CRANFIELD UNIVERSITY

Xxx Alexslis Nyuyfoghan Maindze

Developing a Data and Knowledge Management Approach for Integrated Vehicle Health Management

School of Aerospace, Transport and Manufacturing
IVHM Centre

PhD
Academic Year: 2017 - 2021

Supervisor: Professor Ian K. Jennions
Associate Supervisor: Dr Zakwan Skaf
October 2021

CRANFIELD UNIVERSITY

School of Aerospace, Transport and Manufacturing
Integrated Vehicle Health Management

PhD

Academic Year 2017 - 2021

Xxx Alexslis Nyuyfoghan Maindze

Developing a Data and Knowledge Management Approach for Integrated Vehicle Health Management

Supervisor: Professor Ian K. Jennions
Associate Supervisor: Dr Zakwan Skaf
October 2021

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

In Integrated Vehicle Management (IVHM), research and engineering activities are conducted that generate large amounts of data and content. These activities include simulations, observations, derivation, experiments and referencing. However, IVHM still faces a range of data- and Knowledge Management (KM) challenges ranging from data accuracy to long-term availability for prognostic and diagnostic health management. IHVM is data-centric and therefore requires a robust data life cycle management to supports its data- and Knowledge Management activities. An understanding of the concept of KM is fundamental to addressing the IVHM data and knowledge management issues.

In this regard, this thesis contextualises ‘Knowledge Management’ for IVHM by attempting to resolve the intellectual paradox that has characterised it over the years. It discusses the origins of Knowledge Management as a discipline and addresses its historical inconsistencies. This review of KM and its origins serves as a scoping study guiding a systematic review of data life cycle models. It reviews relevant standards and their role in the data life cycle.

Guided by the V-Model, a Data Life Cycle Model is developed as a result and validated using a multi-technique approach combining peer review and expert insights obtained through a purposive survey. The model is then applied to IVHM centre Knowledge Management System development (KMS). The outcome includes an improved requirements gathering process and a solid foundation for resolving IVHM data and Knowledge Management challenges.

Keywords: Data management, Conceptual Model Validation, Prognostic Health Management, Systems engineering, IVHM

ACKNOWLEDGEMENTS

I would like to extend my profound appreciation to the many people who have committed themselves to helping me achieve my academic goals. It is through their dedication and support that I was able to realize my potential. I am eternally grateful to you all and hope that I can make you all proud, especially Victoria Kernyuy, the mother who made me who I am.

I would first like to express my appreciation to my supervisors who have encouraged and supported me both academically and emotionally throughout this process. **Professor Ian Jennions** has been the most influential person in my academic career. He has been a dedicated mentor to me since my first year in the doctoral program. He has served as a voice of support when I needed encouragement, a voice of breadth and depth when I was thinking too linearly, and a voice of experience when I was limited by my own knowledge. I am forever indebted to him for seeing my ability to fit in at the grown-up table. If it were not for **Dr. Zakwan Skaf** this dissertation would not have been possible.

My gratitude to Professor Mark Easterby-Smith [RIP] for his support and from whom I learnt valuable lessons of humility, integrity, hard work and continuous intellectual curiosity.

I would also like to express my gratitude to Dr Kimbo Edwin and Dr Kih Priscilla for their continuous encouragement.

To my colleagues at the IVHM centre, you were simply fantastic. Special thanks to Sohaib Aslam for his invaluable support.

To those I have not mentioned here, you have not been forgotten. I am indebted to you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation/ Acronym	Meaning
AAAI	Association for the Advancement of Artificial Intelligence
AG	Advisory Generation
ATEC	Army Test and Evaluation Command
BS	British Standard
CBM	Component Business Model
CBM	Condition-based maintenance
CD	Compact Disc
CRUD	Create, Read, Update and Delete
CS	Completion rate
CSMD	Core Scientific Metadata Model
CSMD	Core Scientific Metadata
CV	Curriculum Vitae
DA	Data Acquisition
DCC	Digital Curation Centre
DCMES	Dublin Core Metadata Element Set
DCMI	Dublin Core Metadata Initiative
DDI	Data Documentation Initiative
DLCM	Data lifecycle model
DM	Data Manipulation
EDM	Environmental Data Management
EU	European Union
FGDC	Federal Geographic Data Committee
GE	General Electric
HA	Health Assessment
HDFS	Hadoop Distributed File System
ICPSR	International Leader in Data Stewardship
ICPSR	International Leader in Data Stewardship
IEC	International Electro-technical Commission
IEEE	Institute of Electrical and Electronics Engineers
INCOSE	International Council on Systems Engineering
iRODS	Integrated Rule-Oriented Data System
ISO	International Organization for Standardization
IT	Information Technology
IVHM	Integrated Vehicle Health Management
JISC	Joint Information Systems Committee
KM	Knowledge Management
KMS	Knowledge Management System
KPI	Key Performance Indicator
LAC	Library and Archives Canada

MDLM	master Data life cycle Management
MSc	MSc Science
MSU	Michigan State University
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OEM	Original equipment manufacturer
OSA-CBM	Open Systems Architecture for Condition Based Maintenance
PA	Prognostic Assessment
PHM	Prognostic Health Management
RDF	Resource Description Framework
RR	Response Rate
SD	State Detection
SDI	Software Defined Infrastructure
SODER	simulations, observations, derivation, experiments and referencing
STFC	The Science and Technology Facilities Council
UCSD	University of California San Diego
UK	United Kingdom
USGS	United States Geological Survey
XML	Extensible Mark-up Language

CHAPTER 1: INTRODUCTION

1.1 Problem Statement and Motivation

Knowledge Management over the years has become a central foundation for business improvement, sustainability, and competitiveness. This occurred against the backdrop of organisations seeking to manage their hard and soft assets better through greater integration with information technologies. A specific area in which the management of data, information and knowledge is fundamental for its success, is the field of Integrated Vehicle Health Management (IVHM). The first stage of the IVHM cycle is the collection of data about an asset. IVHM delivers value to stakeholders and reduces the cost of delivery by monitoring the health of an asset and making decisions based on the data collected (Jennions, 2011). Consequently, IVHM relies predominantly on the availability of high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health. The data has to be accurate, complete, timely, context-relevant, reliable and explicit (Dibsdale, 2011). The data-driven approach uses statistics and probability for analysing current and historical data (Mathew and Pecht, 2013). Model-based techniques facilitate the understanding of component failure and mode progression. The models can include physics-based models, Autoregressive Moving-Average (ARMA) techniques, Bayesian filtering algorithms, and empirical-based models. They use changes in essential assets or system properties over time, and use mathematics, computation and computer simulation to analyse the models. Hybrid computational analysis or fusion approach is an amalgamation data-driven and model-based techniques (Goebel, Vachtsevanos and Orchard, 2013).

The IVHM Centre at Cranfield University has existed since 2008 and runs an active research program that generates high-value knowledge through theoretical and experimental work. The Centre produces a wide range of algorithms and processes to capture and analyse data from experiments carried out on rigs and other facilities. The centre has worked on more than 40 projects, generating more than 120 technical papers, publishing six books, and generating lots of experimental data from its ten experimental rigs with the associated dynamic programming algorithms (Skaf and Jennions, 2017). With this growth in the creation of research data, data automating algorithms, technical papers, reports and theses, the centre needs both a data management model and a knowledge management system that facilitate the storing, organising and sharing of its research outputs and which is secure and scalable with a high level of cross-platform transferability.

However, the IVHM Centre does not yet have a systematic and coherent approach to IVHM knowledge and data management. The absence of a data life cycle model and knowledge management system means that valuable knowledge is lost or is very difficult to find. Data visualisation is fragmented and done on a project by project basis, which increases cost. There is insufficient algorithm documentation and communication for new researchers to rapidly get up to speed in these complex areas. A knowledge management system is therefore needed to improve the visibility of the Centre's range of research projects and the outputs from these projects. A system that presents information to enable browsing by a range of categories and searching on specific terms is desired.

The review of the literature highlighted, first of all, the need to understand the concept of knowledge management at a fundamental level and the relationship between data, knowledge management and data life cycle models. A detailed literature review that studies data life cycle models, frameworks, standards and process models pertaining to Knowledge Management in the context of IVHM has been suggested. A pilot Knowledge Management System was built on the existing University's SharePoint infrastructure to determine its suitability for the IVHM knowledge management system. A system that presents information to enable browsing by a range of categories and searching on specific terms is desired.

1.2 Overall Aim and Individual Objectives

The aim of this research is to develop a Knowledge Management System for IVHM using an appropriate data life cycle model to capture, not only reporting (datasets, processes, technical papers, etc.) but algorithms and experimental data.

The specific objectives of this research are:

1. To review the literature on data/ knowledge management and data life cycle models.
2. To develop and validate an appropriate Data Life Cycle Model for IVHM.
3. To use the Data Life Cycle Model as a foundation to build a Knowledge Management System (KMS) for IVHM.

1.3 Research Methodology

There are four phases to the methodological approach adopted for this research; Literature review, requirements gathering, conceptual modelling and systems development (Figure 1.1). The literature review adopted a combination of traditional and systematic reviews. The traditional review was used in the scoping study in order to create the protocol for planning searching and screening for a systematic review. The R0-MSc KMS Requirements document was the initial high level document specifying the overall system requirements to govern the development and implementation of the IVHM KMS. This was developed as part of an MSc project at the IVHM Centre. A review of the R0-MSc KMS requirements document led to the definition of the revised and improved functional specification (R1). This updated version of the requirements document is referred to as R1-MSc Requirements (Figure1.1). This update provided clarity of features and rationale of the KMS.

The methodological approach is depicted below (Figure 1.2). A systematic review is important because it is characterised by being objective, systematic, transparent and replicable. The requirements gathering supported through gap analysis characterised the second phase, which ran in parallel with the 1st, 2nd and 3rd phases. Model Integration and conceptual modelling techniques were used to develop the proposed Data Life Cycle model and the systems development model forming the implementation approach for the finalised Knowledge Management System requirements.

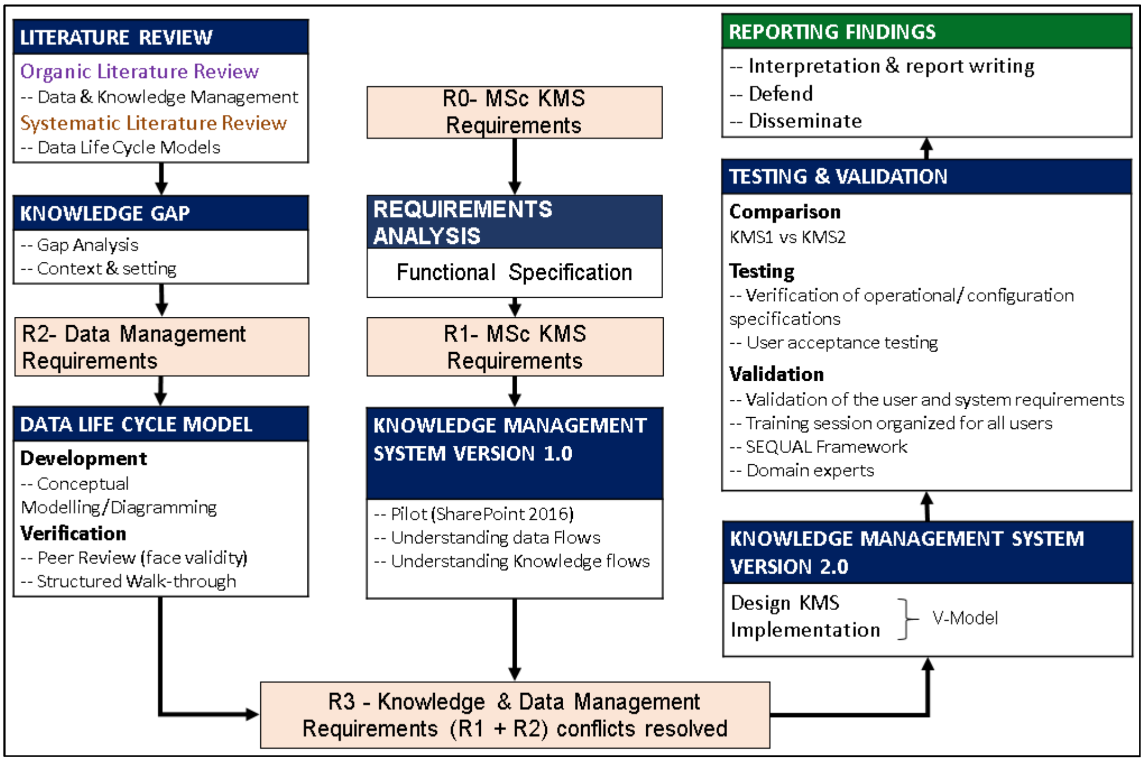


Figure 1-1 Research Methodology Flow Diagram

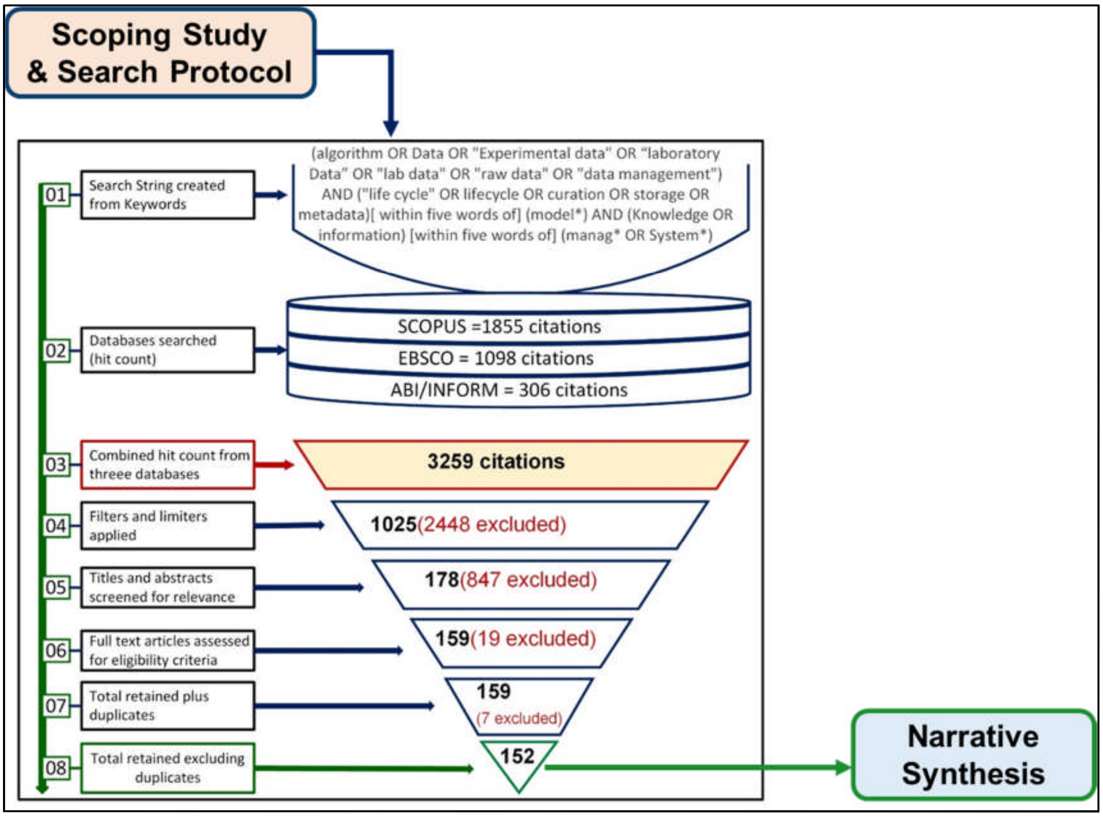


Figure 1-2 Systematic Literature

1.4 Quality Assurance

To ensure quality and reliability of the project, a project plan is developed, including the holding regular weekly meetings with supervisors. These regular meetings help in refining the research methodology, identifying problems and issues with procedures and operations and recommending measures towards addressing them. These supervisory meetings also ensure that the tools and techniques used have been validated as fit for purpose before use. In addition, peer review took place including presentation of findings at major academic events and publication of outcomes, and work is grounded through the access of peer-reviewed work indexed in recognised academic databases such as Scopus, ABI/Inform and ProQuest for literature.

1.5 Research Scope

It is important to stress that this study has been primarily concerned with the data life cycle model and how it can serve as a base for:

- Improved research, data- and knowledge and knowledge management activities.
- Improved requirements gathering and specification for systems development as per the requirements but does not delve into system verification and validation as the system is based on a pre-existing platform.

The thesis also delves into the approach for validating conceptual models as a process for validating the proposed model.

1.6 Organisation of Thesis

The organization of this thesis is in eight chapters using the paper-format. Furthermore, some of the published papers have been reformatted for conformity with the thesis template and to provide consistency. The papers that have been so formatted include Chapters 2, 5 and parts of Chapter 7. Part of Chapter 7 was published as an extended abstract. As a result, these are likely to appear as selfplagiarism on Turnitin. To the exception of Chapter 1, the thesis is organised as follows:

Chapter 2: Knowledge Management Yesterday And Tomorrow: Exploring an 'Intellectual Paradox'.

This chapter introduces the concept of Knowledge Management and its definitional challenges. It highlights the difficulty of a definitional consensus provides and proposes a solution to the predicament.

This chapter is a reformatted conference paper published in *“Advances in Manufacturing Technology XXXI -2017”*

Chapter 3: A History of Knowledge Management The 1960s To 2019: The Missing Bits.

This chapter discusses the history of Knowledge Management and the inconsistencies surrounding it. It sheds light on the philosophical pillars of knowledge management and how this was influenced by technological developments over the years. It shows how these developments impacted the narration of the history of Knowledge Management.

This chapter is the reformatted version of the paper under review at the *“International Journal of Knowledge Management Studies”.*

The above two chapters were a result of the scoping study. The scoping study is the first step towards the implementation of systematic reviews. It supports the creation of the review protocol that is detailed in chapter 4. These two chapters are addressing **‘Objective 1’** and supports **‘Objective 2’.**

Chapter 4: Scoping, Systematic Review, Design And Validation: A Methodological Approach.

This chapter defines the core methodology of the thesis. The chapter focuses on the systematic reviews of data life cycle models as the primary technique. It also includes part of the organic literature supporting the research work. This chapter addresses ‘**Objective 1’** and supports **‘Objectives 2 and 3**’.

Chapter 5: Towards an Enhanced Data- and Knowledge Management Capability: A Data Life Cycle Model Proposition for Integrated Vehicle Health Management.

Chapter 5 develops and presents the Data life cycle model for IVHM. It is the product of '**Objective 2'**. It details the phases of the proposed data life cycle and their descriptions. This chapter represents the core of the thesis. It details the IVHM data sources and challenges, the links between the data life cycle model, OSA-CBM architecture and key standards with IVHM data- and Knowledge Management difficulties.

This chapter is also a reformatted paper published at *"Annual Conference of the* Prognostic Health Management Society (PHM) *Society, 11."*

Chapter 6: The Importance of Data Life Cycle Model in IVHM Research and Engineering Practice: A Validation Survey.

