, are *Master Data Error Detection and Correction*, *Identification of Inconsistent and Missing Master Data Fields*, and *Master Data Duplicate Detection and De-Duplication*.

6.4 CONCLUSION

This section evaluated the findings of the two respective SLRs carried out in Chapter 4 and Chapter 5. The *Master Data Management (MDM) Functions* and *Artificial Intelligence (AI) Capabilities Applicable to Master Data Management (MDM) and Data Quality Management (DQM)* identified from the literature were presented to data management professionals with corresponding questions in the form of a survey, which was then analysed.

The purpose of the survey was to determine the impact of each MDM function on master data quality, establish which MDM functions require the most human involvement and which are the most time-consuming, and evaluate which AI capabilities are the most important for organisations to adopt and which could be most easily implemented into an MDM system. The survey questions were designed with the diffusion of innovation (DOI) theoretical lens to establish the general complexity and perceived advantage for organisations to adopt AI tools in the context of MDM systems.

Based on the survey responses, the diagrams established which MDM functions have the most significant impact on master data quality, require the most human involvement and are most time-consuming. We could also see that the top three AI capabilities which are most easily implemented and most important for organisations to adopt, according to the data management professionals, are *Master Data Error Detection and Correction*, *Identification of Inconsistent and Missing Master Data Fields*, and *Master Data Duplicate Detection and De-Duplication*.

All the data management professionals indicated that an AI-Enabled Master Data Management reference framework would be helpful to industry professionals as guide for implementing AI tools into MDM systems. In the next chapter of the research, the final research artefact is constructed by utilising the findings from this chapter to create the *AI-Enabled MDM Framework* in section 7.3.

7 AI-ENABLED MDM FRAMEWORK – RESEARCH ARTEFACT

The figure below shows the layout and position of Chapter 7 in relation to the rest of the research paper.

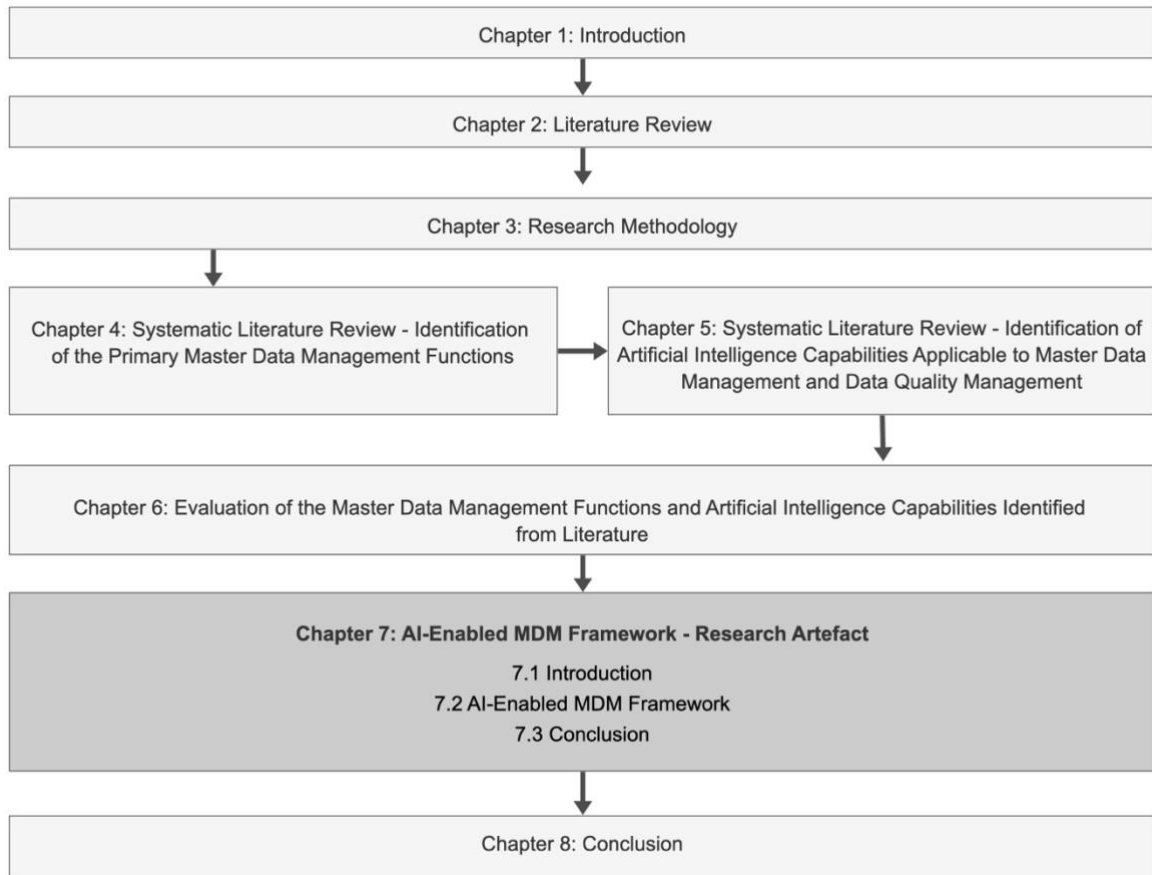


Figure 34: Structure of Research Paper - Position of Chapter 7

7.1 INTRODUCTION

This section produces the final research artefact by utilising the survey responses and combining the findings to create the *AI-Enabled MDM Framework*. The most prominent MDM functions and AI capabilities identified for MDM and DQM are logically mapped to produce the final research artefact, *AI-Enabled MDM Framework* that categorises MDM functions and tasks to appropriate AI data management capabilities and provides suggested AI tools and methods identified from the literature, as can be viewed in section 7.3.

7.2 AI-ENABLED MDM FRAMEWORK

This section contains the research findings and the final research artefact called the *AI-Enabled MDM Framework*. The research findings are presented in tables. The first table classifies the Master Data

Management (MDM) functions identified from the literature in order of prominence based on the number of times each MDM function was mentioned and the percentage of survey responses per survey question for each identified MDM function. The second table contains the most prominent AI capabilities for MDM and DQM that have been established by classifying the AI data management capabilities identified according to the highest percentages in the survey responses and the highest number of references in the literature. The third table logically maps the MDM functions and the AI capabilities identified. The *AI-Capabilities Framework* contains the logical groupings of the MDM functions to AI capabilities based on the functional architecture for master data management Pontello et al. (2021). The MDM functions are grouped into *Master Data Lifecycle Management*, *Master Data Quality Assurance*, *Master Data Cross Functions*, and *Master Data Administration*. The primary design science research contribution is presented in section 7.3, namely the *AI-Enabled MDM Framework* artefact.

7.2.1 MOST PROMINENT MASTER DATA MANAGEMENT FUNCTIONS IDENTIFIED

The most prominent Master Data Management (MDM) functions have been established by classifying the MDM functions identified according to the highest percentages in the survey responses and the highest number of references in the literature. The following table contains the organised list of MDM functions in order of prominence. The most prominent MDM function identified is *Master Data Creation and Maintenance*, and the least prominent capability identified is *Master Data Analysis and Reporting*.

MDM Functions	Survey Responses			Literature References	
	Biggest Impact on Master Data Quality	Requires Most Human Involvement	Most Time-Consuming	Number of References	References
Master Data Creation and Maintenance	13 (72.2%)	16 (88.9%)	11 (61.1%)	12	(Maedche, 2010), (Knapp & Hasibether, 2011), (Zhang & Ye, 2011), (Knolmayer & Röthlin, 2006), (O'Farrell et al., 2013), (Szívós & Orosz, 2014), (Schäffer & Leyh, 2017), (Vilminko-Heikkinen & Pekkola, 2019), (Weber et al., 2009), (Tambo, 2018), (Malihi et al., 2016), (Prokhorov & Kolesnik, 2018)
Master Data Enrichment and Cleansing	14 (77.8%)	11 (61.1%)	14 (77.8%)	5	(Maedche, 2010), (Prokhorov & Kolesnik, 2018), (Zhang & Ye, 2011), (Knapp & Hasibether, 2011), (Grangel-González et al., 2023)
Master Data Validation	14 (77.8%)	9 (50%)	7 (38.9%)	7	(Knolmayer & Röthlin, 2006), (Knapp & Hasibether, 2011), (Szívós & Orosz, 2014), (Schäffer