This chapter emanates from **'Objective 2'**, the part that further supports the validation of the IVHM data life Cycle model. The chapter presents the results of the validation survey sent out purposively selected expert community of academics and engineers. The chapter highlights the significance of peer reviews in the validation of conceptual models.

Chapter 7: Developing a Knowledge Management System Using a Data Life Cycle Model: The IVHM Centre Case.

The chapter represents the use case or case study for application. After the application of the data life cycle models, the requirements were updated, and the Knowledge Management System built. This chapter showcase the main features of the system. The chapter is the response to **'Objective 3'** and the overall data management challenges of the Centre. It demonstrates the architecture before application and the architecture after application. This chapter also explains the relationship between the data layer, the knowledge management layer of the knowledge management pyramid and the Knowledge Management System.

Parts of this paper are reformats of the extended abstract published at the *'Annual Conference of the Prognostic Health Management Society (PHM), 2018'.*

Chapter 8: Discussion and Contribution.

This chapter discusses the research objectives, showing how they have been addressed. The chapter also highlights some of the major contributions of the thesis. It presents the theoretical and practical contributions and potential future research area themes.

1.7 Overview of Published and Submitted Work

1.7.1 Published papers

- Maindze, A., Skaf, Z., & Jennions, I. (2019). Towards an Enhanced Data- and Knowledge Management Capability: A Data Life Cycle Model Proposition for Integrated Vehicle Health Management: A Data Life Cycle Model Proposition for Integrated Vehicle Health Management. Annual Conference of the PHM Society, 11(1). https://doi.org/10.36001/phmconf.2019.v11i1.842
- Maindze, A. (2018). Developing a Knowledge Management System for Integrated Vehicle Health Management Using a Data Life Cycle Model. Annual Conference of the PHM Society, 10(1). https://doi.org/10.36001/phmconf.2018.v10i1.710
- Maindze A, Jennions I and Skaf Z. Knowledge Management Yesterday and Tomorrow: Exploring an 'Intellectual Paradox'. Advances in Transdisciplinary Engineering: Volume 6: Advances in Manufacturing Technology XXXI. Proceedings of the 15th International Conference on Manufacturing Research, Incorporating the 32nd National Conference on Manufacturing Research, September 5–7, 2017, University of Greenwich, UK.https://doi.org/10.3233/978-1-61499-792-4-522

1.7.2 Submissions under peer review

- Maindze, A., Skaf, Z., & Jennions, I.: A History of Knowledge Management: The Missing Bits from 1960s to 2019. Submitted to International Journal of Knowledge Management Studies

1.8 References

Dibsdale, C. (2011) ‘Integrated Vehicle Health Management Operations Rooms’, in Jennions, I. K. (ed.) Integrated Vehicle Health Management: Perspectives on an Emerging Field. SAE International, pp. 113–123.

Mathew, S. and Pecht, M. (2013) ‘IVHM System Design’, in Jennions, I. (ed.) Integrated Vehicle Health Management: The Technology. SAE International, pp. 157–176.

Goebel, K. ;, Vachtsevanos, G. and Orchard, M. E. . (2013) ‘Prognostics’, in Jennions, I. (ed.) Integrated Vehicle Health Management: The Technology. SAE International, pp. 49–70.

Jennions, I. K. (2011) ‘Introduction’, in Jennions, I. K. (ed.) Integrated vehicle health management: perspectives on an emerging field. SAE International, pp. 1–7.

Skaf, Z. and Jennions, I. (2017) Knowledge management system for the IVHM centre.

CHAPTER 2: KNOWLEDGE MANAGEMENT YESTERDAY AND TOMORROW: EXPLORING AN 'INTELLECTUAL PARADOX'.

(This chapter is a reformatted conference paper published in "Advances in Manufacturing Technology XXXI -2017")

Abstract. Knowledge Management continues to be characterized by strong contextual application with diversity of techniques, tools and applications which practitioners far and wide seem to agree and adopt. However, when it comes to its philosophical distinctness, it is yet to achieve something as seemingly easy as a common definition. There is significant agreement on fluidity and methods of application but limited consensus on philosophical interpretation. Furthermore, that we know what it is, acknowledge its impact, functional relevance and yet cannot articulate a common methodology points to what this paper terms an '*intellectual paradox'*.

An intellectual paradox is the phenomenon whereby professionals and academics acknowledge a concept, practice it, write about it, and promote its relevance individually but as a collective lack a consensus on exactly what it is. This paper seeks to explore this phenomenon in detail and to propose a philosophical framework. It further explores the role of the traditional composition; people, process and technology in sustaining this suggested conundrum. This phenomenon seems to tie neatly with the tacit form of knowledge on the basis of the difficulty in articulating a common definitional framework of perception, though it could be argued that it is merely exhibiting characteristics of 'Tacit' knowledge management; thereby justifying the status quo. Some authors point to "*descriptive frameworks*" and insufficient addressing of learning including structural differences in organisations. This difficulty per some writers, results from the use of multiple and variable methods, tools techniques and strategies. Their alternative proposition views for a both 'descriptive and prescriptive' framework still did not yield a consensus either. This paper seeks to explore the problem and to propose a new definition.

Keywords. Knowledge Management, Intellectual Paradox, definition, contextual application, philosophical distinctness, limited consensus, common methodology

2.1 Introduction

The significance of knowledge management (KM) as a highly treasured intangible asset in today's economy cannot be emphasized enough; not only has KM become a household name, it has also become the vital substructure for learning, growth, increased efficiency and effectiveness, and competitive advantage in all sectors of the economy (Thomas and Laurence, 1998; Jennex, Smolnik and Croasdell, 2012; Dalkir, 2013b; Omotayo, 2015; Maldonado-Guzman, Marin-Aguilar and Pinzon-Castro, 2017; Rhem, 2017; Serrat, 2017). Although the relevance of KM in today's business in propagating the concept of knowledge as a competitive resource remains firm in theory as well as in practice, it suffers from what we call an 'Intellectual Paradox'.

An intellectual paradox is the phenomenon whereby professionals and academics acknowledge a concept, practice it, write about it, and promote its relevance individually but as a collective lack a consensus on exactly what it is. Academics and practitioners broadly agree on the diversity of techniques, tools and application regardless of fluidity and context in their respective and individual functional spheres. However, when it comes to the question of collective or broad agreement on its definitional framework there is no consensus. It is this interaction between individualized consensual agreement on functional application and collective disagreement on definitional perception that creates the 'Intellectual paradox'. In other words, there is universal agreement on 'individual forms' and universal disagreement on 'collective structure' and is rooted in the KM history described in the next section.

2.2 A Short History of Knowledge Management and the Birth of a Paradox

The initial introduction of the expression knowledge management (KM) can be traced to the mid-70s. In May 1976 (Berry and Cook, 1976) distinguished between data, information and knowledge. An "*effective knowledge management* *for an enterprise*" (Berry and Cook, 1976) required maintaining repositories files… and other resources (Kellogg, 1983). However, it was not until 1982 that the phrase KM was again re-introduced by Charles Kellogg in (Kellog, 1982) and (Kellogg, 1983). In the latter, the phrase is mentioned seven times. KM gained more prominence with the introduction of the KM concept in 1986 by Dr. Karl Wiig at the United Nations (Jasimuddin, 2006; Dalkir and Liebowitz, 2011; Dalkir, 2013b). KM was further propelled into the academic and non-academic scene by Nonaka and Takeuchi [9] who studied the processes of organizational knowledge creation including dissemination using Japanese companies as case studies. They drew a distinction between explicit and tacit forms of knowledge and proposed a design which translates into people, process and technology; the trilogy that has dominated the expression of KM. In this this history to date, many perspectives on what KM is have emerged without any generating consensus.

2.3 Definitional Perspectives and Problem

The definitional problem was identified by Fahey and Prusak (1998) as number one in which they argued that it was *"… a critical error. Not developing a working definition of knowledge."* (Fahey and Prusak, 1998). The alternative suggestion by this paper is that though they elaborated on the processes of knowledge creation, conversion and the requisite conditions for knowledge creation and the management of knowledge, they did not provide what some authors have described as a 'prescriptive'(Rubenstein-Montano *et al.*, 2001) definition of KM thus creating the foundations for lack of consensus.

Furthermore, Nonaka and Takeuchi (Nonaka and Takeuchi, 1995) explained organizational knowledge creation as "…*The capability of a company as whole to create new knowledge, disseminate it throughout the organization, and embody it in products, services and systems."* (Nonaka and Takeuchi, 1995) . This process was built on the spiral interaction between tacit and explicit forms of knowledge at individual, group and organization level. It was the amalgam of the continuous interaction between tacit and explicit knowledge and the conversion processes that they called knowledge creation (see Figure 2.1).

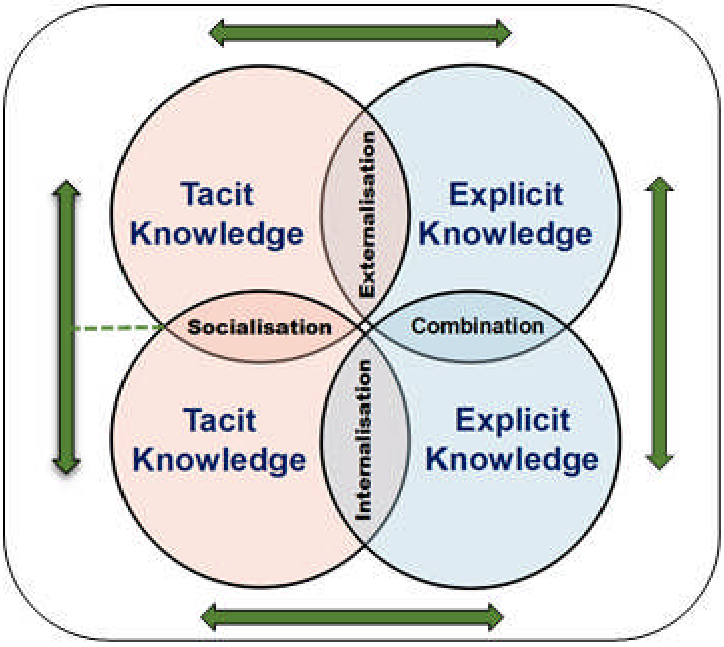


Figure 2-1 Organizational Knowledge Creation & Knowledge Conversion
[interpreting(Nonaka and Takeuchi, 1995)**]**

The next section describes early attempts at addressing the definitional problem.

2.4 Earlier Attempts at Resolving the Problem

Many researchers have considered and proposed solutions to the definitional problem. Collison and Parcel (2004) concluded that KM was very difficult to define. Jennex, Smolnik and Croadsdell (2012) used a consensus – building approach by surveying an Expert panel to propose a successful definition of KM. They concluded that a generally acceptable KM definition would be: *“KM success is a multidimensional concept. … KM success is measured using the dimensions of impact on business….”* (Jennex, Smolnik and Croasdell, 2012). By defining KM in terms of its success criteria, created yet another foundation for diverse perception of KM and sustenance of the status quo.

After informally studying more than 100 definitions of KM, Mishra wrote, “*KM is a multidisciplinary field of study that covers a lot of ground*” (Mishra, 2009). It is this concoction of approaches from diverse disciplines, researchers and professionals that sustains the lack of a definitional consensus to KM (Demarest, 1997; Maier and Remus, 2003; Mishra, 2009). This lack of consensus testifies to the *"Three Blind and an Elephant*" ailment that the KM field is enduring and continues to endure (Dalkir and Liebowitz, 2011; Serrat, 2017)(Kellog, 1982; Demarest, 1997; Natarajan and Shekhar, 2000).

In a more recent study, Girard & Girard (Girard and Girard, 2015) formally studied and analysed word composition of more than 100 openly accessible definitions of KM with applied orientation from 13 countries and 23 knowledge domains. This categorization according to these authors was down to the idea expressed across the literature that KM is diverse and draws from many disciples. The results show that KM is consistently defined in terms of Create, Share, manage, and knowledge process, Organization and information. They observed that two definitions could be carved out of these commonalities. They include:

- "*Knowledge Management is the process of creating, sharing, using and managing the knowledge and information of an organization.*

- *Knowledge Management is the management process of creating, sharing and using organizational information and knowledge.*" (Girard and Girard, 2015).

These definitions incorporate the basic elements of a good KM definition which some authors (Mishra, 2009; Dalkir and Liebowitz, 2011) have suggested should include combined capturing, storing and valuing of intellectual assets.

2.5 Proposition: Overcoming the Paradox

In this review, the authors found that the major lenses through which professionals and academics have viewed and understood KM throughout its history included multidisciplinary, science, processes, environment, technology, knowledge creation, value creation and retention. On the bases of the history, definitions or descriptions of KM to date, we concluded that knowledge is a multidisciplinary science (Alavi and Leidner, 2001; Becerra-Fernandez and Sabherwal, 2014; Cummings et al., 2013; Girard and Girard, 2015; Handzic, 2015; Nonaka, 1994; Nonaka et al., 2006; Nonaka and Takeuchi, 1995; Russel, 1961; Spender, 2015; Standard, 2005) and adapts to the organizational context (Maldonado-Guzman, Marin-Aguilar and Pinzon-Castro, 2017). This seems to have been a significant piece missing from the interpretations of the earlier works of KM to date. We, therefore, put forth a context agnostic proposition with high organizational interoperability that sums these characteristics of KM. The proposal follows thus: *Knowledge management is a* ***multidisciplinary science*** *and* ***process*** *of organizational knowledge* ***creation and retention*** *that engages* ***people****,* ***process****,* ***environment*** *and* ***technology*** *to create, retain or increase* ***value****.* Figure 2.2 below is used to depict this definition of KM and approaches proposed by this paper.

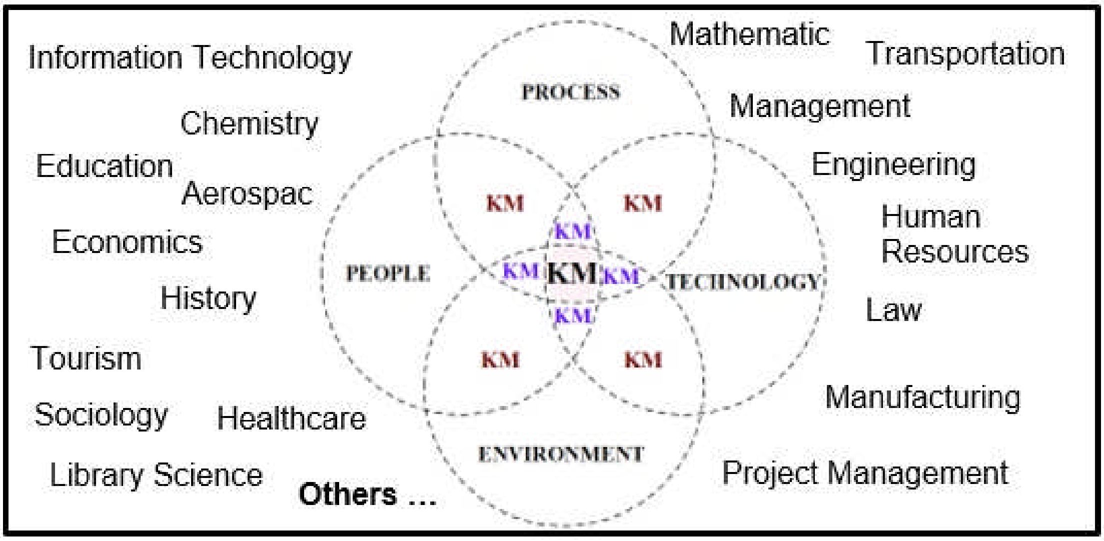


Figure 2-2 Illustration of KM; interpreting Nonaka & Takeuchi (Maindze, Jennions and Skaf, 2017)

Turning now to further evidence in the literature that supports the new definition. In the next two sections, we present the core lenses of KM perception comprising the proposed definition and how they are captured in existing works.

2.6 Knowledge Management as Multidisciplinary Science

Knowledge is a "*Universal*" and therefore constitutes a concept of general awareness and understanding (Nonaka and Takeuchi, 1995). By this explanation, knowledge belongs to all disciplines and therefore managing cannot be isolated from its respective diverse disciplinary affiliation and presence, thus making KM a multidisciplinary endeavour. Nonaka and Takeuchi (1995) confirmed the multi-disciplinarity of KM by highlighting the fact that "*socio-*

economists …[and other researchers] *in the fields of industrial organization, technology management, management strategy, and organizational theory have begun to theorize about management of Knowledge"* (Nonaka and Takeuchi, 1995). This affirmation demonstrates that these authors already construed KM as a multidisciplinary science without directly stating so.

Furthermore, KM is a concatenation of two or more disciplines and processes (Nonaka, 1994; Nonaka and Takeuchi, 1995; Standard, 2005; Nonaka, Von Krogh and Voelpel, 2006; Cummings *et al.*, 2013; Girard and Girard, 2015; Handzic, 2015; Spender, 2015). In their book, Dalkir et al. (2013a) observed that "*KM has its roots in a variety of Disciplines.*" They listed at least 12 fields that KM transcends. This multidisciplinary view is supported by other researchers who describe KM as a multifaceted, multi-sourced, ambiguous, scientific discipline with a multidisciplinary ownership, with a fragmented history and perception (Cummings *et al.*, 2013; Handzic, 2015; Spender, 2015). Girard and Girard (2015) categorized the definitions of KM by disciplines because *"... knowledge management is a multidisciplinary field drawing from many subject areas"* (Girard and Girard, 2015), as has been expressed in the literature.

2.7 KM – Engaging People, Process, Technology & Environment

This section shows how KM has been perceived and portrayed in the literature as fundamentally involving people, process, technology and the environment. It sheds light in how people, process and technology influenced the philosophical distinctness of KM.

In their research, Nonaka and Takeuchi (1995) presented the management of knowledge as the process of interaction between the epistemological and the ontological dimensions of knowledge. Though the environment is not captured in their design, it forms part of a key enabler of the organizational knowledge creation they referred to as– "Fluctuation and creative chaos" (Nonaka and Takeuchi, 1995). Nonaka and Takeuchi (1995) observed that "*An environmental fluctuation often triggers a breakdown within the organization, out of which new knowledge can be create*d". Their assessment and interpretation set the foundation what would become today's definitional challenge. Their publication captured all the elements or principles of KM espoused existing KM literature to date.

They drew attention to the process of knowledge conversion that is characterized by the interaction between explicit and tacit knowledge in a spiral process. KM as a process characterized by people, technology and environment is widely supported in the literature on (Alavi and Leidner, 2001; Cross, 2004; Cummings *et al.*, 2013; Girard and Girard, 2015; Handzic, 2015; Spender, 2015).

2.8 Conclusion

Altogether, the emphasis on the creation of new knowledge without including existing knowledge that is strongly considered in later approaches to knowledge management adds to the difficulty of defining knowledge management in a way that would be generally acceptable. Because existing knowledge is context or discipline specific, attempts at including this element has therefore led to extensive contextualization and proliferation of knowledge management definitions without a generally accepted one.

This paper has therefore proposed a definition that captures all the characteristics of knowledge management as expressed in various research papers and books from the 1970s to date.

2.9 References

Alavi, M. and Leidner, D. E. (2001) ‘Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues’, Miss. Q. Management Information Systems Research Center, University of Minnesota, 25(1), pp. 107–136. doi: 10.2307/3250961.

Becerra-Fernandez, I. and Sabherwal, R. (2014) Knowledge Management: Systems and Processes. Routledge.

Berry, J. F. and Cook, C. M. (1976) Managing Knowledge as a Corporate Resource. doi: ADA029891.

Collison, C. and Parcell, G. (2004) Learning to Fly: Practical knowledge Management from Leading and Learning Organisations. Capstone Ltd.

Cross, J. (2004) An Informal History of eLearning. On the Horizon, 2004. 12 (3). pp. 103.

Cummings, S. et al. (2013) ‘Proposing a fifth generation of knowledge management for development: investigating convergence between knowledge management for development and transdisciplinary research’, Knowledge Management for Development Journal. journal.km4dev.org, 9(2), pp. 10–36. Available at: http://journal.km4dev.org/index.php/km4dj/article/view/170.

Dalkir, K. (2013a) Knowledge Management in Theory and Practice. Routledge.

Dalkir, K. (2013b) ‘Organizational Memory in Institutions of Higher Education: A Case Study’, Knowledge Management: An International Journal, 12(1).

Dalkir, K. and Liebowitz, J. (2011) Knowledge Management in Theory and Practice. MIT Press.

Demarest, M. (1997) ‘Understanding knowledge management’, Long Range Plann. Elsevier, 30(3), pp. 321–384. doi: 10.1016/S0024-6301(97)00017-4.

Fahey, L. and Prusak, L. (1998) ‘The eleven deadliest sins of knowledge management’, California Management Review, University of California Press Journals, 40(3), pp. 265–276. Available at: http://cmr.ucpress.edu/content/ucpcmr/40/3/265.full.pdf.

Girard, John and Girard, JoAnn (2015) ‘Defining knowledge management: Toward an applied compendium’, Online Journal of Applied Knowledge Management, 3(1), p. 20.

Handzic, M. (2015) ‘A Descriptive Analysis of Knowledge Management Research: Period from 1997 to 2012’, in Bolisani, E. and Handzic, M. (eds) Advances in Knowledge Management. Springer International Publishing (Knowledge Management and Organizational Learning), pp. 45–64. doi: 10.1007/978-3-319-09501-1_3.

Jasimuddin, S. M. (2006) ‘Disciplinary roots of knowledge management: a theoretical review’, International Journal of Organizational Analysis. emeraldinsight.com, 14(2), pp. 171–180. doi: 10.1108/10553180610742782.

Jennex, M. E., Smolnik, S. and Croasdell, D. (2012) ‘Towards a Consensus Knowledge Management Success Definition’, in Organizational Learning and Knowledge: Concepts, Methodologies, Tools and Applications. IGI Global, pp. 403–415.

Kellog, C. (1982) ‘Knowledge Management: A Practical Amalgam of Knowledge and Data Base Technology’, in Proceeding AAAI’82 Proceedings of the Second AAAI Conference on Artificial Intelligence. Pittsburgh, PA, USA: AAAI Press, p. Pages 306-309.

Kellogg, C. (1983) ‘Intelligent Assistants for Knowledge and Information Resources Management’, IJCAI. Citeseer, pp. 170–172. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.109.6258&rep=rep1&type=pdf.

Maier, R. and Remus, U. (2003) ‘Implementing process-oriented knowledge management strategies’, Journal of Knowledge Management. Emerald, 7(4), pp. 62–74. doi: 10.1108/13673270310492958.

Maldonado-Guzman, G., Marin-Aguilar, J. T. and Pinzon-Castro, S. Y. (2017) ‘Knowledge Management in Mexican Manufacturing Small Business’, International Journal of Business Administration, 8(2), p. 47. Available at: http://www.sciedu.ca/journal/index.php/ijba/article/view/11220/6873.

Mishra, J. K. (2009) Knowledge Management: Complexity, Learning & Sustainable Innovation. Global India Publications.

Natarajan, G. and Shekhar, S. (2000) Knowledge management: Enabling business growth. Tata McGraw-Hill.

Nonaka, I. (1994) ‘A Dynamic Theory of Organizational Knowledge Creation’, Organization Science, 5(1), pp. 14–37. doi: 10.1287/orsc.5.1.14.

Nonaka, I., von Krogh, G. and Voelpel, S. (2006) ‘Organizational Knowledge Creation Theory: Evolutionary Paths and Future Advances’, Organization Studies. journals.sagepub.com, 27(8), pp. 1179–1208. doi: 10.1177/0170840606066312.

Nonaka, I. and Takeuchi, H. (1995) The knowledge-Creating Company : How Japanese Companies Create the Dynamics of Innovation. New York: Oxford University Press.

Omotayo, F. O. and Authors (2015) ‘Knowledge Management as an important tool in Organisational Management: A Review of Literature’. Available at: http://digitalcommons.unl.edu/libphilprac/1238/.

Rhem, A. J. (2017) Knowledge management in practice. Florida: CRC Press, Taylor & Francis Group.

Rubenstein-Montano, B. et al. (2001) ‘A Systems Thinking Framework for Knowledge Management’, Decision support systems. Elsevier, 31(1), pp. 5–16.

Russel, B. (1961) ‘Part V:The Epistemologist.’, in Dennon, E. R. (ed.) The Basic Writings of Bertrand Russell.

Serrat, O. (2017) ‘What Does Knowledge Management Mean To You?’ Available at: https://www.academia.edu/25630177/What_Does_Knowledge_Management_Mean_To_You.

Spender, J. C. (2015) ‘Knowledge Management: Origins, History, and Development’, in Bolisani, E. and Handzic, M. (eds) Advances in Knowledge Management: Celebrating Twenty Years of Research and Practice. Cham: Springer International Publishing, pp. 3–23. doi: 10.1007/978-3-319-09501-1_1.

Standard, A. (2005) ‘5037-2005: Knowledge Management- A Guide’, Standards Australia, September.

Thomas, D. and Laurence, P. (1998) Working knowledge: How organizations manage what they know, Harvard Business School Press, Boston, Massachussettes.

CHAPTER 3: A HISTORY OF KNOWLEDGE MANAGEMENT THE 1960s TO 2019: THE MISSING BITS.

(This chapter is the reformatted version of the paper under review at the *“International Journal of Knowledge Management Studies”.)*

Abstract: The history of Knowledge Management (KM) seems to generate as many differing perspectives as the definition of the concept. As Knowledge Management emerged as a veritable force in both industry and academia, many scholars and professionals have written about its origins. The history of KM in the literature is fragmented (Spender, 2015; Maindze, Jennions and Skaf, 2017) and inconsistent on when it was created or who created it. This paper examines the history of Knowledge Management and It tells a sequential story of the concept from the building blocks to the mature concept of ‘Knowledge Management’, including the challenges that came with it as well some of the major events that appear to have had an influence on both its history.

3.1 Introduction and Scope

The history of Knowledge Management (KM) seems to generate as many differing perspectives as the definition of the concept. As KM emerged as a veritable force in both industry and academia, many scholars and professionals have written about its origins. Scholars and professionals have been debating “Knowledge Management” over the years and have had no consensus on a common definition; with each carving some early definitions to meet the wants and needs of their particular constituents (Peter, 2009; Girard and Girard, 2015). This lack of consensus is also reflected in the way that the history of KM has been captured in the literature to date. The history of KM in the literature is fragmented (Spender, 2015; Maindze, Jennions and Skaf, 2017) and inconsistent on when it was created or who created it. It tells a sequential story of the concept from the building blocks to the mature concept of ‘Knowledge Management’, including the challenges that came with it as well some of the major events that appear to have had an influence on its history. Although a variety of definitions of the term

Knowledge Management have been suggested, this paper will use the definition suggested by (Maindze, Jennions and Skaf, 2017) who described KM as "*… a multidisciplinary science and process of organizational knowledge creation and retention that engages people, process, environment and technology to create, retain or increase value."* This definition presents KM as a field of knowledge/study or academic discipline enriched by organisational processes. In (2017), Sahoo, Pati, and Mohanty used a structured framework outlining an academic discipline on KM and discovered that KM meets and supports all of the established criteria because its own body of knowledge, specialised journals, academic curricula, professional societies, and being a progressive academic discipline and scholarly field of study with its own tradition and history. Thus, this paper treats KM as a discipline and encapsulates the characteristics as found in the literature.

The history of KM indicates that KM was treated not only as power, but as a state secret that conjured dominance, competitive edge and influence. In early years of KM, it echoed a strong inclination to technology to the extent that it was almost synonymous to implementing a piece of technology, be it hardware or software (Davenport, Long and Beers, 1998; Mårtensson, 2000; McMahon, Lowe and Culley, 2004; Sandars, 2004) and in other words, the success of KM initiatives depended on these technologies (Maier, 2007). This is because most organisations predominantly adopted the technology-driven approach to KM (Mertins, Heisig and Vorbeck, 2003). This paper presents new insights into the history of KM from the early 1960s and early 1970s when the coinage of the term 'Knowledge Management' is found in some literatures. It reveals how KM over the years was shrouded in secrecy and subsumed by technology. The paper also finds a history of KM filled with scepticism and engulfed by an 'intellectual paradox'. It further explains the origins of the commonly known trilogy of people, process, and technology that is characteristic of Knowledge Management discourse, research and practice.

This article does not delve into the discussion and debate surrounding the concept of knowledge. It tries to narrate the origins of the KM [Discipline] from a deliberately short and holistic view. This paper traces the introduction of the phrase and concept of “Knowledge Management” into functional and theoretical literature and discusses its evolution from then to the year 2019. It attempts to make a clear distinction between Knowledge Management “a collective phrase for a group of processes and practices used by organizations to increase their value by improving the effectiveness of the generation and application of their intellectual capital” (Marr *et al.*, 2003, p. 773) and “Knowledge Management” discipline and draws attention to the events leading up to the emergence of KM as a discipline. The linguistic form of KM, as a discipline, practice or process has been partly the source of the challenges expressed above. Since the coinage of the concept knowledge management, there has been no clear distinction between the process of managing knowledge - “knowledge management” and the “Knowledge Management” Discipline. As a result, it is hardly obvious which of both is being referred to in the literature. This probably explains why many authors often engage and try to answer the question of “what is knowledge?” when writing about the Knowledge Management discipline. The definitional problem or “intellectual paradox” can be argued is or was a function of this phenomenon. The intellectual paradox is defined as *“…the phenomenon whereby professionals and academics acknowledge a concept, practice it, write about it, and promote its relevance individually but as a collective lack a consensus on exactly what it is”* (Maindze, Jennions and Skaf, 2017:522). In this paper, the following form: ‘Knowledge Management’ refers to the domain or discipline of Knowledge Management and not the process.

In summary, this paper answers the question on the origins and evolution of the Knowledge Management discipline; thus providing a new perspective on the history of the discipline. The paper provides new empirical evidence on earlier origins of “Knowledge Management”, and its initial foundational views – People, Process and Technology. It also shows the implications of political and technological innovations on the evolution and propagation of “Knowledge Management” literature. Understanding the history of the KM Discipline can help provide a solid foundation for facilitating the designing of study materials, organising perceptions of KM and the development of KM strategies and KM systems.

3.2 State of the Art

The history of Knowledge Management and its origins has remained contradictory (Jasimuddin, 2006), fragmented and varied amongst scholars and professionals (Spender, 2015) as has been its definition for the last 20 plus years (Maindze, Jennions and Skaf, 2017). Scholars have disagreed about the origins of “Knowledge Management” as a term or discipline. This has been as a result of the field’s multidisciplinary/ multi-domain nature (Gupta, Sharma and Hsu, 2004; Wallace, 2007). According to some scholars, the KM concept has existed since the days of the Greek Philosopher Plato (Nazim and Mukherjee, 2016) but it was not until 1986 that Karl Wiig brought the concept to light (Sveiby, 2001; Grant, 2003; Jasimuddin, 2006; Edwards, 2015) by first using it at the International Labour Organisation Conference (Liebowitz, 2005; J S Edwards *et al.*, 2009; Easterby-Smith and Lyles, 2011; Lambe, 2011; Nazim and Mukherjee, 2016; Razmerita, Phillips-Wren and Lakhmi C. Jain, 2016). Karl Wiig (2000), Koenig, and Neveroski (2008), Dalkir and Liebowitz (2017) suggested that the term KM was gradually introduced in the 1980s amidst management uncertainty.

However, in an earlier publication Karl Wiig (1997b) suggested that knowledge management practice can be traced to Chaparral Steel in 1975 and later, Prusak (2001) concludes that the term knowledge management can be rooted in the early 90s, in an article titled “Where did Knowledge management Come From”. This position is further supported by Syed et al. (2018) and by Gandhi (2004) who describes KM as a business trend that emerged in the 1990s, and by Razmerita et al (2016). Easterby-Smith and Lyles (2011)(Day, 2001; Frankland, Amjad and Nolas, 2005) support the 1990s origins perspective, but maintain that concept of knowledge management did not exist until the mid-1990s. McInerney and Koenig (2011) suggest that from an operational perspective it only appeared in the Mid-90s, but claim that the term “knowledge management” was coined in the 1980s by the Dean of the School of Information Studies at Syracuse University and further attributes the first applications of knowledge management to Ernst & Young citing Davenport (1994).

According to Parsons (2004)**,** "Knowledge Management" is a product of Peter Drucker's 1988 management contributions embedded in his ideas of the knowledge worker and the knowledge-based economy. This assertion, by implication, equates the combination of "knowledge worker" and "knowledge-based" economy to knowledge management. This hints at the challenge of defining knowledge management and its origins. Parsons further maintains that it was after these contributions that the term knowledge management become a regular term in business lexicon leading to the main knowledge management works of the 1990s such as Nonaka and Takeuchi's *The knowledge Creating Company* in 1995, and Davenport and Prusak's *Working Knowledge* in 1997. Table 3.1 below details some of the key milestones in KM literature and practice. Noticeably missing are the contribution of Holsapple (1987), Kellog (1982) and Wiig (1986) as will be shown in the next sections. This highlights the challenges surrounding knowledge management and its history.

Table 3.1 Key Milestones in Knowledge Management 1971-2018 (Parsons [2004:p.20] adapted/updated)

Year	Book /Journal/article/Conference	Author/Publisher/Sponsor
1971	VINE	MCB UP Ltd
1982	'Knowledge Management: A Practical Amalgam of Knowledge and Data Base Technology'	Kellogg C
1986	'Management of Knowledge: Perspectives of a New Opportunity.'	Wiig, K. M.
1987	The First Book Relating to KM is Published in Europe (Sveiby & Lloyd: "Managing Knowhow").	Wiig, K. M.
	'Adapting Demons to Knowledge Management Environments'	Holsapple, C.W
1988	"The Coming of the New Organisation"	Drucker P.
1989	An International Knowledge Management Network is established in Europe.	Wiig, K. M.
1991	"Brainpower"	Stewart T. A.
1992	The ACM Conference on Information and Knowledge Management (CIKM), Baltimore, November 8-11	ISMM and the University of Maryland Baltimore County
1993	Knowledge and Process Management	Wiley
	First Conference specifically devoted to Knowledge Management, Boston	Prusak L.
1995	*The Knowledge Creating Company*	Nonaka and Takeuchi
	The Knowledge Management Forum is started on the Internet.	Wiig, K. M.
	Wellspring of Knowledge	Leonard
1996	*Start of The European Knowledge Management Association*	Wiig, K. M.
1997	*The New Organisational Wealth*	Sveiby K. E
	Intellectual Capital: The New Wealth of Organizations	Stewart T. A
	Intellectual Capital	Edvinsson and Malone
	Knowledge and Process Management (KPM)	Wiley Online Library
	The Journal of Knowledge Management (JKM)	Emerald (United Kingdom)
1998	*Working Knowledge*	Davenport and Prusak
2000	*Enabling Knowledge Creation*	Von Krogh, Ichijo and Nonaka
	The Knowing-Doing Gap	Pfeffer and Sutton
2001	*The Wealth of Knowledge*	Stewart T. A
2002	*Journal of Information and Knowledge Management*	World Scientific Publishing Co.
2003	*Knowledge Management Research & Practice*	Taylor & Francis
	Electronic Journal of Knowledge Management (EJKM)	Academic Conferences Limited (England)
2005	*International Journal of Knowledge Management*	IGI Global
	Encyclopaedia of Knowledge Management	Schwartz, David , IGI Global
2006	*Interdisciplinary Journal of Information, Knowledge, and Management (IJIKM)*	Information Science Institute
2009	*International Journal of Knowledge Management Studies (IJKMS)*	InderScience Publishers
	Knowledge Management and ELearning (KM&EL)	University of Hong Kong
2013	*Knowledge Management*	Common Ground Research Networks
2018 (Nov)	*ISO 30401:2018 Knowledge Management Systems — Requirements*	International Organization for Standardization (ISO)

According to Parsons, knowledge management in the 1990s was about conference organisers and software vendors- championed crusades using technology for capturing and disseminating tacit and explicit knowledge. Apart from the proliferation of articles, books, and conferences, many multinationals such as Shell, British Petroleum (BP), Chevron, Hewlett Packard were early adopters of knowledge management (Quintas, 2005). Writing about the basics of knowledge management, Liebowitz (2010:p.3) argues "*Even though the term was coined in the early 1980s, the underlying principles really weren't adequately conveyed until the mid-1990s when Web-based and intranet technologies were becoming more commonplace in organizations*." The author further recognised the challenge of a definitional consensus but pointed out that most practitioners and academics do agree that the capture, sharing, application and creation constitute the four main processes of knowledge management.

Other studies, in their discussion of the history of knowledge management, seem to have used Knowledge Management (Discipline) and 'Knowledge Management (Process) interchangeably and concluding that "the history of managing knowledge goes back to the earliest civilisation" Ives, Torrey and Gordon (1997, p. 269) citing Karl Wiig's 1997 publication. Others have argued that KM came about in the 1990s (Tuomi, 1986; Mertins, Heisig and Vorbeck, 2003; Schultze, 2006; Spender, 2015). Some scholars agree with this view but suggest that the origins of knowledge management rest on the business improvement process (Metaxiotis, Ergazakis and Psarras, 2005) and then revert to the debate around Knowledge. The idea that much of Knowledge Management (discipline) existed before the actual term came into popular use (Liebowitz, 2005; Dalkir, 2013a) also points to the process versus discipline phenomenon in Knowledge Management literature.

Together these studies provide important insights into the theoretical and operational history of knowledge management. The studies confirm the challenges of defining knowledge Management, determining its origins as a discipline and distinguishing between KM and the concept of Knowledge. In the following sections, this paper tells the story of knowledge management from the

1960s to 2019. The section is organized as follows: Section 3: The Early Years of Knowledge Management– 1960 -1979, Section 4: Dissemination Era

3.3 The Early Years of Knowledge Management: 1960-1979?

This period was characterised by the digital revolution, the birth of the information economy, and the emergence of home computing, and the Cold War. The “email” was invented in 1971 by Raymond (Ray) Tomlinson (Spicer, 2016), Microsoft in 1975 (Naveed, Watanabe and Neittaanmäki, 2018; Oppitz and Tomsu, 2018; Kumar, 2019), and Apple computers in 1976 (Campbell-Kelly, 2018; Naveed, Watanabe and Neittaanmäki, 2018) . These developments had both a direct and indirect impact on the development and propagation of knowledge management. One may also argue that the Cold War accounts for the classification of knowledge management ideas that were developed at the time. Knowledge seems to have been treated as power to the extent that Nicholas L Henry remarked, “It has become a commencement address cliché in the technological societies that knowledge really is power” Nicholas L. Henry (1974, p. 189).