MDM Functions	Survey Responses			Literature References	
	Biggest Impact on Master Data Quality	Requires Most Human Involvement	Most Time-Consuming	Number of References	References
and Business Rules					& Leyh, 2017), (Grangel-González et al., 2023), (Prokhorov & Kolesnik, 2018), (Malihi et al., 2016)
Master Data Duplicate Detection and De-duplication	12 (66.7%)	4 (22.2%)	6 (33.3%)	9	(Knolmayer & Röthlin, 2006), (Knapp & Hasibether, 2011), (Prokhorov & Kolesnik, 2018), (Maedche, 2010), (Zong et al., 2017), (Vilminko-Heikkinen & Pekkola, 2019), (Harrison et al., 2019), (Szívós & Orosz, 2014), (Siren et al., 2015)
Master Data Classification and Taxonomy Management	7 (38.9%)	9 (50%)	10 (55.6%)	4	(Knapp & Hasibether, 2011), (Zong et al., 2017), (Ender Hancioglu, 2020), (Prokhorov & Kolesnik, 2018)
Master Data Model and Workflow Customising	6 (33.3%)	8 (44.4%)	5 (27.8%)	8	(Zhang & Ye, 2011), (O'Farrell et al., 2013), (Maedche, 2010), (Vilminko-Heikkinen & Pekkola, 2019), (Szívós & Orosz, 2014), (Prokhorov & Kolesnik, 2018), (Weber et al., 2009), (Dittrich et al., 2009)
Master Data Import and Distribution	5 (27.8%)	5 (27.8%)	5 (27.8%)	9	(Knolmayer & Röthlin, 2006), (Maedche, 2010), (Prokhorov & Kolesnik, 2018), (Vilminko-Heikkinen & Pekkola, 2019) (Grangel-González et al., 2023), (Siren et al., 2015), (Schäffer & Leyh, 2017), (Dittrich et al., 2009), (Zhang & Ye, 2011)
Master Data Roles and Authorisation Management	7 (38.9%)	9 (50%)	0 (0)	7	(Prokhorov & Kolesnik, 2018), (Weber et al., 2009), (Maedche, 2010), (Siren et al., 2015), (Vilminko-Heikkinen & Pekkola, 2019), (Dittrich et al., 2009), (Szívós & Orosz, 2014)
Master Data Consolidation	5 (27.8%)	7 (38.9%)	5 (27.8%)	5	(Maedche, 2010), (Szívós & Orosz, 2014), (Siren et al., 2015), (Prokhorov & Kolesnik, 2018),

MDM Functions	Survey Responses			Literature References	
	Biggest Impact on Master Data Quality	Requires Most Human Involvement	Most Time-Consuming	Number of References	References
					(Vilminko-Heikkinen & Pekkola, 2019)
Master Data Analysis and Reporting	1 (5.6%)	4 (22.2%)	4 (22.2%)	1	(Prokhorov & Kolesnik, 2018)

Table 14: Most Prominent MDM Functions

7.2.2 MOST PROMINENT ARTIFICIAL INTELLIGENCE CAPABILITIES IDENTIFIED FOR MASTER DATA MANAGEMENT AND DATA QUALITY MANAGEMENT

The most prominent AI Capabilities for MDM and DQM have been established by classifying the AI Data Management Capabilities according to the highest percentages in the survey responses and the highest number of references in the literature. The following table contains an organised list of AI Data Management Capabilities in order of prominence. The most prominent AI Data Management Capability identified is *Master Data Duplicate Detection and De-Duplication*, and the least prominent capability identified is *Master Data Cataloguing*.

AI Data Management Capabilities Identified	Survey Responses		Literature References	
	Most Important for Organisations to Adopt	Most Easily Implemented into an MDM System	Number of References	References
Master Data Duplicate Detection and De-Duplication	14 (77,8%)	10 (55,6%)	6	(Riesener et al., 2022), (Leary, 2014), (Zhang et al., 2021), (Guo et al., 2018), (Spreitzenbarth et al., 2024), (Liu & Anderson, 2024)
Master Data Error Detection and Correction	15 (83,3%)	12 (66,7%)	3	(Zhang et al., 2021), (Tverdal et al., 2024), (Rangineni et al., 2023)
Identification of Inconsistent and Missing Master Data Fields	14 (77,8%)	12 (66,7%)	2	(Pakkala et al., 2024), (Jonathan & Frank, 2023)
Automated Master Data Quality Assessment and Checks	12 (66,7%)	10 (55,6%)	1	(Rangineni et al., 2023)
Master Data Pattern and Anomaly Detection	10 (55,6%)	5 (27,8%)	3	(Liu & Anderson, 2024), (Rangineni et al., 2023), (Jonathan & Frank, 2023)
Automated Master Data Harmonising, Enrichment and Cleansing	11 (61,1%)	3 (16,7%)	2	(Pakkala et al., 2024), (Jesmeen et al., 2018)
Automated Master Data Maintenance	10 (55,6%)	5 (27,8%)	1	(Riesener et al., 2022)

AI Data Management Capabilities Identified	Survey Responses		Literature References	
	Most Important for Organisations to Adopt	Most Easily Implemented into an MDM System	Number of References	References
Data Attributes Prediction	6 (33,3%)	6 (33,3%)	3	(Riesener et al., 2022), (Pakkala et al., 2024), (Rangineni et al., 2023)
Master Data Analysis and Monitoring	5 (27,8%)	4 (22,2%)	2	(Kulkov, 2021), (Jonathan & Frank, 2023)
Mass Data Maintenance	6 (33,3%)	4 (22,2%)	1	(Riesener et al., 2022)
Master Data Cataloguing	4 (22,2%)	2 (11,1%)	2	(Jonathan & Frank, 2023), (Roth et al., 2022)

Table 15: Most Prominent AI Data Management Capabilities

7.2.3 LOGICAL MAPPING OF ARTIFICIAL INTELLIGENCE CAPABILITIES TO MASTER DATA MANAGEMENT FUNCTIONS

The following table logically maps the Master Data Management functions identified from the literature to the AI capabilities applicable to Master Data Management and Data Quality Management. This mapping forms part of the basis for the final research artefact created, namely the *AI-Enabled MDM Framework*, which can be viewed in section 7.3.

		AI Capabilities Applicable to MDM and DQM										
		Master Data Duplicate Detection and De-Duplication	Master Data Error Detection and Correction	Identification of Inconsistent and Missing Master Data Fields	Automated Master Data Quality Assessment and Checks	Master Data Pattern and Anomaly Detection	Automated Master Data Harmonising, Enrichment and Cleansing	Automated Master Data Maintenance	Data Attributes Prediction	Master Data Analysis and Monitoring	Mass Data Maintenance	Master Data Cataloguing
MDM Functions												
Master Data Creation and Maintenance				X				X	X		X	X
Master Data Enrichment and Cleansing				X	X	X	X		X			
Master Data Validation and Business Rules			X	X	X				X			
Master Data Duplicate Detection and De-duplication		X								X	X	X
Master Data Classification and Taxonomy Management								X			X	X

AI Capabilities Applicable to MDM and DQM												
MDM Functions	Master Data Duplicate Detection and De-Duplication	Master Data Error Detection and Correction	Identification of Inconsistent and Missing Master Data Fields	Automated Master Data Quality Assessment and Checks	Master Data Pattern and Anomaly Detection	Automated Master Data Harmonising, Enrichment and Cleansing	Automated Master Data Maintenance	Data Attributes Prediction	Master Data Analysis and Monitoring	Mass Data Maintenance	Master Data Cataloguing	
Master Data Model and Workflow Customising							X			X		
Master Data Import and Distribution												
Master Data Roles and Authorisation Management									X			
Master Data Consolidation	X			X						X		
Master Data Analysis and Reporting									X			

Table 16: Logical Mapping of AI Capabilities to MDM Functions

In the following section, the Master Data Management functions are grouped according to a functional architecture in research Pontello et al. (2021). It is important to note that the MDM Function category *Master Data Import and Distribution* does not match any of the identified AI Capabilities and was subsequently removed from the final research artefact.

7.2.4 AI CAPABILITIES MAPPED TO CORRESPONDING MASTERD DATA MANAGEMENT FUNCTIONS

The *AI-Capabilities Framework* below contains the logical groupings of the MDM functions to AI capabilities based on the functional architecture for Master Data Management by Pontello et al. (2021).