3.3.1 The Building Blocks

The management of knowledge is rooted in earliest civilizations (Wiig, 1997) and “the study of human knowledge is as old as human history itself” (Nonaka and Takeuchi, 1995, p. viii). In a study of the relations between knowledge and action, Dewey J. (1929) intimated that ideas, values, hypotheses, and others are tools for the reforming the environment in order to produce solutions to problems. He argued that "*knowing is intelligible only as the outcome of the activities by which we come to know the things we do*" (Dewey, 1929, p. xi). His perspective laid emphasis on the practical side of knowledge application. Clearly, prior to 1960s knowledge management existed rather as an unconsolidated set of principles. These principles can be traced back to ancient times. They were in use in ancient Greece more than 4000 years ago and Greek philosophers like Plato and Aristotle debated the character or qualities of empirical knowledge and learning in 400-300BC (Serenko, 2021). Under these principles, knowledge was treated as a valuable consultancy strategies (Wilson, 2002). Some of these principles include:

- Storytelling – a narrative techniques which spans the history of humans,
- Meetings, teams and communities
- Organisation, interpretation and accessing (Koenig and Neveroski, 2008; Dalkir and Liebowitz, 2011)
- Communication (Khasseh and Mokhtarpour, 2016)
- Capture, Storing and sharing (Ives, Torrey and Gordon, 1997)

Storytelling and meetings form the base of tacit knowledge which is among the most important issues or concepts in the field of KM (Khasseh and Mokhtarpour, 2016). The US army for example has storytelling and predominantly paper based knowledge management for the past 200 years and the US army is designed to easily transform tacit to explicit knowledge and transfers with ease through the ranks (Lausin, Desouza and Kraft, 2003).

Furthermore, these foundational principles are captured in the recent discoveries in Elba-Syria and Alexandria-Egypt. In Syria, 4000 years old archives of Sumer and Akkad are some of the recent discoveries representing early attempts at organisation records of civilization, government and commerce in order to prevent loss and transfer high value information from generation to generation. The desire to preserve knowledge for posterity is further exemplified by the great Library of Alexandria in Egypt, a library of antiquity founded in the Third Century BC and lasted almost 1,000 years. The library at its prime, boosted more than 500, 000 of hand-written works that were copied and shared globally. New technologies were also regularly developed to improve the efficiency and effectiveness of capture, storage and dissemination of knowledge (Ives, Torrey and Gordon, 1997). These laid the foundations for explicit knowledge management.

In the 60s, organizations were already concerned with how they managed their data and knowledge resources, particularly the capture, storage, and organisation of information for easy access. Also, there was also an ongoing debate about data, information, and knowledge- their definitions and differences. According to Kellogg (1960) at this time, companies were already experiencing increased information needs resulting from sustained growth. The loads of paper-based information had seen diminished analytical, integration and usage ability. He then proposed a logical linguistic structure for capturing, storing and managing information, that would allow for timely search and retrieval of specific and reliable information when and where needed. The proposed structure was referred to as the “Fact Compiler” – which introduced the concepts of metadata creation, organizational charts and file plans.

Two years later Machlup (1962, p. 14,15), defined knowledge as “*a state of knowing … produced by activities such as talking plus listening, writing plus reading, but also by activities such as discovering, inventing, intuiting*.” In addition, the author explained knowledge as a commodity and suggested that 29% of the US GDP was a function of the Knowledge industry. Machlup further drew a distinction between knowledge and information using their respective verb forms. To inform: “to inform is an activity by which knowledge is conveyed.” To Know: “to know may be the result of having been informed” (Machlup, 1962, p. 15). He described “Information” as that which is being communicated becomes identical with “knowledge” in the sense of that which is known. According to him, information can be used in three different ways: intellectual, pastime, and unwanted knowledge. “*All information in the ordinary sense of the word is knowledge, though not all knowledge may be called information*” (Machlup, 1962, p. 15). He further differentiated between practical knowledge, intellectual knowledge, small-talk and pastime knowledge, spiritual knowledge and unwanted knowledge. Michael Polanyi (1967, p. 4) argued that we should start from the fact that ‘*we can know more than we can tell*‘. He called the pre-logical phase of knowing as ‘tacit knowledge’.

By 1967, the challenges of creating and sharing knowledge were growing. In their review of knowledge creation, dissemination and use in education, Eidell (1967) pointed out that making existing knowledge available to potential users had become a challenge across various institutions in the US – the military, public welfare and education. They argued for the development of document managements systems and agents to address the challenges. They proposed specialist roles to facilitate knowledge dissemination- getting the right information to the right people when needed. They called this specialized role “Linker” – responsible for facilitating the flow of knowledge between creators and users.

Furthermore, Marschak (1970) introduced the notion of communication economics. They argued that delays in the creation and dissemination of information would increase costs by way of diminished value resulting from decisions made on obsolete or distorted information. The cost of a slow pace of transmission would reduce the benefits of knowledge regardless of its high value profile. They concluded that it was imperative to have an effective knowledge-creating, storing, dissemination and usage system.

3.3.2 The Introduction of Knowledge Management as a Discipline

The noun “Knowledge Management” is introduced in May/June 1974 by Nicholas L. Henry (1974, p. 189) in a report for the US Department of Commerce that defines it as “*public policy for the production, dissemination, accessibility, and use of information as it applies to public policy formulation”.* This context-specific, non-prescriptive definition seemed to have set the precedence for various subsequent definitions of Knowledge Management and the foundations for a lack of consensus. Also, he argued that information technologies were important to the extent that they could not be separated from policy outcomes and policy-making styles. He further added that we must first appreciate the role of knowledge in influencing policy results in order to understand the width and breadth of Knowledge Management including the importance of information technology. Again using Knowledge Management as a noun, Nicholas L. Henry (1974, p. 194) writes “*Knowledge Management is a, if unfamiliar, field*”, thus officially giving birth to, or describing Knowledge Management as a domain/discipline.

The emergence of the Knowledge Management can be traced back to 1976 In a classified report titled “Managing Knowledge as a Corporate Resource” submitted to the US Department of Defence in May, Berry (1976) offered a second definition of knowledge management, distinguished between data, information and knowledge and further proposed three views of knowledge management. To them, knowledge management helps maintain the value of data and went on to describe knowledge management as that which is *concerned with the formal structuring of the processes of knowledge preservation and sharing with future generations* (Berry and Cook, 1976, p. 6). This report suggests that knowledge management had emerged earlier than has been reported in most of the knowledge management literature. This was at the peak of the Cold War and the need to sustain a competitive economic, political and technological advantage shaped the treatment of knowledge. Peter Drucker (1995, p. 16) sums up the extent of value given to Knowledge Management supporting this perspective as follows *Knowledge has become the key resource, for a nation's military strength as well as for its economic strength"* ". This probably explains why knowledge management practice and documentation by the US Government was treated as a state secret and thus 'classified' and was only declassified in later years. The distribution statement of the key report from the Department of Defense Washington DC website shows the report was only "Approved For Public Release" (DTIC, 2021). This lable is applied to documents or reports that had previously been subject to "Any information or material that has been determined by the United States Government, pursuant to an executive order, statute, or regulation, to require protection against unauthorized disclosure for reasons of national security and any restricted data, as defined in paragraph r or section 11 of the Atomic Energy Act of 1954 [42 USC 2014(y)]."

Furthermore, Berry (1976, p. 4) they define data as "*a collection of attributes and values which describe an entity of interest to the enterprise*." They added that it is crucial in the implementation of knowledge management to understand the difference between processing notions and the ideas of entities and attributes. Defining information as "*the aggregate of the associations and relationships among the entities in the database …the relationships among real-world objects which are represented in the database…objects which the enterprise seeks to store and retrieve*", Berry (1976, p. 4,5) argued that data by itself was worthless, information had some value, but the life of the organisation was dependent on knowledge. And this knowledge they defined as "*the data, the relationships that exist among the data items, the semantics of the data (i.e., the use to which the information is to be put), and the rules and conditions which have been* *established as applying to the data of the enterprise*" (Berry and Cook, 1976, p. 4,5). Organisations receive data, produce information and then abstract knowledge as depicted in Figure 3.1 below, as suggested by Berry and Cook (1976).

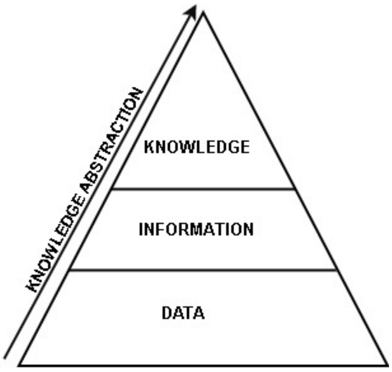


Figure 3-1 The Knowledge Abstraction Pyramid (Berry and Cook, 1976)

3.4 The Pillars of Knowledge Management

The trilogy of People, Process and Technology was introduced by Berry and Cook (1976, p. 10) to explain the data management views that form the foundations of knowledge abstraction. To these authors, one group's information outputs may be another group's data and through a different process, the knowledge created by one group can similarly be disseminated as information to other groups. They proposed three views to knowledge management from the foundation of data management that would characterise the discussion and practice of Knowledge Management: the Structural view, the Functional view, and the Physical view outlined below:

- *The Functional View*: It is based on the organisational roles and functions. By this view, the knowledge of the organization must service several different kinds of users. It emphasizes the various kinds of user-functions performed with and Enterprise regardless of the equipment or organisation involved. This represents the **People** element.
- *The Structural View*: This view is based on architectural models. It describes the organisation which the enterprise uses to accomplish the management of its knowledge resource. Flow of data from process to

process. This view represents the **Process** component of Knowledge Management.

- *The Physical View*: It reflects the current hardware/software environment of the organization. This view depicts the physical or the machine environment in which the data is processed. It lists the tools which are available to the various people identified in the structural view to enable them to do their jobs. The physical view is characterised by **Technology**.

These knowledge management views gave rise to the trilogy of knowledge management, commonly referred to as people process and technology depicted below. The views are usually represented by a Venn diagram (Figure 3.2) which is found in most if not all Knowledge Management literature and represents a reference for knowledge management practitioners and professionals.

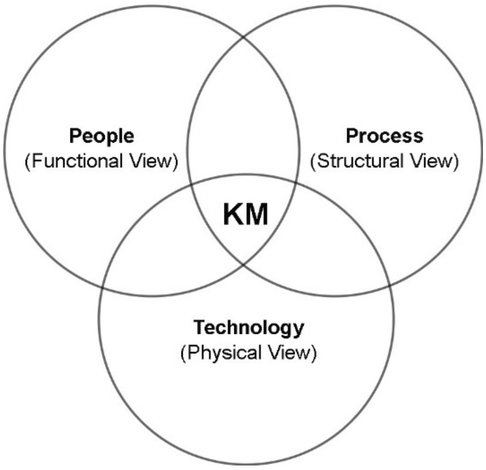


Figure 3-2 The Trilogy of Knowledge Management (interpreting Berry & Cook, 1976).

Finally, it is clear from this view that technology was an inherent part of Knowledge Management. This makes Knowledge Management today inseparable from technology (Holsapple, 2005). The basic building blocks of People, Process and Technology were defined from the outset during 1960s and early 1970s. It is therefore not surprising given that period was characterised by numerous technological innovations and disruptors such as the creation of Microsoft and Apple computers. Two organisations actively pursuing Knowledge

Management, as we know it today, were the US Department of Commerce (Nicholas, 1974) and the Department of Defence, Washington DC (Berry and Cook, 1976). However, the practice of KM by these organisations was classified (Nicholas, 1974).

In a technical report on information management, Athanassiades (1978) presents the debate on the 'knowledge revolution', its impact on libraries and other information systems and their respective managers. The debate yielded two opposing views: to either reduce the need for information or increase the capacity for its processing. Furthermore, in this debate the term Knowledge Managers is also used interchangeably with “librarians”, “information managers” and “knowledge specialists” whose role in and organisation would be “… *to identify, find, and deliver appropriate information to the teams or institutions they would serve*” (Athanassiades, 1978, p. 1). The role of the knowledge manager is clearly defined and would provide the basis or template for defining future Knowledge Management roles in most organisations.

By the end of this first period, though most of the key features of knowledge management were defined, it had not yet hit the mainstream as a discipline, process, or practice. It was the events of the subsequent period 1980-1999 that would transform and strengthen KM as a multidisciplinary discipline, process and practice. As will be demonstrated in the next period, it faced a few challenges: with some being sceptical about its chance of survival to the lack of a universally accepted definition.

3.5 Dissemination Era: 1980 -1999

This period was marked by continuous developments in the Physical View of Knowledge Management, knowledge-based systems or expert decision support systems to support business analytics and decision-making. Although attention was given to other views of KM, Information Technology, data processing and data storage were the main emphasis of KM- researchers, academics and practitioners (Hislop *et al.*, 2018). The first internet service providers started servicing customers (Johnson and Maltz, 1996) during this time. It was during this period that Tim Berners-Lee submitted a proposal for a distributed system leading to what is known today as the www (World Wide Web) in 1989. In 1991 the www software was released along with own browser - ‘line-mode’ browser (Berners-Lee and Fischetti, 2001). The impact of these developments on the development and propagation of Knowledge Management cannot be overemphasized. The basic tenets of Knowledge Management and organisational learning were developed during this period (Sveiby, 2001), which also saw a significant growth and innovations in information technology. Furthermore, it can be argued that the foundation for transitioning to a knowledge-based economy, though doubted by management (Wiig, 1999, 2000) were laid during these years.

The pioneers both unacknowledged (Lambe, 2011) and acknowledged were active in their respective domains writing and discussing “Knowledge Management”. However, all of them did not gain the same visibility, leading to discrepancies on the reporting surrounding the origins and naming of knowledge management. For instance, the first use of the term “Knowledge Management” in a presentation is attributed to Karl Wiig’s keynote speech at a European Management Conference sponsored by the International Labour Organisation of the United Nations (Beckman, 1999; Aguirre, Brena and Cantu, 2001; Nunamaker, Romano and Briggs, 2001; J S Edwards *et al.*, 2009; John S Edwards *et al.*, 2009; Ismail and Yusof, 2010; Grant, 2011). Although it propelled Knowledge Management further to a wider and influential audience, it was not the first use of the term (Sveiby, 2001; Lambe, 2011)*.* Other key players or pioneers with significant contributions have not been given the deserved acknowledgement regarding the creation and propagation of the concept of knowledge management (Lambe, 2011). This may be due to the fact that the print media was the main avenue of knowledge and information dissemination at the time. Though the world was entering the ‘information age’, it was transmitting academic contributions at ‘snail pace’. Worth noting here is the fact much of what happened in the late 1960s and the early 1970s classified and treated knowledge as “power” – or as put by Sir Francis Bacon “Knowledge is power” in (Dziuban, Moskal and Hartman, 2005). A quick survey of two academic data bases ABI/inform and Scopus respectively in Figure 3.3 below shows a slow but steady increase in the 1980s and surge in the Knowledge Management literature in the mid-90s. It therefore is no coincidence that many academics and practitioners have attributed the origins of “Knowledge Management” [the term and discipline] as originating in this period. From the sharp increase in literature, it can be concluded that KM gained global prominence in the mid-90s.

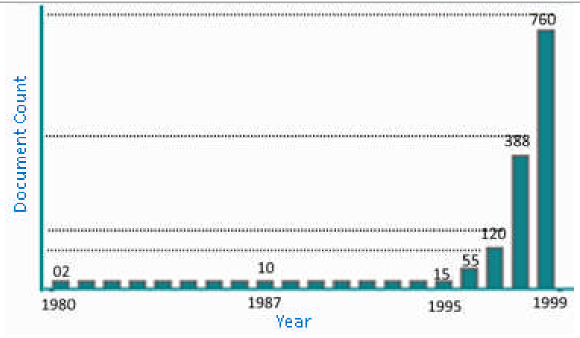
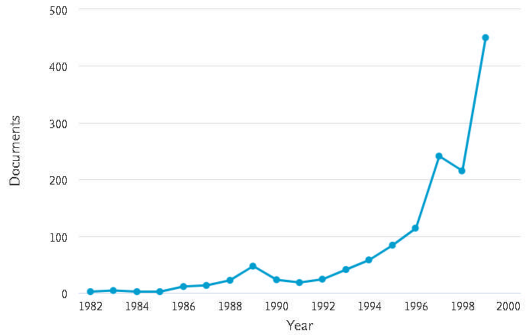
	
ABI/inform -1,396	**Scopus -682**

Figure 3-3 Growth in KM Literature 1980-1999

The search string shown below used to achieve the trend results shown above also revealed some key insights into the history of Knowledge Management. Since the late 1980s and the early 1990s there has been a visible growth in the number of publications about Knowledge Management.

TITLE-ABS-KEY ("KNOWLEDGE MANAGEMENT" OR KNOWLEDGEMANAGEMENT OR KNOWLEDGE-MANAGEMENT) AND PUBYEAR > 1979 AND PUBYEAR < 1990

The literature on the evolution of “Knowledge Management” often points vaguely to the 80s or mid-90s with founding fathers like Peter Drucker, Karl Wiig, Nonaka & Takeuchi, Larry Prusak, Thomas Davenport, Peter Heisig, Karl-Erik Sveiby and Peter Senge. Some writers attribute the first use of the term ‘Knowledge Management’ to Mckinsey in 1987 (McInerney and Koenig, 2011), with others suggesting it did not go public until 1993 (Prusak, 2001). This research shows (Table 3.2 - below) that there were a good number of publications prior to the 1986 outing by Karl Wiig and Karl-Erik Sveiby. In 1982, Charles Kellogg presented a paper titled “*Knowledge Management: A Practical Amalgam of*

Knowledge and Data Base Technology" at the Second AAAI Conference on Artificial Intelligence. This publication focused on the implementation (Figure 3.4 below) of the Structural View of Knowledge Management as an effective strategy for knowledge and expert systems.

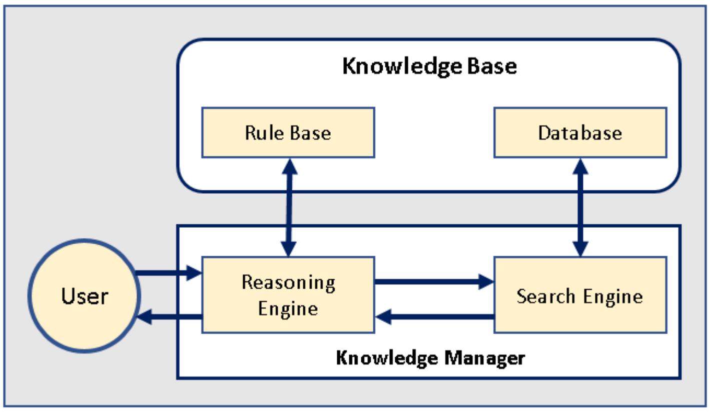


Figure 3-4 Knowledge Management System Architecture (Kellogg, 1986)

By 1999, Holsapple had published more than 14 Publications from 1979 focusing on the structural elements of Knowledge Management. He wrote extensively on decision support and information systems and issues relating to their implementations. He sought to clarify the relationships between information processing, decision making, and decision support and proposed a framework for perceiving organisational information processing and decision making. There was a minimum of three publications with explicit focus and mention of the term "Knowledge Management" in both title and content, some of which are shown in Table 3.2 on next page. In an IEEE Computer Society "Trends and Applications" event, Lewis & Lynch presented "*GETREE: A Knowledge Management Tool*" a new interactive "Knowledge Management System" and probably the first to be labelled as such, that was implemented in GE Corporate Research.

Table 3.2 Knowledge Management Publications of the 80s (Sorted by year only).

AUTHOR/YEAR	EXPLICIT TITLE
Bonczek, R & Holsapple, C (1981)	Foundations of Decision Support Systems
Charles Kellogg (1982)	Knowledge Management: A Practical Amalgam Of Knowledge And Data Base Technology.
Lewis, J.W. and Lynch, F.S. (1983)	GETREE: A Knowledge Management Tool
Harold Borko (1983)	Information and knowledge Worker Productivity
Akscyn, Robert (1984)	Advanced Distributed Operating Environment for Document Development and Management.
Dolk and Konsynski (1984)	Knowledge Representation for Model Management Systems
Dan Kogan (1984)	The Manager's Assistant. An Application of Knowledge Management
Kellog Charles (1986)	From Data Management to Knowledge Management. The components of a Knowledge Management System
Van Lohuizen (1986)	Knowledge Management and Policymaking
Karl-Erik Sveiby & Anders Risling (1986)	"Kunskapsföretaget - seklets viktigaste ledarutmaning?" (The Knowledge Company - The Most Important Leader Challenge of The Century?)
Karl Wiig (1997)	"*The concept of 'Management of Knowledge: Perspectives of a new opportunity' is introduced in a keynote address at a European management conference sponsored by the International Labour Organisation of the United Nations*. (source: Karl Wiig 1999/2000 : p.10)

Because this early period was characterised by developments in the Structural View of Knowledge Management, it is no surprise that many of the publications focused on the development of decision support systems, expert systems and frameworks for their respective requirements gathering and development. Holsapple (1987) proposed a number of knowledge management techniques which he argued were key to the success of decision support systems. The techniques included database management, programming, spreadsheet analysis, and rule set management and automated inference. These techniques are still used today in knowledge management practices. In the same year, similar propositions emerged, focused on requirements gathering for Knowledge Management and systems integration. One such proposition came from

Applegate et al. in 1987 (Applegate *et al.*, 1987) who proposed a knowledge management requirements for organisational planning characterised by the integration of data management, model management and process management with a decision support environment. This approach was supported by Shen's 1987 (Shen, 1987) knowledge management framework that emphasized the value of relations and decision models. Another data-based framework on classification, interpretation and diagnosis as the primary method for knowledge-based development was recommended by Frediani and Saitta (1987). However, manufacturing companies had a different experience as they had embraced computer integration in their information management strategies, creating the need to integrate data and knowledge in their respective enterprise information models (Kellog, 1983; Hsu and Skevington, 1987). As a result of these developments, Hsu (1988) proposed a meta-database technique for data and knowledge management leading to the development of a Global Information Resources Dictionary for manufacturing facilities. The author formalized the structure of contextual knowledge and mapped the relationships between data and knowledge resources. Furthermore, the emergence of large-scale industrial systems presented a new challenge for managers of management information systems, related to the difficulty of developing large-scale databases. Filteau (1988) sought to resolve the issues by proposing a methodology for large-scale database development using an incremental development approach that involved two key steps depicted below (Figure 3.5)

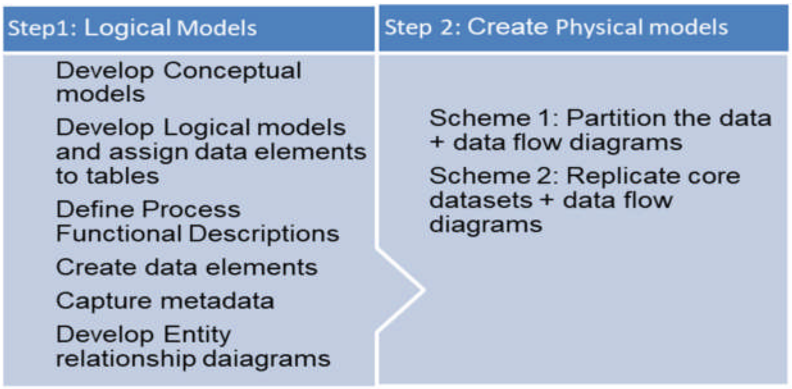


Figure 3-5: Large Scale Database Development Methodology (Filteau, Kassicieh and Tripp, 1988)

The interaction between the various views and disciplinary roots led to one of the early characterisations of Knowledge Management by Hannabuss (1987, p. 1) as "*a multi-disciplinary field*". Adding to the significance of the discipline in improving workforce performance and organisational productivity, Grossman (2007, p. 37) held the view that Knowledge Management should to be researched and included in the curriculum as an imperative "discipline". This view is later supported by Serenko *et al.* (2010) who sort its establishment as "scholarly field". At the time, Professor Clyde W. Holsapple in January (1989), following on earlier work, reviewed the significance of varied knowledge management methods and their application to the management of specific knowledge categories. Holsapple (1989) also presented an object-oriented paradigm as a basis for an integrated and harmonised manipulation of various knowledge management techniques in a unique environment. Using this context, Holsapple proposed two strategies to make the techniques available to knowledge and information professionals, specifically in "*skeletal environments*" and "*furnished environments"* (1989, p. 37). On one hand, the skeletal environments offer a 'do-it-yourself' framework to knowledge management and it is up to the knowledge management professional to provision the environment with any type of entity that might be required. On the other hand, the furnished environment gives the knowledge management professional a computerised knowledge representation and processing system. Furthermore, Huang (1989) explored the role of Knowledge Management in the creation of hierarchical models. Huang proposed a collection of knowledge bases to assist in the creation of hierarchical models. Huang provided an in-depth synthesis of knowledge acquisition and representation framework, a model generation, a functional simulation model creation, including a rule-based automatic analysis. Using a combination of entity based representation with production rules and frames acquisition process and a universal model database ensured a faster development of models for promoting the enhancements.

One of the key developments in the late 1980s, resulting from the application of Knowledge Management, was the introduction of the confirmation of the concept of 'Organisational Knowledge Management'. Paradice and Courtney in June (1989), discussing the application of Knowledge Management in organisational setting in in their book; “Organizational Knowledge Management” concluded that a substantial return on investment opportunity existed for organisations that effectively exploited artificial intelligence technologies for the management of their organizational knowledge.

Figure 3-6: Steps in Developing a Knowledge Management System (Paradice and Courtney, 1989)

They introduced three basic tenets (Figure 3.6) of knowledge management: knowledge acquisition, knowledge representation and knowledge utilisation in the context of specific organisational knowledge challenges and requirements. They explained knowledge acquisition as the process of “extracting” knowledge about issues from tacit and explicit sources – people and print and other media. Knowledge representation was referred to as the organisation and storage in computer systems, supported by artificial intelligence. Knowledge utilization was defined as the final stage that takes place after acquisition, organisation and storage: then available thereafter to people in the organisation. These principles represented the essential steps in the development of Knowledge Management Systems. Paradice and Courtney (1989) proposed a series of important areas where these innovations can be used in organisational contexts. These areas of application of fundamental technology being developed for knowledge management practices included the following:

☐ Document preparation
☐ Organisational learning
☐ Electronic mail
☐ Hypertext
☐ Intranets/Extranets
☐ Information technology management, and
☐ Group decision support systems.

Organizational challenges relating to staff resistance, competitive edge and cost of implementing information management technologies are also discussed.

Furthermore, to confirm or validate the principles of organization, storage, and utilization of knowledge, Schmoldt (1989) used two case studies, relating to fire characteristics prediction and initial-attack force dispatch, developed for wildfire prevention planning. Scarce, largely intuitive and subjective knowledge is offered to a broader community of users through expert systems that rely on the experience or training of a few important individuals. Such systems fill a gap in situations where knowledge is uncommon or scarce by providing several copies of individual expertise, and allow for the preservation of experience that would otherwise be lost through retirement or transition. Knowledge-based programs make it possible to use the latest methods and study findings more effectively. Schmoldt concluded that there was the capacity to reflect and exploit knowledge in both tacit (human) and explicit (literature and methods) forms. To be most useful, the author suggested that we must have multidimensional knowledge of real-world objects; that is, we must be able to communicate with other objects and ideas in diverse number of possible ways. Knowledge-based and expert systems have showed potential as mechanisms for processing information in a more structured and more useful decision-making format in the areas of transfer of expertise, translation of research results and application of management tools that existed.

The history of Knowledge Management (KM) might not be complete without mention of Intellectual Capital(IC) and its relationship to KM. This is because the history of IC runs in parrallel with that of KM. Since its emergence as a field in the 90s KM/IC continues grow in peer review journals and academic programs. Although the field gained its roots in the 1990s, (Serenko and Bontis, 2017) the term ‘Intellectaul Capital’ can be traced to about the same time KM was gaining momentum to becaoming a discipline. Galbraith (1967) was the first writer to use the term Intellectual Capital (Bontis, 1998) to describe a high reasoning, sharp-witted knowledge processing and usage capability. To understand the significance of IC in KM, it is important to define IC. "Intellectual Capital is the sum of everything everybody in a company knows that gives it a compatitive edge" (Stewart, 1997, p. Xix). It is "…is intellectual material---knowledge, information, intellectual property, experience---that can be put to use to create wealth" (Stewart, 1997, p. xx). Intellectual capital can be split into three broad categories: structural capital, human capital and customer capital (Karl M Wiig, 1997a; Stewart, 1997; Hussi, 2004; Seleim and Khalil, 2011), and the practice of managing IC was led by Skadia – a swedish Insurance company (Karl M Wiig, 1997a). KM represnets the management of knowledge whereas IC represnets the stock of knowledge (Alexeis *et al.*, 2020) and is functional to KM practices (Paoloni *et al.*, 2020).

IC and KM complement each other (Karl M Wiig, 1997a; Marr *et al.*, 2003; Hussi, 2004) are closely related to each other(Koenig, 1997) as depicted in figure 3.7. The three braod categories of intellectual capital are pulled together by knowledge Management to create value (Hussi, 2004). This realtionship bwteen KM and IC according to Hsu and Sabherwal (2012) (Cabrilo and Dahms, 2018) is grounded in the knowledge based view of the firm and that IC had an impact on KM na ddynamic capabilities. Furthermore, on how the knowledge-based issues relating to IC assets and KM practices affect organizational performance, Kianto et al. (2014) concluded that on one hand, IC assets mediated on the effects of KM practices on organisational perfomamnce, whiulst on the other KM practices mediated on the effects of IC assets (Nielsen, 2018) on organisational performance. This assessmment of this relationship between KM and IC is in Sync with obervations made by other authors more than a decade ealier. It shows the continued growth and relationship between KM and IC.

They only differ on their approaches to building and governing intellectual assets. Whilst KM pursues the tactical and operational perspective - planning, implementing, operating and monitoring all the knowledge-related activities such as creation, elicitation, capture, transformation and use for effective intellectual capital management, IC focuses on strategic and enterprise-iwde governance of intellectual assets. It thus clear from Figure 3.7 below that "*There is considerable overlap in the scope of intellectual capital management and knowledge* *management."* (Karl M Wiig, 1997a, p. 400). For instance the Funtional View of KM's user fucntions and expertise are similar to IC's individual's knowledge, experiences and skills which are elements of the Human Capital. There is also an overlap between Structural Capital [IC], Pyshical View [KM] as they use the same technological tools and techniques. On the whole, IC and KM seem to generate the outputs such as improving organisational performance, facil;itating knowledge sharing, creating/adding/retaining vaklue and increasing the organisation's competitive edge

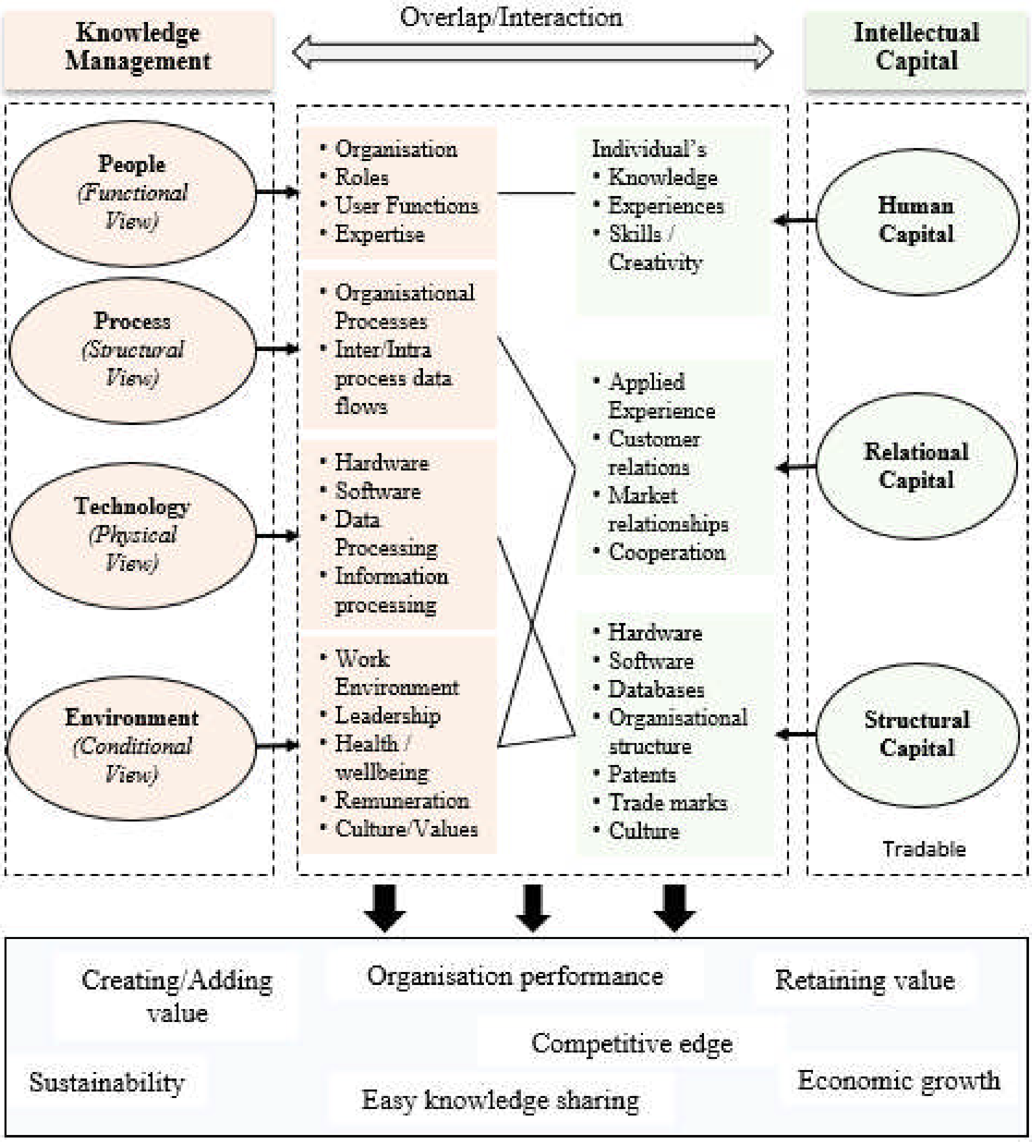


Figure 3-7: Relationship Between IC and KM

Intellectual Capital (IC) emphasizes the reconditioning and optimization of the value of intellectual assets throughout the business organisation. Knowledge Management (KM) supports advanced organisations to pursue deliberate strategies to coordinate and exploit intellectual assets by concentrating on comprehensive methodical, unambiguous methods, techniques, overlap and interaction between IC management and KM. Those strategies create balanced intellectual capital portfolios that they implement with KM approaches and tools from IC management standpoints (Karl M Wiig, 1997a; Marr *et al.*, 2003). By the year 2000, IC had gained more recognition as a true strategic asset leading to an increase in Organizational Knowledge Management Systems (OKMS) for managing intellectual capital (Meso and Smith, 2000). It had also become increasingly fundamental to measure intellectual capital winthin Knowledge Management Communities (Liebowitz and Suen, 2000). In their article "*A Critical Review of Knowledge Management as a Management Tool*" Martensson (2000) conlcuded that Knowledge Management was essentially the management of IC controlled by the enterprise.This view was upported by others who viewed KM as the activity for obtaining, growing and sustaining IC in organisations (Marr *et al.*, 2003) or an independent element of intellectual capita (Hussi, 2004). This perspective further highlights the complimentarity that continue to exist between KM and IC. Figure 3.8 below shows a three way relationship between KM and IC and strenthening the case for strong complementarity between the two concepts. Wiig in (Fink, 2003; Serenko and Bontis, 2004) concluded that it was absolutly imperative for organisations to integrate IC and KM.

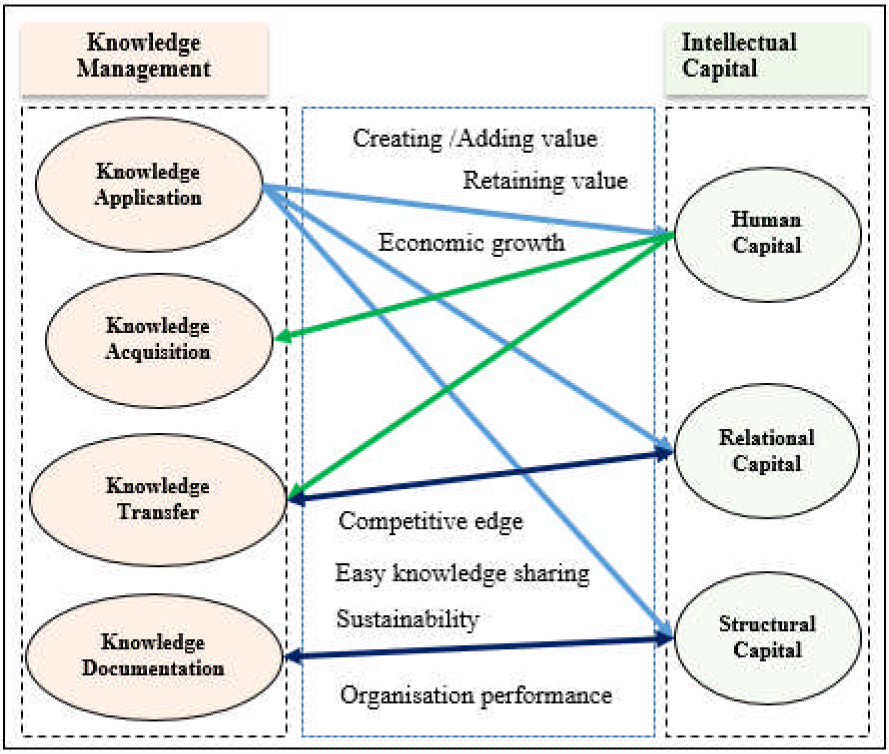


Figure 3-8 Three Way Relationship Between KM & IC [Interpreting (Seleim and Khalil, 2011)**]**

KM and IC were growing at 50% per annum since the early 1990s (Serenko and Bontis, 2004) and by 2009 it became clear that KM/IC was growing at an accelerated rate (Serenko and Bontis, 2009) and by 2010 the field boosted 20 pier-reviewed journals at its embryonic stage. In a study aimed at showing that "managing knowledge" and "leveraging intellectual capital" can yield results and agenda items, Stewart (1997) concluded that Wal-Mart, Microsoft and Toyota became great companies not because they were richer organisations with greater physical or financial assets but rather because they had Intellectual Capital. The more knowledge management practices are used, the more new knowledge of improved levels of IC assets are created leading to more innovative or improved organisational performance as the ultimate KM outcome (Handzic and Durmic, 2015). Paoloni et al. (2020) analysing the relationship between KM and intellectual capital (IC) and entrepreneurship concluded that IC is functional to KM practices and that both IC and KM deal with issues related to the knowledge-based view of the firm. Moreover, they observed that whilst IC has provides a base to evaluate the strategic and significant intangible resources, KM lays emphasis on the decision-making, management processes and practices. This relationship and interaction between KM and IC has remained an active component of knowledge management research and practice to date.