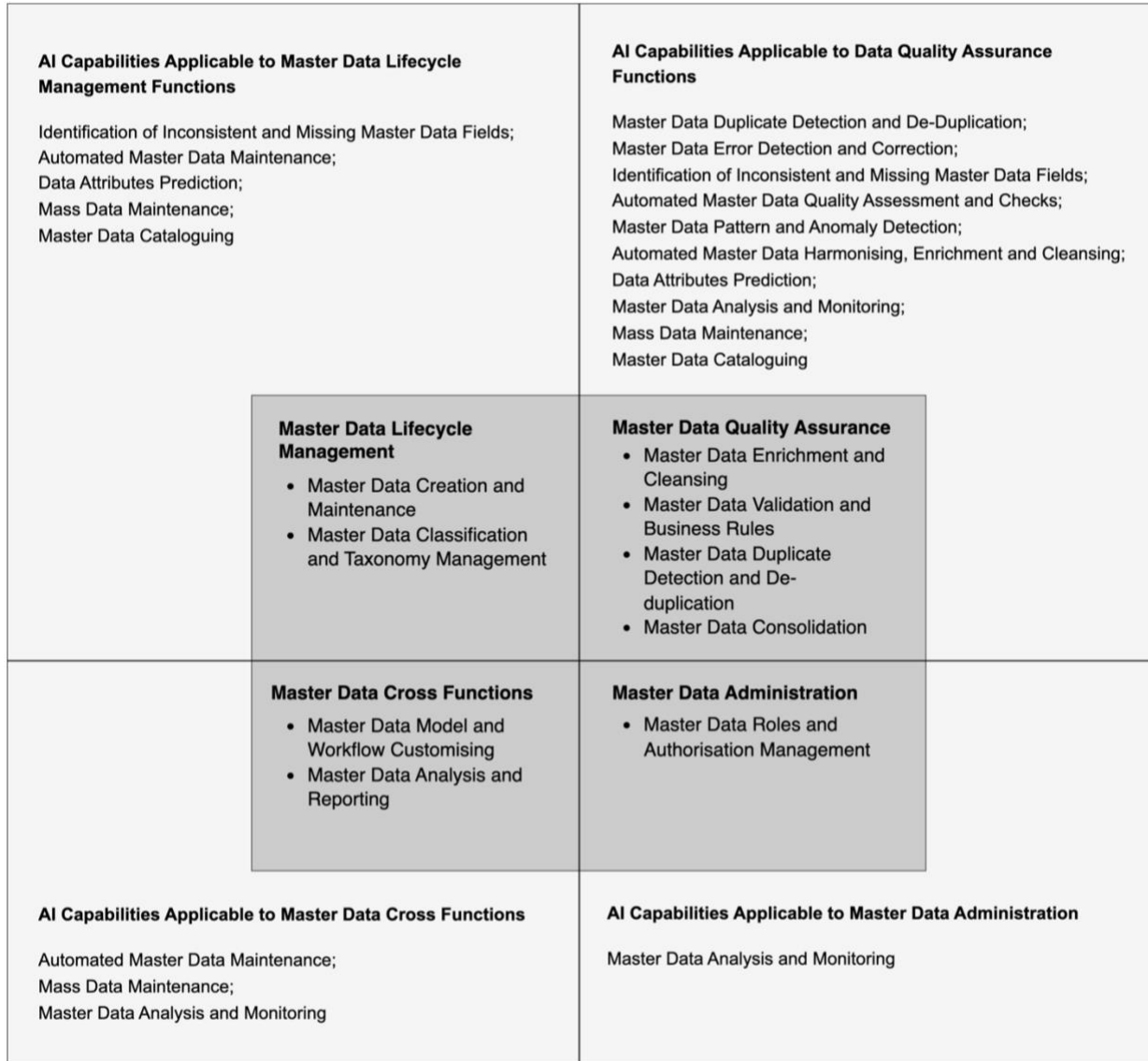


Figure 35: AI-Capabilities Framework

The MDM functions are grouped into *Master Data Lifecycle Management*, *Master Data Quality Assurance*, *Master Data Cross Functions*, and *Master Data Administration*. The Master Data Lifecycle Management grouping contains the *Master Data Creation and Maintenance* and *Master Data Classification and Taxonomy Management* MDM functions. Master Data Quality Assurance grouping contains *Master Data Enrichment and Cleansing*, *Master Data Validation and Business Rules*, *Master Data Duplicate Detection and De-duplication*, and *Master Data Consolidation* MDM functions. The Master Data Cross Functions grouping encompasses the *Master Data Model and Workflow Customising* and

Master Data Analysis and Reporting MDM function. The Master Data Administration grouping incorporates the *Master Data and Authorisation Management* MDM function. The MDM function *Master Data Import and Distribution* has been left out of the final artefact since no AI capability is identified from literature that matches this function.

The AI capabilities applicable to the MDM function category *Master Data Lifecycle Management* comprise *Identification of Inconsistent and Missing Master Data Fields*, *Automated Master Data Maintenance*, *Data Attributes Prediction*, *Mass Data Maintenance*, and *Master Data Cataloguing* AI capabilities. The AI capabilities applicable to the MDM function category *Master Data Quality Assurance* comprises *Master Data Enrichment and Cleansing*, *Master Data Validation and Business Rules*, *Master Data Duplicate Detection and De-duplication* and *Master Data Consolidation*.

The AI capabilities applicable to the MDM function category *Master Data Cross Functions* comprise *Master Data Model and Workflow Customising* and *Master Data Analysis and Reporting*. The AI capability applicable to the final MDM function category, *Master Data Administration*, is *Master Data Roles and Authorisation Management*.

7.3 AI-ENABLED MASTER DATA MANAGEMENT FRAMEWORK

This section comprises the primary research contribution, which was created based on the findings and frameworks developed in the previous sections. It describes the artificial intelligence (AI) tools and methods identified from the literature that can be applied to each of the AI capabilities identified that apply to the Master Data Management (MDM) functions identified by the literature.

The research findings' usefulness as a design science research (DSR) artefact is motivated by four case studies of how the *AI-Enabled MDM Framework* can be applied to specific organisational master data issues. Industry professionals can use this research artefact to assess which MDM functions can be enhanced by helping them discover an adequate AI tool or method to apply to the corresponding MDM function issue that the organisation is facing. The DSR artefact also serves as a substantial research finding for future research. Future researchers in MDM and AI can utilise the findings from this study to further investigate the impact of AI on the field of MDM, as discussed in the following sections.

The following table consists of the *Master Data Management Functions*, the *Functional Grouping* based on Pontello et al. (2021), the AI capabilities that were identified as applying to these MDM functions, and the AI tools and methods that were identified from literature that can be applied to these MDM functions for enhanced data quality in an ERP system that makes use of MDM for master data quality management. The AI tools and methods suggested range from specific approaches, such as utilising Bayesian probabilistic methods for matching master data models, to using natural language understanding for mapping similar master data fields in a master data model (Behera & Panda, 2023).

MDM Function Grouping	MDM Functions	AI Capabilities Relevant to the MDM Functions	AI Tools and Methods Identified from Literature	References
Master Data Lifecycle Management	Master Data Creation and Maintenance; Master Data Classification and Taxonomy Management	Identification of Inconsistent and Missing Master Data Fields	Inconsistencies and hidden relationships in master data can be identified using unsupervised Machine Learning (ML) techniques such as clustering and anomaly detection (Liu & Anderson, 2024). Data inaccuracies and inconsistencies can be predicted by employing Machine Learning techniques such as predictive analytical models (Rangineni et al., 2023).	(Liu & Anderson, 2024); (Rangineni et al., 2023)
		Automated Master Data Maintenance	Master data maintenance tasks such as data quality assessment can be automated by employing AI tools such as deep-learning methods (Rangineni et al., 2023). Natural Language Processing (NLP) techniques can be employed to master data searching functionality to better understand user queries and retrieve correct master data records (Jonathan & Frank, 2023).	(Rangineni et al., 2023); (Jonathan & Frank, 2023)
		Data Attributes Prediction	Utilising Machine Learning algorithms such as a random forest based regression algorithm to predict and classify data attributes.	(Riesener et al., 2022)
		Mass Data Maintenance	Large-scale master data processing can be assisted with deep learning methods which have the potential to automate data quality assessment tasks, according to Rangineni et al. (2023).	(Rangineni et al., 2023)
		Master Data Cataloguing	Automated data acquisition and categorisation is made possible with a Machine Learning engine, employing Named Entity Recognition (NER) and Natural Language Understanding (NLU), which can accurately map master data into master data model structures.	(Behera & Panda, 2023)
Master Data Quality Assurance	Master Data Enrichment and Cleansing; Master Data Validation and Business Rules; Master Data Duplicate Detection and De-duplication; Master Data Consolidation	Master Data Duplicate Detection and De-Duplication	Categorisation of the Master Data utilising Machine Learning techniques such as Named Entity Recognition (NER) and Natural Language Understanding (NLU) is suggested by Behera and Panda (2023). These methods assist in discovering similar fields in master data models and can be used to detect and correct duplicate data in the MDM system. Machine Learning techniques such as a data matching and de-duplication is suggested by Liu and Anderson (2024). Pattern and anomalies in master data can be recognised by employing supervised learning programs such as support vector machines and random forests, which enables better data matching and de-duplication.	(Behera & Panda, 2023); (Liu & Anderson, 2024)
		Master Data Error Detection and Correction	Employing Machine Learning techniques to Master Data Management present predictive abilities for anomaly detection and data error detection (Rangineni et al., 2023).	(Rangineni et al., 2023)
		Identification of Inconsistent and Missing Master Data Fields	Inconsistencies and hidden relationships in master data can be identified using unsupervised Machine Learning (ML) techniques such as clustering and anomaly detection (Liu & Anderson, 2024). Data inaccuracies and inconsistencies can be predicted by employing Machine Learning techniques such as predictive analytical models (Rangineni et al., 2023).	(Liu & Anderson, 2024); (Rangineni et al., 2023)
		Automated Master Data Quality Assessment and Checks	Pontello et al. (2021) suggest utilising a supervised learning method such as a regression model to assist in improved quality management of master data.	(Pontello et al., 2021)
		Master Data Pattern and Anomaly Detection	Machine Learning tools can be applied to identify patterns in master data residing in large databases and in estimating new data attributes (Riesener et al., 2022). Employing supervised learning programs such as support vector machines and random forests for data matching is suggested by Liu and Anderson (2024), which can assist in data pattern and anomaly detection. AI techniques such as Machine Learning and Deep Learning algorithms can be applied to master data for enhanced data anomaly detection and pattern prediction, according to Jonathan and Frank (2023). Deep Learning Models such as Neural Networks (NN) and Recurrent Neural Networks (RNN) show strong potential for assisting in real-time master data anomaly detection (Jonathan & Frank, 2023).	(Riesener et al., 2022); (Liu & Anderson, 2024); (Jonathan & Frank, 2023)

MDM Function Grouping	MDM Functions	AI Capabilities Relevant to the MDM Functions	AI Tools and Methods Identified from Literature	References
		Automated Master Data Harmonising, Enrichment and Cleansing	As an AI tool for data cleansing, Pontello et al. (2021) suggest utilising a supervised learning method such as a regression model to assist in improved quality management of master data.	(Pontello et al., 2021)
		Data Attributes Prediction	Utilising Machine Learning algorithms such as a random forest based regression algorithm to predict and classify data attributes.	(Riesener et al., 2022)
		Master Data Analysis and Monitoring	Employing the AI method K-means clustering, which is used for pattern recognition and clustering, to identify correlations in master data, as per the paper by Pontello et al. (2021).	(Pontello et al., 2021)
		Mass Data Maintenance	Deep learning methods have the potential to automate data quality assessment tasks according to Rangineni et al. (2023). These deep learning methods can also assist in large-scale master data processing.	(Rangineni et al., 2023)
		Master Data Cataloguing	Automated data acquisition and categorisation is made possible with a Machine Learning engine, employing Named Entity Recognition (NER) and Natural Language Understanding (NLU), which can accurately map master data into master data model structures.	(Behera & Panda, 2023)
Master Data Administration	Master Data Roles and Authorisation Management	Master Data Analysis and Monitoring	Employing the AI method K-means clustering, which is used for pattern recognition and clustering, to identify correlations in master data, as per the paper by Pontello et al. (2021).	(Pontello et al., 2021)
Master Data Cross Functions	Master Data Model and Workflow Customising; Master Data Analysis and Reporting	Automated Master Data Maintenance	Master data maintenance tasks such as data quality assessment can be automated by employing AI tools such as deep-learning methods (Rangineni et al., 2023). Natural Language Processing (NLP) techniques can be employed to master data searching functionality to better understand user queries and retrieve correct master data records (Jonathan & Frank, 2023).	(Rangineni et al., 2023); (Jonathan & Frank, 2023)
		Mass Data Maintenance	According to Rangineni et al. (2023), large-scale master data processing can be enhanced with deep learning methods which have the potential to automate data quality assessment tasks.	(Rangineni et al., 2023)
		Master Data Analysis and Monitoring	Employing the AI method K-means clustering, which is used for pattern recognition and clustering, to identify correlations in master data, as per the paper by Pontello et al. (2021).	(Pontello et al., 2021)

Table 17: AI-Enabled MDM Framework (Design Science Research Artefact)

7.3.1 THE AI-ENABLED MDM FRAMEWORK AS A USABLE DSR ARTEFACT FOR INDUSTRY PROFESSIONALS

This section motivates why the *AI-Enabled MDM Framework* is an adequate tool for industry professionals to use in discovering how AI tools can be applied in the field of Master Data Management for enhanced master data quality. The framework acts as a helpful tool in assessing which MDM functions can be enhanced by helping them discover an adequate AI tool or method to apply to the corresponding MDM function issue that the organisation is facing.

In addition to using the framework as a tool for matching AI Capabilities with the AI Tools and Methods identified from literature that can be investigated as a potential solution for the organisation's specific MDM function issue, the framework also serves as a substantial research finding that can assist future researchers in the field of Master Data Management and Artificial Intelligence.

Case Study 1: Application of the AI-Enabled MDM Framework in the Scenario of Multiple Duplicate Master Data Records

Scenario:

The following describes how the *AI-Enabled MDM framework* can be applied in the scenario where a company is making use of an MDM system and finds that there are numerous duplicate master data records in the system, such as multiple versions of a customer record with varying data that need to be identified and consolidated.

Utilising the AI-Enabled MDM Framework:

In this scenario, a data management industry professional can use the *AI-Enabled MDM Framework* to determine which AI capabilities may match the organisation's problem.

The *MDM Function Grouping* will be *Master Data Quality Assurance*. The corresponding *MDM Functions* will be *Master Data Duplicate Detection and De-duplication* and *Master Data Enrichment and Cleansing*, and the corresponding *AI Capabilities Relevant to the MDM Functions* will be:

- Master Data Duplicate Detection and De-Duplication, and
- Automated Master Data Harmonising, Enrichment and Cleansing.

The corresponding AI tools and methods identified from the literature that can form part of a solution for the organisation's master data problem can then be identified in the *AI Tools and Methods Identified from Literature* column of the framework. According to the framework, the following solutions can be applied in this scenario:

- Master Data Duplicate Detection and De-Duplication:
“*Categorisation of the Master Data utilising Machine Learning Techniques such as Named Entity Recognition (NER) and Natural Language Understanding (NLU) is suggested by Behera and*

Panda (2023). These methods assist in discovering similar fields in master data models and can be used to detect and correct duplicate data in the MDM system.

Machine Learning techniques such as a data matching and de-duplication is suggested by Liu and Anderson (2024). Pattern and anomalies in master data can be recognised by employing supervised learning programs such as support vector machines and random forests, which enables better data matching and de-duplication.”

- Automated Master Data Harmonising, Enrichment and Cleansing:

“As an AI tool for data cleansing, Pontello et al. (2021) suggest utilising a supervised learning method such as a regression model to assist in improved quality management of master data.”

Case Study 2: Application of the AI-Enabled MDM Framework in the Scenario of Automating Master Data Maintenance Processes

Scenario:

The following describes how the *AI-Enabled MDM Framework* can be applied when an organisation wants to automate the master data maintenance residing in their ERP systems.

Utilising the AI-Enabled MDM Framework:

In this scenario, a data management industry professional can use the *AI-Enabled MDM Framework* to determine which AI capabilities may match the organisation's problem.

The *MDM Function Grouping* will be *Master Data Lifecycle Management*. The corresponding *MDM Function* will be *Master Data Creation and Maintenance*, and the *AI Capabilities Relevant to the MDM Functions* will be:

- Automated Master Data Maintenance.