At the close of the 1980s, the International Knowledge Management (KM) Network was started in Europe in 1989 (Karl M. Wiig, 1997). The decade that followed would become the period that dominates the knowledge management literature as its foundation years, detailed in the first section of this paper. It was also during this period specifically in the 1990s that the IC concept started gaining significant grounds. The concept of Knowledge Management gained momentum and more popularity in both industry and academia in what is referred to herein as the rebirth of Knowledge Management as a discipline.

3.6 The Rebirth of Knowledge Management

Dominated by globalisation, which in itself was founded on innovations in transport and communications technology (Rennen and Martens, 2003), the 1990s saw the dominance of the Functional View of Knowledge Management underpinned by the worldwide web, wikis and communities of practice and the creation of the world's first search engine - Archie; tools that continued to play a fundamental role in knowledge management and organisational learning. Karl-Erik Sveiby (1990) published the first book with Knowledge Management in the title (Craven, no date; Wilson, 2002) and Peter Senge (1990) published the Fifth Discipline which had significant impact on the rebirth of Knowledge Management. In a personal interview with Alistair Craven (Craven, no date), Sveiby maintains that the 1990s was the decade of the internet. By this time, the relationship with Technology became even stronger. Another key development of the 90s was the formation of a consortium to provide a technological base for knowledge management called Initiative for Managing Knowledge Assets (IMKA) in the USA. Following the development of the hypertext system by Tim Berners-Lee and the subsequent launch of the worldwide web, which became publicly available in August 1991 there was increase communication. Shortly after, in 1992, Microsoft released Windows 3.1. The Knowledge Management Forum was started on the internet in 1995 (Karl M Wiig, 1997b).

Furthermore, at the start of the 1990s based on his ‘systems thinking' claim, Senge (1990) argued that organisational learning is only beneficial if it is grounded in an understanding of how the entire organisational structure is interconnected. He contrasted first class and second-class learning. He drew a distinction between 'adaptive' and 'generative' learning in distinction to Argyris (1976), who used 'single loop' and 'double loop' learning. Specifically, he argued for adaptive learning, reacting to past errors and changing future behaviour, while on the other, he argued for the imaginative aspect of generative learning and the ability to alter courses of action using new knowledge. He further suggested for organizations to become core learning organizations, 'learning disciplines‘ such as system thought, individual mastery, mental models, collective vision and team learning should be included in their activities. Personal mastery, mental models and shared vision have specific application for the individual expert, and the last two have group relevance (Senge, 1990). The five disciplines are fundamental tools for the encouragement learning within an organization and for constructing and sustaining learning organizations. In a learning organization, the individual struggles to make his or her own contributions toward a shared mission/vision of the organisation. Normal modifications in mental models are constantly challenged and refined in search of personal mastery. Individual, organisational goals are associated with the mission/vision of the organization. The key questions that followed this perspective on knowledge management was: How is knowledge created?

In this regard, Nonaka (1994) classified knowledge into explicit and tacit categories. Tacit knowledge deals with a person’s experiences and knowledge and is therefore not easy to articulate, imitate or access. This is the intangible expertise, which resides in persons’ heads or memory, and shared organizational practices. Tacit knowledge is shared through team working and collaboration in networks of people, meetings, articulated anecdotal experiences and other forms of communications. It is regarded as the most important source, which could be with an individual or organization. On the contrary, the former knowledge (explicit) comes in the form of objects that we can see such as books, documents, white papers, databases, standards and policy manuals (Nonaka, 1994). This is a more articulated and generalised type of knowledge. Despite being still used widely today, the dissimilarity between tacit and explicit knowledge has been criticised as being too simplistic (Hicks, 1995) and relative (Dalkir, 2005).

According to Nonaka and Takeuchi (1995), the prerequisite for successful knowledge management programs is the conversion of tacit knowledge - what is in our heads- into explicit- codified knowledge. Their logical conceptualisation considers organisational knowledge creation as the processes of interaction between tacit and explicit knowledge constitution. These processes refer to a continuous interchange between tacit and explicit domains of knowledge. This continuous interchange between knowledge domains is known as the SECI process or the SECI model proposed by Nonaka and Takeuchi (1995); it represents socialization, externalization, combination, and internalization. A growing spiral surges as knowledge moves through various levels. They identified four means of knowledge creation in an organisation: socialization-networking, social communication and shared experiences; externalization-verbalization of best practice or vital lessons learned; internalization-understanding resulting from interpretation written material or discussion; and combination- processing of existing explicit knowledge (1994).

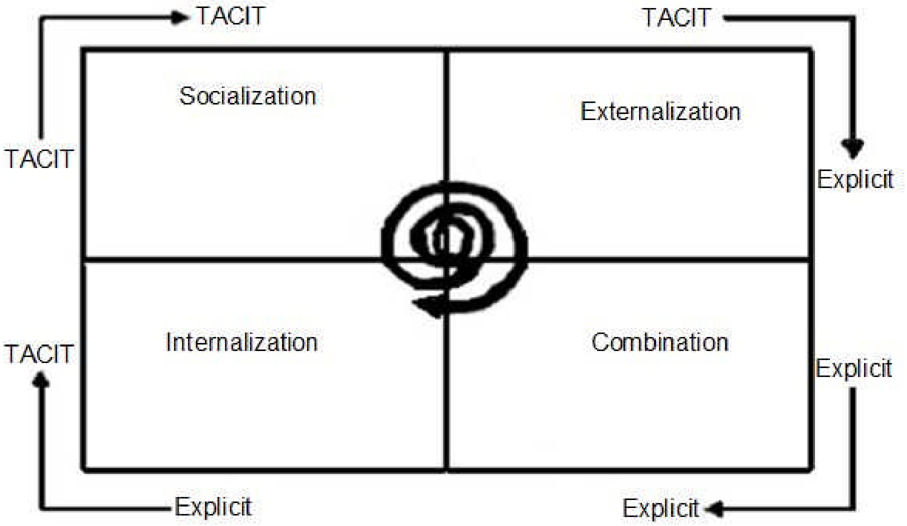


Figure 3-9: The SECI Process (Nonaka, Von Krogh and Voelpel, 2006)

Figure 3.9 above depicts the spiral interaction of the four modes of knowledge creation form a continuous cycle that is driven by changes between the different modes. Nonaka, Von Krogh and Voelpel (2006) maintain their argument that the changes are triggered by processes such as knowledge vision, driving objectives, interaction, dialogues, practices, knowledge assets, knowledge leadership or environment (Nonaka, Von Krogh and Voelpel, 2006, pp. 1180–1192). This spiral increases in size as it moves up the knowledge levels from individual to group to organization through inter-organization.

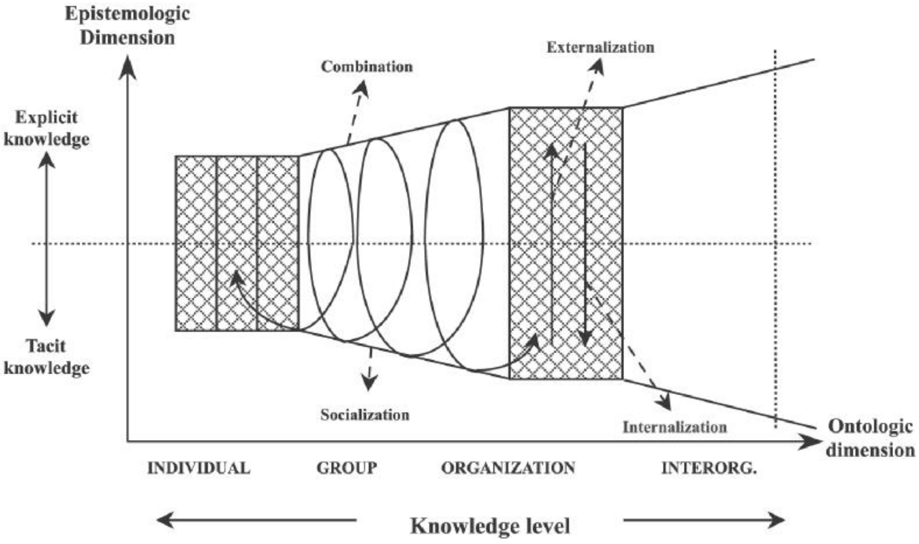


Figure 3-10: Spiral Knowledge Creation (Source: Adapted from Nonaka and Takeuchi (Nonaka, 1994)**)**

Figure 3.10 shows the upward ontological spiral creation developed by Nonaka (1994). Though the model provided for knowledge processing and addressed the static and passive perception generated, it did not take into account the type of knowledge the organizations created.

3.7 Trends in Knowledge Management Publications

It was this publication that put the authors at the centre of Knowledge Management, have since been dubbed by many as the founders of knowledge management. This however is not surprising given that soon after their publication; Knowledge Management started experiencing significant growth in publications. The evolution of Knowledge Management and publications from 1970 to early 2020 is depicted in Figure 3.11 below. This figure is indicative of the overall trend but it is by no means exhaustive because it based only on the analysis of the EBSCO database. The graph shows little or no growth in knowledge management publications in period between 1970 and 1988. During this period only a couple hundred publications were registered. In 1989, KM publications surged by about 1033% from the previous period from just a little over 150 publications to 931 publications. Another milestone in KM was the founding of the International Network of Knowledge Management in Europe. These developments were followed in 1990 by a further growth by 30% in KM publications. Overall, the graph shows a steady rate of yearly increases in KM publications. Although the increase was stable from 1990 to 1995, it soared in 1996 by 50% against the period 1990-1995. Soon after Nonaka, Knowledge Management experienced another significant shift in the number of publications from hundreds to Thousands. Another key point in this trend was 1998 that recorded 171% increase in KM publications from the year before from 1,770 to 4,797 publications.

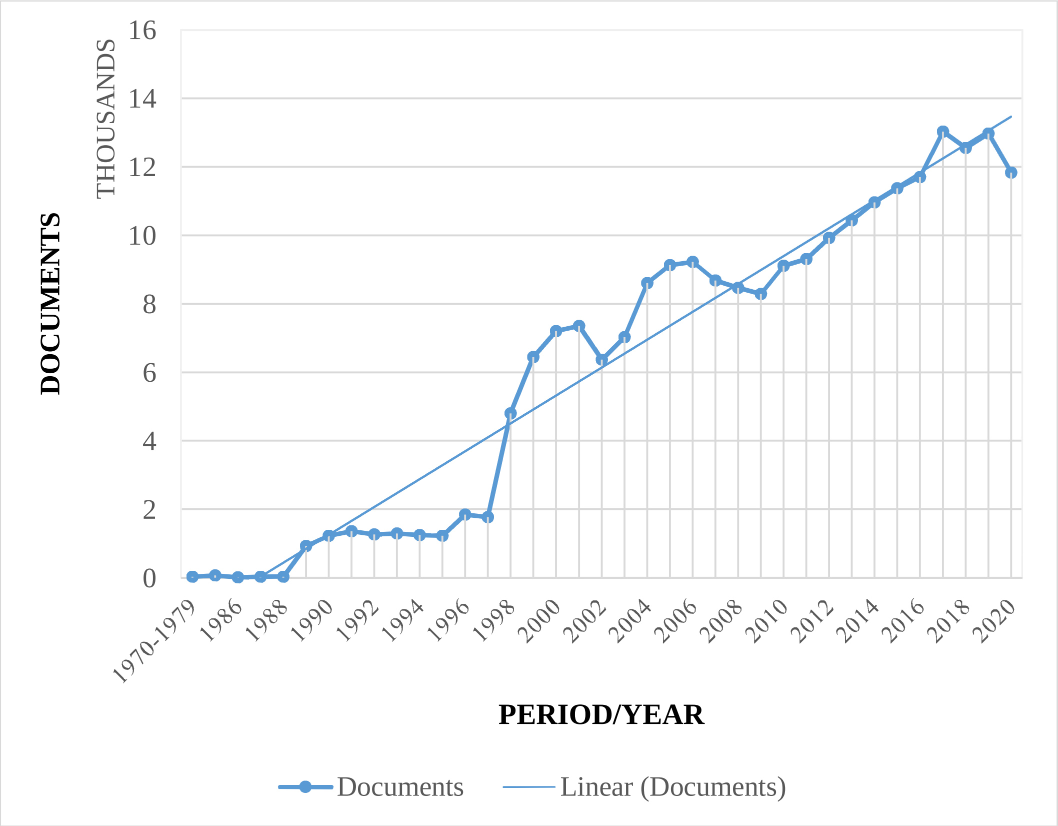


Figure 3-11: Knowledge Management Publication Trend 1970 - 2020

Furthermore, as developments in communication technologies grew, the internet being key amongst them allowed knowledge management to gain more exposure. Figure 3-11 shows surge in knowledge management publications soon after the technological innovations. Just three years after “The Knowledge Creating Company”, knowledge management gained a steady growth in the number of publications from 1,229 in 1995 to 6,443 in 1999, and by 2000 Knowledge Management publications hit more than 7000 published documents annually. This sustained and stable increase in KM publications can be argued was a result of an increased number of relevant KM and Intellectual Capital (IC) peer-reviewed journals. Some of the key KM journals established within this period includes: *The learning Organisation*, 1994; *Journal of Knowledge Management,* 1997; *Knowledge and Process management: The Journal of Corporate Transformation,* 1997; and the *Journal of Knowledge Management Practice*, 1998. By the year 2000, there were already more than 18 KM/IC peer reviewed journals in circulation and the Journal of Knowledge Management has since maintained a leading influence in KM research and practice (Serenko and Bontis, 2017). Despite the small dip in publications in 2002 Knowledge Management publications have continued to increase at a steady rate annually. This trend suggests this growth will either continue or remain stable in the future.

There seems to be a correlation between political and economic developments on the growth and reach of Knowledge Management. For instance, the cold war ended in 1991 following the fall of the Iron curtain in 1989, making globalisation a truly global phenomenon (Vanham, 2019). It was in the same year -1989 that KM publications registered the highest percentage increase in a single year to date. A key technological milestone is achieved with the creation of the internet 1991. The link between technology, and particularly the internet and knowledge dissemination was further strengthened and perhaps accounting for the continuous growth in KM publications. Furthermore, there was also an increase economic interaction between the eastern and the western countries as a result, with China joining the World Trade Organisation (WTO) in 2001. These increase interaction and economic growth meant increased middle class and academic inquiry. Considering it a coincidence or not, the growth in publications seem to be in sync with global economic and political changes. Nonetheless, this it not to say there is a direct correlation, though it can easily be associated with these global events.

3.8 Knowledge Management Developments 2000 – 2019

This period was characterised by developments in KM strategies, frameworks, culture and behaviour change approaches (Shu-Mei, 2010; Javier and Fátima, 2011; Mills and Smith, 2011) and related tools for KM. These changes were a result of the recognition of organisational culture as a major barrier in leveraging knowledge assets in early 2000s. David W. De Long and Liam Fahey (2000) and Gold, Malhotra and Segars (2001) found that culture shaped assumptions, defined relationships, created context for social interaction, and shapes the processes of organisational knowledge creation and dissemination. Consequently, there was a drive in this period to develop KM strategies and specific cultural changes to align with organisational visions, goals in order to support more effective and efficient creation and sharing of knowledge. Michael Earl (Earl, 2001) proposed a taxonomy of strategies, or "schools," for knowledge management. The proposed framework focused on provisioning executives with options for starting KM projects. These choices range from goals to organizational character, and technological, behavioural, or economic biases.

In addition, the adoption of best practices was a key feature in organisational knowledge practices. By 2000, KM was equally experiencing a proliferation of prescriptive, descriptive, or Hybrid (a combination) frameworks for Knowledge Management. Wenger and Snyder (2000) proposed Communities of Practice as an alternative form of organisation that was going to complement existing structures and radically encourage knowledge sharing, learning and transformation. They went on to define communities of practice as "…*groups of people informally bound together by shared expertise and passion for a joint enterprise…* (2000, p. 139)". It is a knowledge translation strategy that brings people together, people with a common issue in order to expand their knowledge and expertise (Liza *et al.*, 2018). Two years later in a book titled, *Cultivating Communities of Practice: A Guide to Managing Knowledge* (Wenger, McDermott and Snyder, 2002)*,* the authors provided a hands-on guidance on how to making knowledge work inside an organization using Communities of Practice. Rubenstein-Montanoa et al (2001) outlined a sample of 25 frameworks for knowledge management and a criteria for evaluating them. During this period, there was a growing recognition that those on the front line had critical knowledge that was not being captured by the experts. This lead to most organisations seeking frameworks and strategies for capturing and sharing knowledge with particular interest in its Tacit form. This is exemplified by Chris Collison, Geoff Parcell (2001) in their book *Learning to Fly: Practical Knowledge Management from Leading and Learning Organizations,* in which they proposed practical tools and techniques for implementing knowledge management akin to peer reviews. Using their experiences at BP they proposed the following tools and techniques like Connecting Sharers with Learners – Using Self-Assessment and Learning From Peers – Somebody Has Already Done It. Others included Learning Whilst Doing – Time to Reflect; Learning After Doing – When it's All Over; Finding the Right People – If Only I Knew Who; Networking and Communities of Practice; and Leveraging What We Have Learned – Capturing Knowledge.