The corresponding AI tools and methods identified from the literature that can assist the organisation in finding an AI solution for the master data problem can then be identified in the *AI Tools and Methods Identified from Literature* column of the framework. According to the framework, the following solutions can be applied in this scenario:

- Automated Master Data Maintenance:

“Master data maintenance tasks such as data quality assessment can be automated by employing AI tools such as deep-learning methods (Rangineni et al., 2023).

Natural Language Processing (NLP) techniques can be employed to master data searching functionality to better understand user queries and retrieve correct master data records (Jonathan & Frank, 2023).”

Case Study 3: Application of the AI-Enabled MDM Framework for Master Data Analysis

Scenario:

An organisation would like to discover hidden patterns in their data as part of their master data analysis capabilities in the MDM system.

Utilising the AI-Enabled MDM Framework:

In this scenario, a data management industry professional can use the *AI-Enabled MDM Framework* to determine which AI capabilities may match the organisation's problem.

The *MDM Function Grouping* will be *Master Data Cross Functions*. The corresponding *MDM Function* will be *Master Data Analysis and Reporting*, and the *AI Capabilities Relevant to the MDM Functions* will be:

- Master Data Analysis and Monitoring.

The corresponding AI tools and methods identified from the literature that can assist the organisation in finding an AI solution for the master data problem can then be identified in the *AI Tools and Methods Identified from Literature* column of the framework. According to the framework, the following solutions can be applied in this scenario:

- Master Data Analysis and Monitoring:
“Employing the AI method K-means clustering, which is used for pattern recognition and clustering, to identify correlations in master data, as per the paper by Pontello et al. (2021).”

Case Study 4: Application of the AI-Enabled MDM Framework for Detecting Patterns in Master Data

Scenario:

An organisation would like to discover hidden patterns in their data as part of their master data enrichment and cleansing capabilities in the MDM system.

Utilising the AI-Enabled MDM Framework:

In this scenario, a data management industry professional can apply the *AI-Enabled MDM Framework* to determine which AI capabilities may match the organisation's need for detecting patterns in the master data.

The *MDM Function Grouping* will be *Master Data Quality Assurance*. The corresponding *MDM Function* will be *Master Data Enrichment and Cleansing*, and the *AI Capabilities Relevant to the MDM Functions* will be:

- Master Data Pattern and Anomaly Detection.

The corresponding AI tools and methods identified from the literature that can assist the organisation in finding an AI solution for the master data problem can then be identified in the *AI Tools and Methods*

Identified from Literature column of the framework. According to the framework, the following solutions can be applied in this scenario:

- Master Data Pattern and Anomaly Detection:

“Machine Learning tools can be applied to identify patterns in master data residing in large databases and in estimating new data attributes (Riesener et al., 2022).

Employing supervised learning programs such as support vector machines and random forests for data matching is suggested by Liu and Anderson (2024), which can assist in data pattern and anomaly detection.

AI techniques such as Machine Learning and Deep Learning algorithms can be applied to master data for enhanced data anomaly detection and pattern prediction, according to Jonathan and Frank (2023). Deep Learning Models such as Neural Networks (NN) and Recurrent Neural Networks (RNN) show strong potential for assisting in real-time master data anomaly detection (Jonathan & Frank, 2023).”

The scenarios laid out in the above examples are not an exhaustive list of application scenarios but give the reader a general sense of how the *AI-Enabled MDM Framework* can be employed as a helpful tool to assist industry professionals in the field of Master Data Management in exploring how to apply AI to the MDM functions identified in this study.

These scenarios also motivate the value of the research findings as a usable design science research artefact. This artefact does not encompass all the possibilities for each AI Capability applicable to the identified MDM functions, and organisations would need to allocate internal researchers to further investigate the application of AI tools and methods based on the initial direction provided for each AI capability applicable to MDM functions identified in this study.

7.3.2 THE AI-ENABLED MDM FRAMEWORK AS A RESEARCH TOOL FOR FUTURE RESEARCH IN THE FIELD OF AI AND MDM

Future research must further investigate the AI tools and methods identified from the literature to create artefacts such as implementation methodologies, reference models, and frameworks that can be useful to academic research and also to industry professionals for implementing each of the suggested tools and methods for the corresponding challenges in managing master data in an MDM system. Specific AI, ML and NLP tools should be analysed to determine their usefulness in addressing the master data quality problems in an MDM system.

As mentioned, the research artefact does not include all the possibilities for each AI Capability applicable to the identified MDM functions. Future research can explore further AI tools and methods that can be employed for each identified AI capability. More specific AI implementation scenarios should also be identified and researched to assist complex design problems in integrating AI tools into Master Data Management systems.

7.4 CONCLUSION

This chapter produced the final research artefact using the survey responses and combining the findings to create the *AI-Enabled Master Data Management Framework*. The most prominent MDM functions and AI capabilities identified for MDM and DQM were combined to produce the final research artefact, namely the *AI-Enabled MDM Framework* that categorises MDM functions and tasks to appropriate AI data management capabilities and provides suggested AI tools and methods identified from the literature.

The most prominent Master Data Management (MDM) functions were defined by classifying the MDM functions identified according to the highest percentages in the survey responses and the highest number of references in the literature. The most prominent AI Capabilities for MDM and DQM were also defined by classifying the AI Data Management Capabilities Identified according to the highest percentages in the survey responses and the highest number of references in the literature.

The findings of this research are useful, as demonstrated by the case study examples. They are a valuable portion of research on which future researchers can base their work in AI and MDM. This section also promotes the value of the research findings as a usable design science research artefact by evaluating specific case study scenarios and demonstrating how the research artefact could be applied in each.

8 CONCLUSION

The figure below shows the layout and position of Chapter 8 in relation to the rest of the research paper.

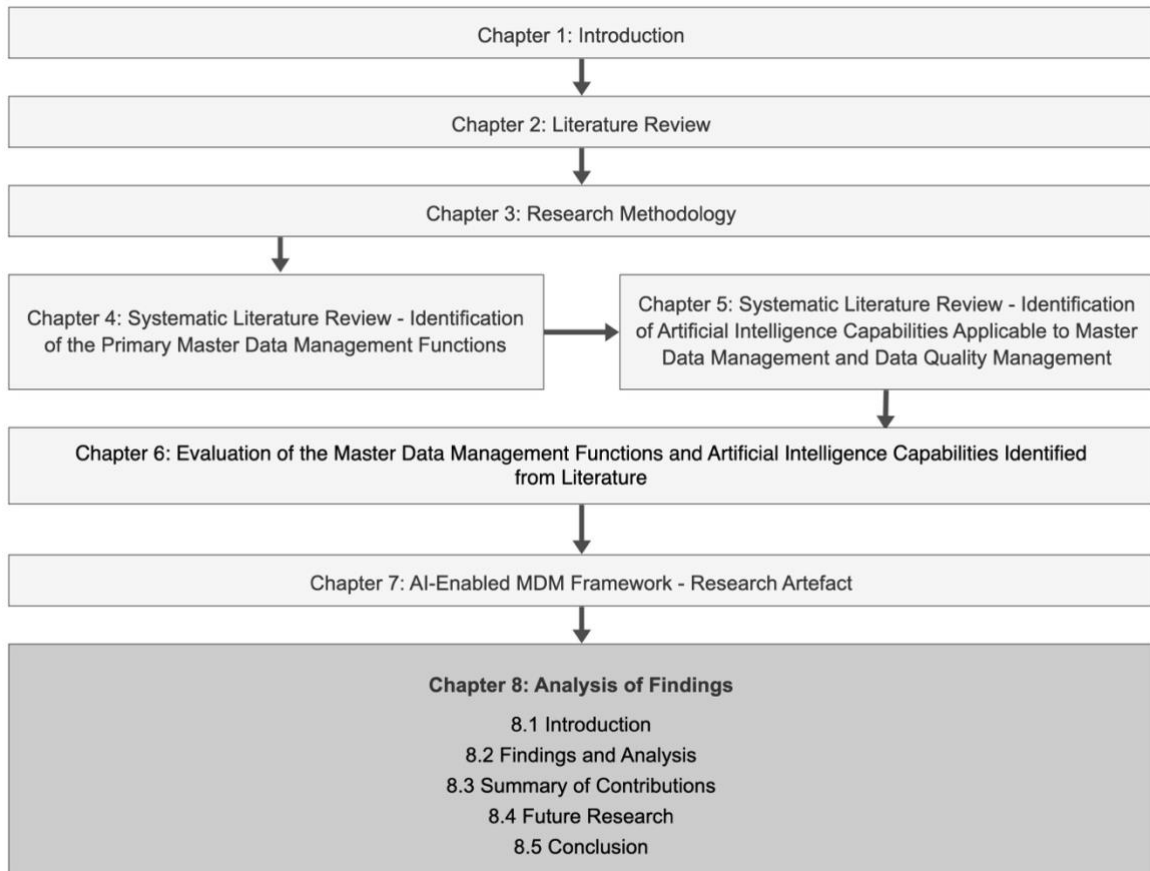


Figure 36: Structure of Research Paper - Position of Chapter 8

8.1 INTRODUCTION

In this section, the research findings are analysed, the summary of contributions is laid out, and future research opportunities are discussed. The primary findings of this research, namely the *AI-Enabled MDM Framework* artefact that categorises Master Data Management (MDM) functions to appropriate AI data management capabilities and applicable AI tools and methods identified from the literature, are analysed and discussed. The research findings are also examined with reference to the diffusion of innovation (DOI) theory.

The primary objective of the research project was to investigate how Artificial Intelligence capabilities can enhance Master Data Management functions to improve the quality of master data within an Enterprise Resource Planning system. The study provides AI researchers and data management professionals with a helpful research artefact that can be used as a guide for adopting AI tools into MDM systems. The study used two Systematic Literature Reviews to extract the data utilised in designing the primary research

artefact. This research's primary contribution to design science knowledge is the *AI-Enabled MDM framework* artefact that categorises MDM functions according to appropriate AI data management capabilities with the corresponding AI tools and methods identified from the literature. The artefact can be viewed in section 7.3.

This research also helps data management professionals perceive the usefulness and value that can be added to an organisation's ERP systems by integrating AI tools with their MDM systems for higher-quality master data. The research shows the importance of investigating AI's impact on an organisation's master data functionality.

8.2 FINDINGS AND ANALYSIS

The findings of this research study include an evaluated and classified framework that maps AI capabilities to MDM functions, as well as subsequent AI tools and methods applicable to the MDM functions. As seen in section 7.3, MDM functionalities that will benefit from AI tools' capabilities are identified and presented in the final research artefact.

The study aimed to discover MDM functionalities that will benefit from applying AI tools for organisations maintaining master data in ERP systems. Enhancing the MDM functionalities of a system will result in better overall master data quality in an organisation's systems and will improve overall business processes and efficiency, as per Riesener et al. (2022).

Findings and Analysis – MDM Functions

An extensive list of Master Data Management functions were identified in Chapter 4. The research survey questions were designed to determine which of the MDM functions identified has the highest impact on master data quality, which are the most time-consuming and which require the most human involvement. According to the diffusion of innovation (DOI) theory, which is discussed in the following sections, the adoption of new technologies is negatively influenced by factors such as complexity and relative advantage. Therefore, it was necessary to determine the general views on the difficulties of managing master data by investigating the MDM functionality from a data management professional's viewpoint.

The MDM functions that were identified as the most problematic were *Master Data Creation and Maintenance*, *Master Data Enrichment and Cleansing*, and *Master Data Validation and Business Rules*. These functions significantly impact data quality, require the most human involvement, and are the most time-consuming. Therefore, the rate at which AI capabilities may be applied to these MDM functions, in particular, would likely be prolonged.

Findings and Analysis – AI Capabilities, Tools and Methods

This study has extracted a comprehensive set of AI capabilities applicable to MDM functions by employing a systematic literature review method in Chapter 5. Corresponding tools and methods appropriate to each of the extracted AI capabilities relevant to the MDM functions, as mentioned in the literature, have also

been included in the study's findings to assist in creating the primary research artefact laid out in section 7.2. This research artefact, the *AI-Enabled MDM Framework*, is further described in the following sections.

The AI capabilities that hold the most significant promise for improved master data quality based on SLR findings and the opinions of industry professionals based on the research survey findings are *Master Data Error Detection and Correction*, *Identification of Inconsistent and Missing Master Data Fields*, and *Master Data Duplicate Detection and De-Duplication*. These AI capabilities were highlighted as the easiest to implement and most important for organisations to adopt. The perceived difficulty of implementing AI capabilities in MDM functions may negatively influence adoption. Thus, AI capabilities such as *Master Data Cataloguing*, *Master Data Analysis and Monitoring*, *Mass Data Maintenance*, and *Automated Master Data Harmonising, Enrichment, and Cleansing* may not be adopted as quickly as the aforementioned AI capabilities that are easier to implement.

Findings and Analysis – MDM Functions and AI Capabilities with Reference to the Diffusion of Innovation Theory

According to the theoretical lens of the study, namely the DOI theory, the adoption of new technologies is influenced by relative advantage, compatibility, trialability, visibility and complexity (Rogers, 2010). The idea of enhancing MDM functionality with the capabilities of AI tools and other AI capabilities, such as Natural Language Processing (NLP) and Machine Learning (ML), is positively influenced by the perceived advantage of implementing AI tools into MDM systems. The compatibility of AI tools in MDM systems, the ability to conduct trials before implementing these AI tools, the visibility of the AI models used by the systems, and the complexity of implementing AI tools in MDM systems play a role in technological adoption.

Section 2.6 laid out the DOI theoretical framework, which serves as the theoretical lens through which this research paper's findings are examined. The DOI theory serves as a benchmark for technological adoption research. The DOI theory was referenced in the design of the research survey questions to determine the perceived complexity and relative advantage of implementing AI tools in MDM Functions. The survey responses of the data management professionals revealed which MDM functions they deem to be most time-consuming, which need the most manual involvement, and which have the most significant impact on master data quality. It also sheds light on the perceived complexity of integrating AI capabilities for specific MDM functions by asking which AI capabilities extracted from literature could be most easily implemented into an MDM system.

The innovation referred to in this research study is the idea of enhancing MDM functionality with AI tools and other AI capabilities such as Natural Language Processing (NLP) and Machine Learning (ML). Thus, the adoption of AI capabilities in Master Data Management systems is perceived as a complex task that requires much research.

Findings and Analysis – AI-Enabled MDM Framework

The primary research contribution is the *AI-Enabled MDM Framework*, which can be viewed in section 7.3. The framework is positioned as a DSR artefact and can be employed as a useful tool to assist industry professionals in Master Data Management in integrating AI tools and methods into the MDM functions identified in this study.

The *AI-Enabled MDM Framework* lays out the MDM functional groupings and the AI capabilities extracted from literature relevant to the MDM functions. The AI tools and methods identified from the literature are provided to assist industry professionals in exploring and finding adequate AI solutions for the issues they face with specific MDM functions in the organisation.

Three case studies were provided to guide how the *AI-Enabled MDM Framework* can be applied to specific business Master Data Management issue scenarios. They motivate the value of the research findings as a usable design science research artefact.

The findings of this study only encompass some of the possibilities for each AI Capability applicable to the identified MDM functions. As mentioned before, organisations must allocate internal researchers to investigate the application of AI tools and methods further based on the initial direction provided for each AI capability applicable to MDM functions identified in this study.

8.3 SUMMARY OF CONTRIBUTIONS

The primary research artefact contributed was the *AI-Enabled MDM Framework* artefact that categorises MDM functions to appropriate AI data management capabilities, with corresponding AI tools and methods identified from the literature. The artefact can be viewed in section 7.3. The study also contains secondary research contributions from the two respective Systematic Literature Reviews in Chapter 4 and Chapter 5.

Chapter 4's first Systematic Literature Review provides the MDM functions and literature trends. It also provides the primary industries utilising MDM systems as extracted from the literature and a table of the master data objects mentioned in each paper.

In the second Systematic Literature Review in Chapter 5, the AI capabilities applicable to MDM and DQM have been extracted from the literature. The initial findings of AI capabilities are classified and later used in Chapter 7 to establish the final research artefact.

8.4 FUTURE RESEARCH

The primary aim of this study was to discover how AI capabilities can be utilised in MDM systems to improve the quality of master data within an ERP system. The findings in this study provide researchers with a theoretical reference framework for further research in AI and MDM. AI Capabilities applicable to each MDM function were categorised into *Master Data Lifecycle Management*, *Master Data Quality*

Assurance, Master Data Cross Functions, and Master Data Administration. The final research artefact also presented the AI capabilities relevant to each of the extracted MDM functions of the research and the corresponding AI tools and methods. Researchers should, therefore, investigate how to implement each identified AI capability and corresponding tools and methods for the MDM functions.