Moreover, these techniques propagated the review of past and current actions, learning and sharing lessons from projects. They advocated After Action Reviews because, simple, timely time efficient regular reviewing of procedures was key in revealing technical expertise, improving information accessibility and learning in organisations. They also advocated 'Peer Assist'- *"…meeting or workshop where people are invited from other organisations and groups to share their experience, insights and knowledge with a team who have requested some help early on in a piece of work*" (Collison and Parcell, 2004, p. 98). This is based on the technique of Peer-assisted Learning (PAL) used in education which has been defined by Onions, (1978) in (Topping and Ehly, 1998, p. 1) as *"… to acquire knowledge and/or skill by study, experience, or teaching. To assist is to aid, help, promote, support, or succor. A peer is an equal in standing or rank, a matched companion*. According to Topping and Ehly (1998) although we had entered the information age where competence in information technology was crucial, computers alone were no panacea to knowledge sharing and learning. They observed that this technique was more cost effective, more democratic, inclusive and empowering. PAL comprises the following methods: Peer Tutoring, Peer Modeling, Peer Education, Peer Counselling, Peer Monitoring and Peer Assessment; all aiming at learning, improving and sharing. According to Krogh (2012) PAL creates the ideal conditions collective acts of knowledge sharing leading to the satisfaction of diverse and distributed knowledge and information interests.

In the early 2000s there was increased awareness of the significant innovations around the Internet and these changes in developments of the internet became known as web 2.0 (Bakardjieva and Gaden, 2012). The term was initially introduced by O'Reilly media in 1999 (O'Reilly, 2009) to emphasize user interaction and collaboration (Lai *et al.*, 2013). By 2004 Web 2.0 services and applications were already facilitating more self-motivated communications between clients and servers, including user friendly interactive webpages and applications. Web 2.0 applications enabled users with little or no technical knowledge to create and share their own knowledge and content, as they do on social networking websites (Harrison and Barthel, 2009; O'Reilly, 2009).Web 2.0 Technologies became very popular in both academic and industrial communities. These technologies had features like wikis, forums, blogs, news, podcasts, calendar, folksonomies, video, social networks, virtual worlds, chat, and glossaries (Becerra-Fernandez and Sabherwal, 2014)(Moria, 2009). The perception within various knowledge management communities was that these technologies facilitate collaboration and communication within most organisations, help with metrics. In addition, these technologies can help improve collaboration and communication across multiple organisation. Clearly, these web 2.0 technologies emphasized the promotion of the human interests- do-it-yourself, personal networks, online communities and social networking sites. These factors are central in the 'Peer Assist' and 'After Action reviews' proposed by Collison and Parcell (2001). It enables the development of agency, a sense of community and communal resources this reducing the social costs of knowledge sharing and minimizing the need for agency (Krogh, 2012). In an article emphasising the people side of KM, Ravi (2002) in addition to providing databases for capturing, storing and sharing knowledge, supported Communities of Practice as an effective and efficient way of exploiting organisational knowledge (Suhwan, Young-Gul and Joon, 2011). According to Maier (2007), Siemens had developed more than 1,600 communities of practice in 2004 and more than 85,000 users of the company's KM system built on the basis of Open Text Livelink. Levy (Moria, 2009) examined and analysed a wide range of articles on web 2.0, Enterprise 2.0 and KM 2.0 to determine whether using WEB 2.0 concepts and tools can result in better integration of knowledge management in organizations. The author concluded as shown in Table 3.3 that WEB 2.0 principles are very close to knowledge management ones and that most of the web 2.0 principles are part of the traditional KM core concepts.

Table 3.3 Web 2.0 Principles Vs Knowledge Management. Adapted from Moria (2009)

WEB 2.0 Principles	KM Matching Principles
WEB as a platform	Technology as a platform Used to integrate interdependent elements, which include People (culture), Process and Technology.
Services development	Web Services The most common way for sharing knowledge and information.
Active participation of users	Active Participation of users It is a necessity for achieving effective and efficient knowledge sharing.
The service improves automatically, the more it is used	Partly correct in Knowledge Management, although increased participation leads to increased and richer content.
Collective intelligence	Knowledge management is based on the collective knowledge of its users.
Content as the core	Content is one of key elements of Knowledge Management

During this period community oriented KM tools or platforms were developed. Such tools included but were not limited to YouTube in 2005; Twitter in 2006; Facebook in 2004 (Lai *et al.*, 2013); Reddit in 2005; and Wikipedia in 2001. An online tool called Moodle in 2002- for creating websites, supporting forums, wikis, databases, collaboration and team learning; and Mambo- an Open Source application written in PHP code, (Lai *et al.*, 2013; Piraquive, García and Aguilar, 2013) were also developed. By the 2006, web 3.0 also referred to as "Web of

Data” had taken root. It was characterised by intelligent categorisation and storage for search and retrieval, Integration of user-generated content and users’ personal preferences. In other words, some of it key features included but were not limited to knowledge connection, semantic web and iGoogle (Lai *et al.*, 2013). Lai et al. (2013) argued that the problem of knowledge sharing, particularly that of tacit knowledge sharing can be resolved through the application of Web 3.0 and inked data in enterprise. Furthermore, another technique proposed by David Gurteen (https://knowledge.cafe/) is called *Knowledge Café* started in 2006. *The Knowledge Café is a conversational process that brings a group of people together to share experiences, learn from each other, build relationships and make a better sense of a rapidly changing, complex, less predictable world to improve decision-making, innovation and the ways in which we work together* (Gurteen, 2015, p. 1). Furthermore, Lefika (2015) in a study of Knowledge cafés outline other techniques developed and practiced during this period. The techniques include the following:

- World cafés: Focuses on the cultivating conversations to trigger knowledge transfer and learning process;
- Technology cafés;- Lays emphasis on discussion of the implementation and use of new technologies;
- Open space technology: Large groups and breakout sessions discussing topics of interest;
- Dialogue meeting: Attention of this technique is on the presentation of questions and communal interpretation and perception; and
- Action learning groups: Groups of individuals with varied levels of expertise and experience in an organisation congregate to analyse problems and develop action plans.

In addition to the development of knowledge sharing techniques, another area of focus in the last two decades was the development of knowledge management strategies and information communication technologies. These included developing KM taxonomies and KM metrics, Big Data, Innovation and KM. de Pablos and Lytas (2018), following a detailed study of 7628 unique research articles from the Web of Science from 1974 to 2017 concluded that a new Data–Knowledge–Wisdom ecosystem will emerge. They suggested that the integration of Big Data, Knowledge Management, and Innovation would be policy driven at a higher level of abstraction, where socially inclusive economic growth and sustainability will become top priorities. Furthermore, KM is developing new forms of technology and communication patterns (Leonardi, 2014; Dalkir, 2017; Archer-Brown and Kietzmann, 2018) that are different from Intranets, databases and social networks. Some of these include social media tools like Twitter, LinkedIn, and Facebook that facilitate the creation and maintenance of large social networks communities(Pasquale *et al.*, 2020). Hemsley and Mason (2013) concluded that social media platforms interacted to produce a new knowledge ecosystem because the these social network communities tend to be developed around diverse collections of content. They concluded that social media could change how organisations perceive knowledge, Knowledge Management, and Knowledge Management Systems.

This period also saw the emergence of the Wikipedia and the Ubuntu knowledge management philosophy all based on organisational culture, behaviours and the concept of social network community. The Ubuntu school of thought rooted in Bantu Africa promotes the interdependency and interconnectedness between people living in the identical communities (Nansubuga and C. Munene, 2019). This philosophy is guides the developments of the Ubuntu open source software operating system that runs on desktop, cloud and internet connected tools. The breadth and depth of Wikipedia and its accessibility and reach, has made it one of the most popular collaborative knowledge repositories (L.G., 2018) in the world.

Despite these developments in Knowledge Management, it remained without a universally accepted definition because it seemed to mean different things to different professionals, academics and organisations. These actors characterised Knowledge Management with respect to their organisational issues, requirements and functional realities. Because these divergent views are individually acceptable and collectively in disagreement, [Anonymous] called this phenomenon an “intellectual paradox”. In proposing a way forward, the authors depicted the fourth view of Knowledge Management (Figure 3.12) below: - the “Conditional View” of Knowledge Management.

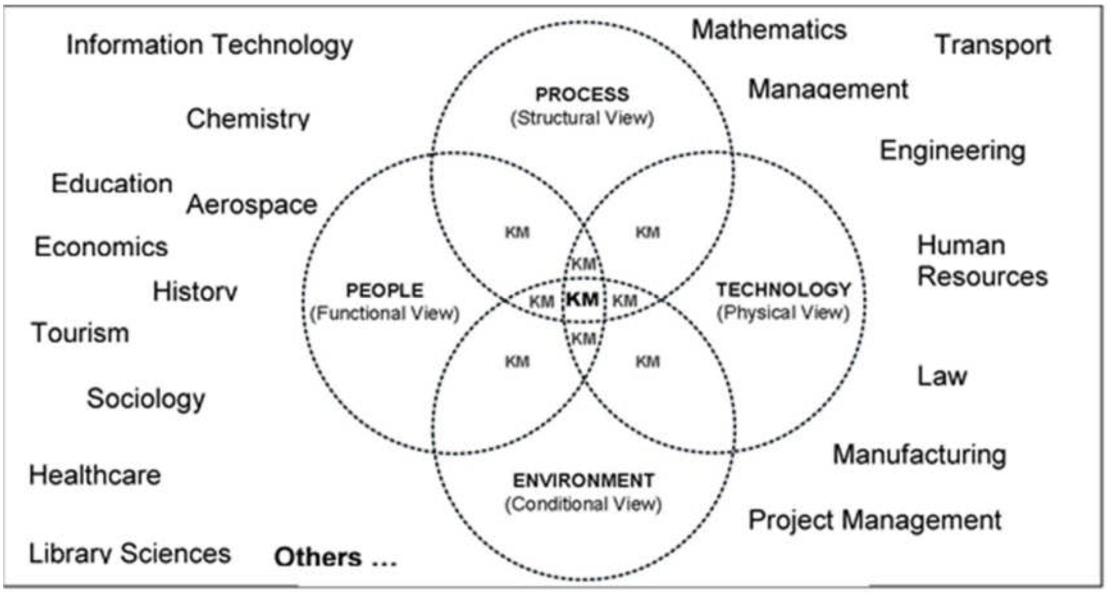


Figure 3-12: A Conditional View of Knowledge Management (Maindze, Jennions and Skaf, 2017)

The Conditional View of knowledge management is shaped by the work environment. The work environment determines the quality, successes, the efficiency and effectiveness of knowledge management techniques, tools and overall implementations. The environmental characteristics that create this conditionality include leadership, employee health and wellbeing, remuneration, organizational culture (Intezari, Taskin and Pauleen, 2017), quality of work tools and space. These and charismatic leadership and contingent reward activities have more impact on all facets of KM practices (Politis, no date; Crawford, 2005; Kumar, 2008; Hai and Sherif, 2011; Donate and Sánchez de Pablo, 2015; Shamim, Cang and Yu, 2019).

3.9 Conclusion

This chapter has discussed key missing bits on the history of Knowledge Management. It highlights the role of information technology as an inseparable part of Knowledge Management. The chapter attempts to demonstrate how potential external events have influenced the narration of the origins of Knowledge Management and its philosophical challenges. At the time of the Cold War, Knowledge Management was a “classified” state secret. This therefore suggests why there has discrepancies in the accounts on the origins of Knowledge Management as a discipline. It is not until fairly recently that the USA’s Department of Defence declassified its reports dating back to the late 1960s and early 1970s robust Knowledge Management programmes and practices.

Furthermore, noteworthy is the fact that during the first two decades (1969-1989) of Knowledge Management as a discipline, ‘snail mail’ was the dominant mode of information dissemination. Therefore, these ideas could not have been communicated at a pace that would gain global attention. As a result, key players like Kellogg, Holsapple and other early knowledge management academics and practitioners missed out in the attribution of credits when the creation or foundations of Knowledge Management are discussed. A typical example of this situation is Berry and Cook who provided the trilogy- “People”, “Process” and “Technology” that is found in almost all Knowledge Management literature but are rarely credited for it. This chapter has shown how soon after Nonaka’s publication of 1995, there was a marked surge in Knowledge Management literature. The trends in the development of technology such as the worldwide web seem to have played a significant part in asserting the origins or foundations of Knowledge Management to the 1990s. This chapter also highlights the bond that exist between KM and Intellectual Capital.

Finally, “The Knowledge Creating Company” came at the right time and brought about the rebirth of Knowledge Management. This justifies the attribution of the creation or foundations of Knowledge Management strongly in favour of its authors and the time of publication.

3.10 References

Aguirre, J.-L., Brena, R. and Cantu, F. J. (2001) ‘Multiagent-Based Knowledge Networks’, *Expert Systems with applications*. Elsevier, 20(1), pp. 65–75.

Akscyn, R. (1984) ‘Advanced Distributed Operating Environment for Document Development and Management.’, in M., A. V. (ed.) *Conference Record - PCC ’84, The Practical Aspects of Engineering Communication*. New York, NY, USA: IEEE.

Alexeis, G.-P. *et al.* (2020) ‘Knowledge Management and Intellectual Capital in Knowledge-Based Organisations: A Review and Theoretical Perspectives’, *Journal of Knowledge Management*. Emerald Publishing Limited, 24(7), pp. 1719–1754. doi: 10.1108/JKM-12-2019-0703.

Applegate, L. M. *et al.* (1987) ‘Knowledge Management in Organizational Planning’, in *Proceedings of the Hawaii International Conference on System Science*. North Hollywood, USA: Western Periodicals Co, pp. 487–496.

Argyris, C. (1976) ‘Single-Loop and Double-Loop Models in Research on Decision Making’, *Administrative Science Quarterly*, 21(3), pp. 363–375. doi: 10.2307/2391848.

Athanassiades, J. C. (1978) ‘Management of the Knowledge Revolution.’ Educational Resources Information Center (ERIC) - ED212265, p. 18.

Bakardjieva, M. and Gaden, G. (2012) ‘Web 2.0 Technologies of the Self’, *Philosophy & Technology*. Springer, 25(3), pp. 399–413.

Becerra-Fernandez, I. and Sabherwal, R. (2014) *Knowledge management: Systems and processes*. Routledge.

Beckman, T. J. (1999) *The Current State of*, *Knowledge management handbook*. CRC press.

Berners-Lee, T. and Fischetti, M. (2001) *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. DIANE Publishing Company.

Berry, J. F. and Cook, C. M. (1976) ‘Managing Knowledge as a Corporate Resource’. Washington DC: Department of Defense, p. 58. doi: ADA029891.

Bontis, N. (1998) ‘Intellectual Capital: An Exploratory Study That Develops Measures and Models’, *Management Decision*. MCB UP Ltd, 36(2), pp. 63–76. doi: 10.1108/00251749810204142.

Borko, H. (1983) ‘Information and Knowledge Worker Productivity’, *Information Processing & Management*, 19(4), pp. 203–212. doi: https://doi.org/10.1016/0306-4573(83)90017-1.

Cabrilo, S. and Dahms, S. (2018) ‘How Strategic Knowledge Management Drives Intellectual Capital to Superior Innovation and Market Performance’, *Journal of Knowledge Management*. Emerald Group Publishing Limited, 22(3), pp. 621–

648. doi: http://dx.doi.org/10.1108/JKM-07-2017-0309.

Campbell-Kelly, M. (2018) *Computer, Student Economy Edition: A History of the Information Machine*. Routledge.

Archer-Brown, C. and Kietzmann, K. (2018) ‘Strategic Knowledge Management and Enterprise Social Media’, *Journal of Knowledge Management*. Emerald Publishing Limited, 22(6), pp. 1288–1309. doi: 10.1108/JKM-08-2017-0359.

Collison, C. and Parcell, G. (2004) *Learning to Fly: Practical knowledge Management from Leading and Learning Organisations*. Capstone Ltd.

Craven, A. (no date) *An interview with Karl-Erik Sveiby - Interview*. Emerald Group Publishing Limited. Available at: http://first.emeraldinsight.com/interviews/pdf/sveiby.pdf?PHPSESSID=1423baeb156c88436a5b11 (Accessed: 20 March 2019).

Crawford, C. B. (2005) ‘Effects of Transformational Leadership and Organizational Position on Knowledge Management’, *Journal of Knowledge Management*. Emerald Group Publishing Limited, 9(6), pp. 6–16. doi: 10.1108/13673270510629927.

Dalkir, K. (2005) *Knowledge Management in Theory and Practice.* Germany: Elsvier Butterworth-Heinemann.

Dalkir, K. (2013) *Knowledge Management in Theory and Practice*. Routledge.

Dalkir, K. (2017) *Knowledge Management in Theory and Practice*. MIT press.

Dalkir, K. and Liebowitz, J. (2011) *Knowledge Management in Theory and Practice*. Massachusetts: MIT press.

Davenport, T. H., Long, D. W. De and Beers, M. C. (1998) ‘Successful Knowledge Management Projects’, *Sloan Management Review*, 39(2), pp. 43–57.

Day, R. E. (2001) ‘Totality and Representation: A History of knowledge management Through European Documentation, Critical Modernity, and Post-Fordism’, *Journal of the American Society for Information Science and Technology*. John Wiley & Sons, Ltd, 52(9), pp. 725–735. doi: https://doi.org/10.1002/asi.1125.

Dewey, J. (1929) ‘The Quest For Certainty: A Study of the Relation of Knowledge and Action’, in Jo Ann Boydston (ed.) *The Later Works of John Dewey, Volume 4, 1925 - 1953*. Volume 4. Southern Illinois University.

Dolk, D. R. and Konsynski, B. R. (1984) ‘Knowledge Representation for Model Management Systems’, *IEEE Transactions on Software Engineering*, 10(6), pp. 619–628.

Donate, M. J. and Sánchez de Pablo, J. D. (2015) ‘The Role of Knowledge-Oriented Leadership in Knowledge Management Practices and Innovation’, *Journal of Business Research*, 68(2), pp. 360–370. doi: https://doi.org/10.1016/j.jbusres.2014.06.022.

Drucker, P. F. (1995) ‘The Age of Social Transformation’, *The Atlantic Monthly*. Centre canadien de gestion, 274(5), pp. 53–80.

DTIC, D. T. I. C. (2021) *ADA029891*, *Accession Number: ADA029891*. Available at: https://apps.dtic.mil/sti/citations/ADA029891 (Accessed: 20 May 2021).

Dziuban, C., Moskal, P. and Hartman, J. (2005) ‘Higher Education, Blended Learning, and the Generations: Knowledge is Power: No More’, *Elements of quality online education: Engaging communities. Needham, MA: Sloan Center for Online Education*, pp. 88–89.

Earl, M. (2001) ‘Knowledge Management Strategies: Toward a Taxonomy’, *Journal of Management Information Systems*. Routledge, 18(1), pp. 215–233. doi: 10.1080/07421222.2001.11045670.

Easterby-Smith, M. and Lyles, M. A. (2011) *Handbook of organizational learning and knowledge management*. John Wiley & Sons.

Edwards, John S *et al.* (2009) ‘Knowledge Management: A Review of the Field and of OR’s Contribution’, *Journal of the Operational Research Society*. Taylor & Francis, 60(sup1), pp. S114–S125.

Edwards, J S *et al.* (2009) ‘Knowledge Management: A Review of The Field and of OR’s Contribution’, *The Journal of the Operational Research Society*. Abingdon: Taylor & Francis Ltd., 60(S1), pp. S114–S125. doi: http://dx.doi.org/10.1057/jors.2008.168.

Edwards, J. S. (2015) ‘Business Processes and Knowledge Management’, in *Encyclopedia of Information Science and Technology*. Third Edit. IGI Global, pp. 4491–4498.