Additionally, future research must investigate the AI tools and methods identified in the literature to create helpful research artefacts such as implementation methodologies, reference models, and frameworks that can be applied to academic research and also employed by industry professionals to implement each of the suggested tools and methods for the corresponding challenges in managing master data in an MDM system.

Additionally, specific AI, ML and NLP tools could be researched to determine their usefulness in addressing master data quality issues in an MDM system.

8.5 CONCLUSION

This section outlined the analysis of the research findings and contributions. The research findings were discussed and analysed with reference to the DOI theory. Employing the theoretical lens of DOI, the adoption of AI capabilities into MDM functions is relative to the perceived complexity, the relative advantage of utilising AI capabilities, the compatibility of AI tools in MDM systems, and the perceived visibility of the application of AI tools to MDM functions.

This research has investigated the application of AI capabilities to MDM systems for enhanced master data quality in Enterprise Resource Planning (ERP) systems. Leveraging AI tools in MDM systems hold promising benefits for data quality improvement. AI capabilities such as *Master Data Duplicate Detection and De-Duplication* and *Master Data Error Detection and Correction* can have a massive impact on increased data quality in organisational master data and, subsequently, an increase in business processes and overall success (Riesener et al., 2022).

Due to the significant impact that master data quality has on business success, more research is still necessary on this topic. AI tools should be studied alongside the identified MDM functions and corresponding AI capabilities in the study to effectively address the pressing data-related challenges and business process issues for organisations with master data residing in ERP systems.

There have been increased opportunities for automating certain data processing functionality in MDM systems. Therefore, additional research must build on this study to investigate the enhancement of MDM functions by designing implementation methodologies, reference models, and frameworks that can be applied to the topic of AI capabilities and tools applicable to MDM functions.

9 REFERENCES

- Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. *Applied Sciences*, 13(12), 7082.
- Behera, T., & Panda, B. (2023). Master Data Management using Machine Learning Techniques: MDM Bot. *Authorea Preprints*.
- Boland, A., Dickson, R., & Cherry, G. (2017). Doing a systematic review: A student's guide.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Cacciattolo, M. (2015). Ethical Considerations in Research. In (pp. 61-79).
https://doi.org/10.1007/978-94-6300-112-0_4
- Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 10(6), 807-815.
<https://doi.org/https://doi.org/10.1016/j.cptl.2018.03.019>
- Chmielarz, G., & Pabian, A. (2020). AI development and application of AI-based solutions in the area of information and data management in Poland. 16th International Conference on Network and Service Management (CNSM),
- Dittrich, Y., Vaucouleur, S., & Giff, S. (2009). ERP Customization as Software Engineering: Knowledge Sharing and Cooperation. *IEEE Software*, 26(6), 41-47.
<https://doi.org/10.1109/MS.2009.173>
- Dutta, A., Deb, T., & Pathak, S. (2021). Automated Data Harmonization (ADH) using Artificial Intelligence (AI) [Article]. *OPSEARCH*, 58(2), 257-275.
<https://doi.org/10.1007/s12597-020-00467-4>
- Ender Hancioglu, S. (2020). A logic to handle, define and process item characteristics in ERP. *CEUR Workshop Proceedings*,
- Fernando, L. K. W., & Haddela, P. S. (2017). Hybrid framework for master data management.
- Galanis, N.-I., Vafiadis, P., Mirzaev, K.-G., & Papakostas, G. A. (2021). Machine Learning Meets Natural Language Processing - The Story so Far. In (pp. 673-686). Springer International Publishing. https://doi.org/10.1007/978-3-030-79150-6_53
- Gantayat, N., Saha, D., Sen, J., & Mani, S. (2019). Goal-based Ontology Creation for Natural Language Querying in SAP-ERP Platform.
- Goldkuhl, G. (2012). Design research in search for a paradigm: Pragmatism is the answer. *Practical Aspects of Design Science: European Design Science Symposium, EDSS 2011, Leixlip, Ireland, October 14, 2011, Revised Selected Papers 2*,
- Grangel-González, I., Rickart, M., Rudolph, O., & Shah, F. (2023). LIS: A Knowledge Graph-Based Line Information System. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*,
- Gröger, C. (2021). There is no AI without data. *Commun. ACM*, 64(11), 98–108.
<https://doi.org/10.1145/3448247>

- Guo, A., Liu, X., & Sun, T. (2018). *Research on Key Problems of Data Quality in Large Industrial Data Environment* Proceedings of the 3rd International Conference on Robotics, Control and Automation, Chengdu, China.
<https://doi.org/10.1145/3265639.3265680>
- Haneem, F., Ali, R., Kama, N., & Basri, S. (2017). Resolving data duplication, inaccuracy and inconsistency issues using Master Data Management.
- Haneem, F., Azmi, A., & Kama, N. (2017). Co-dependence relationship between master data management and data quality: A review [Review]. *Journal of Theoretical and Applied Information Technology*, 95(22), 6323-6335.
<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85036549745&partnerID=40&md5=1cd11efe16281e3b369e27cd8b764db3>
- Hannila, H., Tolonen, A., Harkonen, J., & Haapasalo, H. (2019). Product and supply chain related data, processes and information systems for product portfolio management [Article]. *International Journal of Product Lifecycle Management*, 12(1), 1-19.
<https://doi.org/10.1504/IJPLM.2019.104352>
- Harrison, T., Luna-Reyes, L. F., Pardo, T., Paula, N. D., Najafabadi, M., & Palmer, J. (2019). *The Data Firehose and AI in Government: Why Data Management is a Key to Value and Ethics* Proceedings of the 20th Annual International Conference on Digital Government Research, Dubai, United Arab Emirates.
<https://doi.org/10.1145/3325112.3325245>
- Hechler, E., Oberhofer, M., & Schaeck, T. (2020). Applying AI to Master Data Management. In E. Hechler, M. Oberhofer, & T. Schaeck (Eds.), *Deploying AI in the Enterprise: IT Approaches for Design, DevOps, Governance, Change Management, Blockchain, and Quantum Computing* (pp. 213-234). Apress.
https://doi.org/10.1007/978-1-4842-6206-1_9
- Herrmann, J.-P., Tackenberg, S., Padoano, E., Hartlief, J., Rautenstengel, J., Loeser, C., & Böhme, J. (2022). An ERP Data Quality Assessment Framework for the Implementation of an APS system using Bayesian Networks. *Procedia Computer Science*, 200, 194-204. <https://doi.org/https://doi.org/10.1016/j.procs.2022.01.218>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75-105. <https://doi.org/10.2307/25148625>
- Hikmawati, S., Santosa, P. I., & Hidayah, I. (2021). Improving Data Quality and Data Governance Using Master Data Management: A Review. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 5(3), 90.
<https://doi.org/10.22146/ijitee.66307>
- Ibrahim, A., Mohamed, I., & Satar, N. S. M. (2021). Factors Influencing Master Data Quality: A Systematic Review [Article]. *International Journal of Advanced Computer Science and Applications*, 12(2), 181-192.
<https://doi.org/10.14569/IJACSA.2021.0120224>
- Jaksic, D., Jovanovic, V., & Posic, P. (2017). Integrating evolving MDM and EDW systems by data vault based system catalog.
- Jeong, J. H., Woo, J. H., & Park, J. (2020). Machine Learning Methodology for Management of Shipbuilding Master Data. *International Journal of Naval Architecture and Ocean Engineering*, 12, 428-439.
<https://doi.org/https://doi.org/10.1016/j.ijnaoe.2020.03.005>

- Jesmeeen, M., Hossen, J., Sayeed, S., Ho, C., Tawsif, K., Rahman, A., & Arif, E. (2018). A survey on cleaning dirty data using machine learning paradigm for big data analytics. *Indones. J. Electr. Eng. Comput. Sci*, 10(3), 1234-1243.
- Jonathan, H., & Frank, E. (2023). Real-Time Data Monitoring and Anomaly Detection with AI: A Comprehensive Overview.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and Software Technology*, 51(1), 7-15.
- Knapp, M., & Hasibether, F. (2011). Material master data quality. 2011 17th International Conference on Concurrent Enterprising, ICE 2011 - Conference Proceedings,
- Knapp, M., & Lutz, T. (2021). Master Data Taxonomy - A systematic approach to assess and migrate master data.
- Knolmayer, G. F., & Röthlin, M. (2006). Quality of material master data and its effect on the usefulness of distributed ERP systems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),
- Krieger, R., & Schorr, C. (2019). A reference model for product data profiling in retail ERP systems. DATA 2019 - Proceedings of the 8th International Conference on Data Science, Technology and Applications,
- Kulkov, I. (2021). The role of artificial intelligence in business transformation: A case of pharmaceutical companies. *Technology in Society*, 66, 101629. <https://doi.org/https://doi.org/10.1016/j.techsoc.2021.101629>
- Leary, D. E. O. (2014). Embedding AI and Crowdsourcing in the Big Data Lake. *IEEE Intelligent Systems*, 29(5), 70-73. <https://doi.org/10.1109/MIS.2014.82>
- Li, F., & Xu, G. (2022). AI-driven customer relationship management for sustainable enterprise performance. *Sustainable Energy Technologies and Assessments*, 52, 102103. <https://doi.org/https://doi.org/10.1016/j.seta.2022.102103>
- Liu, C., & Anderson, J. (2024). A Hybrid Framework for Master Data Management: Integrating Machine Learning and Traditional Approaches.
- Maedche, A. (2010). An ERP-centric master data management approach. 16th Americas Conference on Information Systems 2010, AMCIS 2010,
- Malihi, E., Movahedi-Sobhani, F., & Aghdasi, M. (2016). Customisation of master data elements in ERP systems aligned with business goals [Article]. *International Journal of Information Systems and Change Management*, 8(1), 37-59. <https://doi.org/10.1504/IJISCM.2016.077950>
- Miliano, A., Steven, I., Kosim, K. P., Jayadi, R., & Mauritsius, T. (2020). Machine Learning-based Automated Problem Categorization in a Helpdesk Ticketing Application.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*, t. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269.
- Nightingale, A. (2009). A guide to systematic literature reviews. *Surgery (Oxford)*, 27(9), 381-384. <https://doi.org/https://doi.org/10.1016/j.mpsur.2009.07.005>

- Nulhusna, R., Taufiq, N. F., & Ruldeviyani, Y. (2022). Strategy to Improve Data Quality Management: A Case Study of Master Data at Government Organization in Indonesia.
- O'Farrell, J., Purchase, S., & Letch, N. (2013). Ameliorating ERP workflow using a sociomaterial lens. 19th Americas Conference on Information Systems, AMCIS 2013 - Hyperconnected World: Anything, Anywhere, Anytime,
- Otto, B., Hüner, K. M., & Österle, H. (2012). Toward a functional reference model for master data quality management. *Information Systems and e-Business Management*, 10(3), 395-425. <https://doi.org/10.1007/s10257-011-0178-0>
- Pakkala, D., Kääriäinen, J., & Mätäsniemi, T. (2024). Improving efficiency and quality of operational industrial production assets information management in customer – vendor interaction. *Journal of Industrial Information Integration*, 100644. <https://doi.org/https://doi.org/10.1016/j.jii.2024.100644>
- Pontello, V., Beckmann, H., & Lanquillon, C. (2021, 21-23 June 2021). Meta-learning approach for implementation of AI methods in the context of CRISP-DM with case studies from master data management. 2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC),
- Prokhorov, I., & Kolesnik, N. (2018). Development of a master data consolidation system model (on the example of the banking sector). *Procedia Computer Science*, 145, 412-417. <https://doi.org/https://doi.org/10.1016/j.procs.2018.11.093>
- Rangineni, S., Bhanushali, A., Suryadevara, M., Venkata, S., & Peddireddy, K. (2023). A Review on enhancing data quality for optimal data analytics performance. *International Journal of Computer Sciences and Engineering*, 11(10), 51-58.
- Riesener, M., Kuhn, M., Lender, B., & Schuh, G. (2022). Methodology for Automated Master Data Management using Artificial Intelligence. 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM),
- Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.
- Roth, H., Mönch, S. P., & Schäffer, T. (2022). Towards Augmented MDM: Overview of Design and Function Areas—A Literature Review.
- Schäffer, T., & Leyh, C. (2017). Master data quality in the era of digitization - toward inter-organizational master data quality in value networks: A problem identification. *Lecture Notes in Business Information Processing*,
- Shehadeh, K., Arman, N., & Khamayseh, F. (2021). Semi-Automated Classification of Arabic User Requirements into Functional and Non-Functional Requirements using NLP Tools.
- Silvola, R., Jaaskelainen, O., Kropsu-Vehkaperä, H., & Haapasalo, H. (2011). Managing one master data - Challenges and preconditions [Article]. *Industrial Management and Data Systems*, 111(1), 146-162. <https://doi.org/10.1108/02635571111099776>
- Siren, A., Smolander, K., & Jokela, M. (2015). The requirements and needs of global data usage in product lifecycle management. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*,
- Spreitzenbarth, J. M., Bode, C., & Stuckenschmidt, H. (2024). Artificial intelligence and machine learning in purchasing and supply management: A mixed-methods review

of the state-of-the-art in literature and practice. *Journal of Purchasing and Supply Management*, 30(1), 100896.

<https://doi.org/https://doi.org/10.1016/j.pursup.2024.100896>

Szívós, L., & Orosz, I. (2014). The role of data authentication and security in the audit of financial statements [Article]. *Acta Polytechnica Hungarica*, 11(8), 161-176.

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84908539250&partnerID=40&md5=3a56cc7259b0e5a5708fca72c5d4922b>

Tambo, T. (2018). PLM or ERP, the chicken or the EGG: Towards an enterprise level master data management approach for improving innovation and supply chain collaboration. Towards Sustainable Technologies and Innovation - Proceedings of the 27th Annual Conference of the International Association for Management of Technology, IAMOT 2018,

Tong, Q., Ming, X., & Zhang, X. (2020). *The Realization for Automated Warehouse Based on the Integration of ERP and WMS* Proceedings of the 2020 the 7th International Conference on Automation and Logistics (ICAL), Beijing, China.

<https://doi.org/10.1145/3412953.3412954>

Tverdal, S., Goknil, A., Nguyen, P., Husom, E. J., Sen, S., Ruh, J., & Flamigni, F. (2024). *Edge-based Data Profiling and Repair as a Service for IoT* Proceedings of the 13th International Conference on the Internet of Things, <conf-loc>, <city>Nagoya</city>, <country>Japan</country>, </conf-loc>.

<https://doi.org/10.1145/3627050.3627065>

Vaishnavi, V., & Kuechler, W. (2004). Design research in information systems.

Vilminko-Heikkinen, R., & Pekkola, S. (2017). Master data management and its organizational implementation: An ethnographical study within the public sector [Article]. *Journal of Enterprise Information Management*, 30(3), 454-475.

<https://doi.org/10.1108/JEIM-07-2015-0070>

Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, 47, 76-87.

<https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2018.12.017>

Wang, L., Ming, X., & You, J. (2009). The Steps and Methodology of Identifying Master Data from Business Processes. WRI World Congress on Software Engineering,

Wannalai, N., & Mekruksavanich, S. (2019). The Application of Intelligent Database for Modern Information Management.

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All---A Contingency Approach to Data Governance. *J. Data and Information Quality*, 1(1), Article 4.

<https://doi.org/10.1145/1515693.1515696>

Xu, B., Liu, X., Zhao, W., Li, Z., Wang, C., & Zhu, Q. (2021). *Research on the Architecture and Key Technologies of the Ubiquitous Customer Service Operating System for State Grid* 2021 2nd International Conference on Artificial Intelligence and Information Systems, Chongqing, China. <https://doi.org/10.1145/3469213.3470417>

Zhang, M., Schulte, O., & Luo, Y. (2021). *Leveraging Approximate Constraints for Localized Data Error Detection* Proceedings of the Fourth International Workshop on Exploiting Artificial Intelligence Techniques for Data Management, <conf-loc>,

<city>Virtual Event</city>, <country>China</country>, </conf-loc>.

<https://doi.org/10.1145/3464509.3464888>

Zhang, Y., & Ye, Z. (2011, 16-18 Dec. 2011). Data Maintenance and Control Strategy of Group Management. 2011 Second International Conference on Innovations in Bio-inspired Computing and Applications,

Zong, W., Wu, F., Chu, L. K., & Sculli, D. (2017). Identification of approximately duplicate material records in ERP systems [Article]. *Enterprise Information Systems*, 11(3), 434-451. <https://doi.org/10.1080/17517575.2015.1065513>