Eidell, T. L. and Kitchel, J. M. (1967) *Knowledge Production and Utilization in Educational Administration*. Edited by T. L. Eidell and J. M. Kitchel. University Council For Educational Administration And Center For The Advanced Study Of Educational Administration.

Filteau, M. C., Kassicieh, S. K. and Tripp, R. S. (1988) ‘Evolutionary Database Design and Development in Very large Scale MIS’, *Information and Management*, 15(4), pp. 203–2012.

Fink, A. Z. Z. and D. (2003) ‘Knowledge Management and Intellectual Capital: An Empirical Examination of Current Practice in Australia’, *Knowledge Management Research & Practice*. Abingdon: Taylor & Francis Ltd., 1(2), pp. 86–94.

Frankland, Amjad, U. and Nolas, S.-M. (2005) ‘Knowledge Management Processes’, in *Encyclopedia of Knowledge Management*. Schwartz,. IGI Global.

Frediani, S. and Saitta, L. (1987) ‘Knowledge Base Organization in Expert Systems’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 1986; Paris;*, pp. 217–224. doi: 10.1007/3-540-

18579-8_21.

Galbraith, J. K. (1967) *The New Industrial State*. Princeton University Press.

Gandhi, S. (2004) ‘Knowledge Management and Reference Services’, *The Journal of Academic Librarianship*, 30(5), pp. 368–381. doi: https://doi.org/10.1016/j.acalib.2004.06.003.

Girard, John and Girard, JoAnn (2015) ‘Defining knowledge management: Toward an applied compendium’, *Online Journal of Applied Knowledge Management*, 3(1), p. 20.

Gold, A. H., Malhotra, A. and Segars, A. H. (2001) ‘Knowledge Management: An Organizational Capabilities Perspective’, *Journal of Management Information Systems*. Routledge, 18(1), pp. 185–214. doi: 10.1080/07421222.2001.11045669.

Grant, K. (2003) ‘Knowledge Management, An Enduring but Confusing Fashion’, *The Electronic. Journal of Knowledge Management*, 9(2), pp. 117–131.

Grant, K. (2011) ‘Knowledge Management, an Enduring but Confusing Fashion’, *Electronic Journal of Knowledge Management*. Academic Conferences International Limited, 9(2), p. 117.

Grossman, M. (2007) ‘The Emerging Discipline of Knowledge Management’, *International Journal of Knowledge Management (IJKM)*, 18(1), pp. 31–38.

Gupta, J. D., Sharma, S. K. and Hsu, J. (2004) ‘An Overview of Knowledge Management’, in *Creating knowledge based organizations*. Igi Global, pp. 1–28.

Gurteen, D. (2015) ‘Knowledge cafe’, *Inside Knowledge*, 13(3), pp. 8–13.

Hai, N. N. and Sherif, M. (2011) ‘Leadership Behaviors, Organizational Culture and Knowledge Management Practices: An Empirical Investigation’, *Journal of Management Development*. Emerald Group Publishing Limited, 30(2), pp. 206–221. doi: 10.1108/02621711111105786.

Handzic, M. and Durmic, N. (2015) ‘Knowledge Management, Intellectual Capital and Project Management: Connecting the Dots’, *Electronic Journal of Knowledge Management*. Reading: Academic Conferences International Limited, 13(1), pp. 51–61.

Hannabuss, S. (1987) ‘Knowledge Management’, *Library Management*, 8(5), pp. 2–50.

Harrison, T. M. and Barthel, B. (2009) ‘Wielding New Media in Web 2.0: Exploring the History of Engagement with the Collaborative Construction of Media Products’, *New Media & Society*. SAGE Publications, 11(1–2), pp. 155–178. doi: 10.1177/1461444808099580.

Hemsley, J. and Mason, R. M. (2013) ‘Knowledge and Knowledge Management in the Social Media Age’, *Journal of Organizational Computing and Electronic Commerce*. Taylor & Francis, 23(1–2), pp. 138–167. doi:

10.1080/10919392.2013.748614.

Hicks, D. (1995) ‘Published Papers, Tacit Competencies and Corporate Management of the Public/Private Character of Knowledge’, *Industrial and corporate change*. Oxford University Press, 4(2), pp. 401–424.

Hislop, Donald *et al.* (2018) ‘Knowledge Management’, *(Potential) Future Research Directions*. Edited by J Syed et al. Springer International Publishing AG (The Palgrave Handbook of Knowledge Management), pp. 691-703 BT-The Palgrave Handbook of Knowledge Ma. doi: 10.1007/978-3-319-71434-9_28.

Holsapple, C.W. (1987) ‘Adapting Demons to Knowledge Management Environments’, *Decision Support Systems*, 3(4), pp. 289–298. doi: 10.1016/0167-9236(87)90100-X.

Holsapple, C W (1987) ‘Knowledge Base Organization in Expert Systems’, in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*.

Holsapple, C. W. (2005) ‘The Inseparability of Modern Knowledge Management and Computer-Based Technology’, *Journal of Knowledge Management*. Emerald, 9(1), pp. 42–52. doi: 10.1108/13673270510582956.

Holsapple, C. W. and Whinston, A. B. (1989) ‘Knowledge Representation and Processing in Economics and Management’, *Computer Science in Economics and Management*, 2(1), pp. 37–48. doi: 10.1007/BF00454703.

Hsu, C. and Perry, A. (1988) ‘Knowledge Representation Using the Two-Stage Entity-Relationship Approachi A Case Study in Manufacturing Information Management’, in *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 330–335. doi: 10.1109/ICSMC.1988.754307.

Hsu, C. and Skevington, C. (1987) ‘Integration of Data and Knowledge in Manufacturing Enterprises: A Conceptual Framework’, *Journal of Manufacturing Systems*. Elsevier, 6(4), pp. 277–285.

Hsu, I.-C. and Sabherwal, R. (2012) ‘Relationship Between Intellectual Capital and Knowledge Management: An Empirical Investigation’, *Decision Sciences*. Atlanta: American Institute for Decision Sciences, 43(3), p. 489.

Huang, Y.-M., Rozenblit, J. W. and Hu, J. (1989) ‘Role of knowledge Management in Hierarchical Model Development’, in *Winter Simulation Conference Proceedings*. Washington: Publ by IEEE, pp. 584–592.

Hussi, T. (2004) ‘Reconfiguring Knowledge Management - Combining Intellectual Capital, Intangible Assets and Knowledge Creation’, *Journal of Knowledge Management*. Kempston: Emerald Group Publishing Limited, 8(2), pp. 36–52. doi: http://dx.doi.org/10.1108/13673270410529091.

Intezari, A., Taskin, N. and Pauleen, D. J. (2017) ‘Looking Beyond Knowledge Sharing: An Integrative Approach to Knowledge Management Culture.’, *J.*

Knowledge Management, 21(2), pp. 492–515.

Ismail, M. B. and Yusof, Z. M. (2010) ‘The Contribution of Technological Factors on Knowledge Sharing Quality Among Government Officers in Malaysia’, *Knowledge management*. InTech Europe Rijecka, Croatia, pp. 239–254.

Ives, W., Torrey, B. and Gordon, C. (1997) ‘Knowledge Management: An Emerging Discipline with a Long History’, *Journal of Knowledge Management*. MCB UP Ltd, 1(4), pp. 269–274. doi: 10.1108/EUM0000000004598.

Jasimuddin, S. M. (2006) ‘Disciplinary Roots of Knowledge Management: A Theoretical Review’, *International Journal of Organizational Analysis*. Emerald, 14(2), pp. 171–180. doi: http://dx.doi.org/10.1108/13673279910275611.

Javier, D. M. and Fátima, G. (2011) ‘Organizational Factors to Support Knowledge Management and Innovation’, *Journal of Knowledge Management*. Edited by G. M. Castro, P. López-Sáez, and M. Delgado-Verde. Emerald Group Publishing Limited, 15(6), pp. 890–914. doi: 10.1108/13673271111179271.

Johnson, D. B. and Maltz, D. A. (1996) ‘Dynamic Source Routing in Ad Hoc Wireless Networks’, *Mobile computing*. Tomasz Imi. Springer, pp. 153–181.

Kellog, C. (1960) ‘The Fact Compiler: A System for the Extraction, Storage, and Retrieval of Information’, in *Proceeding IRE-AIEE-ACM ’60 (Western) Papers presented at the May 3-5, 1960, western joint IRE-AIEE-ACM computer conference*. San Francisco, California — May 03 - 05, 1960: IRE-AIEE-ACM, pp. 73–82. doi: 10.1145/1460361.1460370.

Kellog, C. (1982) ‘Knowledge Management: A Practical Amalgam of Knowledge and Data Base Technology’, in *Proceedings of the Second AAAI National Conference on Artificial Intelligence*. Pittsburgh, PA, USA: AAAI Press, p. Pages 306-309.

Kellog, C. (1983) ‘Intelligent Assistants for Knowledge and Information Resources Management’, in *Proceeding IJCAI’83 Proceedings of the Eighth international joint conference on Artificial intelligence - Volume 1 Karlsruhe, West Germany — August 08 - 12*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1983, pp. 170–172.

Kellogg, C. (1986) ‘From Data Management to Knowledge Management’, *Computer*, 19(1), pp. 75–84.

Khasseh, A. A. and Mokhtarpour, R. (2016) ‘Tracing the Historical Origins of Knowledge Management Issues Through Referenced publication years spectroscopy (RPYS)’, *Journal of Knowledge Management*. Emerald Group Publishing Limited, 20(6), pp. 1393–1404. doi: 10.1108/JKM-01-2016-0019.

Kianto, A. *et al.* (2014) ‘The Interaction of Intellectual Capital Assets and Knowledge Management Practices in Organizational Value Creation’, *Journal of Intellectual Capital*. Emerald Group Publishing Limited, 15(3), pp. 362–375. doi: http://dx.doi.org/10.1108/JIC-05-2014-0059.

Koenig, M. E. D. (1997) ‘Intellectual Capital and How to Leverage It’, *The Bottom Line*. MCB UP Ltd, 10(3), pp. 112–118. doi: 10.1108/08880459710175368.

Koenig, M. and Neveroski, K. (2008) ‘The Origins and Development of Knowledge Management’, *Journal of Information & Knowledge Management*. World Scientific Publishing Co., 07(04), pp. 243–254. doi: 10.1142/S0219649208002111.

Kogan, D. (1984) ‘The Manager’s Assistant: An Application of Knowledge Management’, in *1984 IEEE First International Conference on Data Engineering*, pp. 592–595. doi: 10.1109/ICDE.1984.7271323.

Krogh, G. Von (2012) ‘Knowledge Sharing in Organizations: The Role of Communities’, in *Handbook of Organizational Learning and Knowledge Management*. (Wiley Online Books), pp. 403–432. doi: https://doi.org/10.1002/9781119207245.ch19.

Kumar, B. R. (2019) ‘Microsoft’s Acquisition of LinkedIn’, in Kumar, B. R. (ed.) *Wealth Creation in the World’s Largest Mergers and Acquisitions: Integrated Case Studies*. Cham: Springer International Publishing, pp. 235–241. doi: 10.1007/978-3-030-02363-8_27.

Kumar, S. S. (2008) ‘Role of Leadership in Knowledge Management: A Study’, *Journal of Knowledge Management*. Emerald Group Publishing Limited, 12(4), pp. 3–15. doi: 10.1108/13673270810884219.

L.G., P. (2018) ‘Community’s Knowledge Need and Knowledge Sharing in Wikipedia’, *Journal of Knowledge Management*. Emerald Publishing Limited, 22(4), pp. 912–930. doi: 10.1108/JKM-09-2017-0412.

Lai, H. *et al.* (2013) ‘The Framework of Web 3.0-Based Enterprise Knowledge Management System’, in Uden, L. et al. (eds) *7th International Conference on Knowledge Management in Organizations: Service and Cloud Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 345–351.

Lambe, P. (2011) ‘The Unacknowledged Parentage of Knowledge Management’, *Journal of Knowledge Management*, 15(2), pp. 175–197.

Lausin, A., Desouza, K. C. and Kraft, G. D. (2003) ‘Knowledge Management in The US Army’, *Knowledge and Process Management*. Wiley Online Library, 10(4), pp. 218–230.

Lefika, P. T. and Mearns, M. A. (2015) ‘Adding Knowledge Cafés to the Repertoire of Knowledge Sharing Techniques’, *International Journal of Information Management*, 35(1), pp. 26–32. doi: https://doi.org/10.1016/j.ijinfomgt.2014.09.005.

Leonardi, P. M. (2014) ‘Social Media, Knowledge Sharing, and Innovation: Toward a Theory of Communication Visibility’, *Information Systems Research*. INFORMS, 25(4), pp. 796–816. doi: 10.1287/isre.2014.0536.

Lewis, J. W. and Lynch, F. S. (1983) ‘Getree: A Knowledge Management Tool.’, in *Proceedings - Trends & Applications (IEEE Computer Society)*. New York, NY, USA: IEEE, pp. 273–277.

Liebowitz, J. (2005) ‘Introduction to Knowledge Management’, in Dalkir, K. (ed.) *Knowledge Management in Theory and Practice*. MIT Press, pp. 1–30.

Liebowitz, J. (2010) ‘The Quick Basics of Knowledge Management’, in Liebowitz;, J., Schieber;, R. A., and Andreadis, J. D. (eds) *Knowledge Management in Public Health*. London: CRC Press: Taylor & Francis, pp. 3–19.

Liebowitz, J. and Suen, C. Y. (2000) ‘Developing Knowledge Management Metrics for Measuring Intellectual Capital’, *Journal of Intellectual Capital*. Bradford: Emerald Group Publishing Limited, 1(1), pp. 54–67. doi: http://dx.doi.org/10.1108/14691930010324160.

Liza, B. *et al.* (2018) ‘Communities of practice to improve public health outcomes: a systematic review’, *Journal of Knowledge Management*. Emerald Publishing Limited, 22(2), pp. 326–343. doi: 10.1108/JKM-03-2017-0111.

Van Lohuizen, C. W. W. (1986) ‘Knowledge Management and Policy Making’, *Science Communication*, 8(1), pp. 12–38.

De Long, D. W. and Fahey, L. (2000) ‘Diagnosing Cultural Barriers to Knowledge Management’, *Academy of Management Perspectives*. Academy of Management, 14(4), pp. 113–127. doi: 10.5465/ame.2000.3979820.

Machlup, F. (1962) *The Production and Pistribution of Knowledge in The United States*. Princeton: Princeton University Press.

Maier, R. (2007) *Knowledge Management Systems: Information and Communication Technologies for Knowledge Management*. Third. Berlin, Heidelberg: Springer Berlin Heidelberg.

Maindze, A., Jennions, I. and Skaf, Z. (2017) ‘Knowledge Management Yesterday and Tomorrow: Exploring an “Intellectual Paradox”’, in James Gao, Mohammed El Souri, S. K. (ed.) *Advances in Transdisciplinary Engineering*. Amsterdam: IOS Press, pp. 522–527. doi: 10.3233/978-1-61499-792-4-522.

Marr, B. *et al.* (2003) ‘Intellectual Capital and Knowledge Management Effectiveness’, *Management Decision*. London: Emerald Group Publishing Limited, 41(8), pp. 771–781. doi: http://dx.doi.org/10.1108/00251740310496288.

Marschak, J. (1970) *Producing, Storing, Transporting, and Using Knowledge.* Presented at the World Econometric Congress (2nd), Cambridge (England) 10 Sep 70.

Mårtensson, M. (2000) ‘A Critical Review of Knowledge Management as a Management Tool’, *Journal of Knowledge Management*. Kempston: Emerald Group Publishing Limited, 4(3), pp. 204–216. doi: 10.1108/13673270010350002.

McInerney, C. R. and Koenig, M. E. D. (2011) *Knowledge Management (KM) Processes in Organizations: Theoretical Foundations and Practice*, *Synthesis Lectures on Information Concepts, Retrieval, and Services*. Morgan & Claypool

Publishers.

McMahon, C., Lowe, A. and Culley, S. (2004) ‘Knowledge Management in Engineering Design: Personalization and Codification’, *Journal of Engineering Design*. Taylor & Francis, 15(4), pp. 307–325. doi: 10.1080/09544820410001697154.

Mertins, K., Heisig, P. and Vorbeck, J. (2003) *Knowledge Management: Concepts and Best Practices*. Second. Edited by K. Mertins, P. Heisig, and J. Vorbeck. Springer Science & Business Media.

Meso, P. and Smith, R. (2000) ‘A Resource-Based View of Organizational Knowledge Management Systems.’, *Journal of Knowledge Management*. Kempston: Emerald Group Publishing Limited, 4(3), p. 224. doi: http://dx.doi.org/10.1108/13673270010350020.

Metaxiotis, K., Ergazakis, K. and Psarras, J. (2005) ‘Exploring the World of Knowledge Management: Agreements and Disagreements in The Academic/Practitioner Community’, *Journal of Knowledge Management*, 9(2), pp. 6–18. doi: 10.1108/13673270510590182.

Mills, A. M. and Smith, T. A. (2011) ‘Knowledge Management and Organizational Performance: A Decomposed View’, *Journal of Knowledge Management*. Kempston: Emerald Group Publishing Limited, 15(1), pp. 156–171. doi: http://dx.doi.org/10.1108/13673271111108756.

Moria, L. (2009) ‘WEB 2.0 Implications on Knowledge Management’, *Journal of Knowledge Management*. Emerald Group Publishing Limited, 13(1), pp. 120–134. doi: 10.1108/13673270910931215.

Nansubuga, F. and C. Munene, J. (2019) ‘Awakening the Ubuntu Episteme to Embrace Knowledge Management in Africa’, *Journal of Knowledge Management*. Emerald Publishing Limited, 24(1), pp. 105–119. doi: 10.1108/JKM-09-2018-0603.

Naveed, K., Watanabe, C. and Neittaanmäki, P. (2018) ‘The Transformative Direction of Innovation Toward an IoT-Based Society - Increasing Dependency on Uncaptured GDP in Global ICT Firms.’, in *Technology in Society*, pp. 23–46. doi: https://doi.org/10.1016/j.techsoc.2017.11.003.

Nazim, M. and Mukherjee, B. (2016) ‘Chapter 2 - History and Evolution of Knowledge Management Systems’, in Rikowski, R. (ed.) *Knowledge Management in Libraries Concepts, Tools and Approaches*. Chandos Publishing, pp. 27–48. doi: https://doi.org/10.1016/B978-0-08-100564-4.00002-8.

Nicholas, L. H. (1974) ‘Knowledge Management: A New Concern for Public Administration’, in *Public Administration Review*. Wiley American Society for Public Administration, pp. 189–198. doi: 10.2307/974902.

Nielsen, C. (2018) ‘Relating Successful Business Models to Intellectual Capital and Knowledge Management Practices’, *Electronic Journal of Knowledge Management*. Academic Conferences International Limited, 16(1), pp. 48–55.

Nonaka, I. (1994) ‘A Dynamic Theory of Organizational Knowledge Creation’, *Organization Science*, 5(1), pp. 14–37. doi: 10.1287/orsc.5.1.14.

Nonaka, I., Von Krogh, G. and Voelpel, S. (2006) ‘Organizational Knowledge Creation Theory: Evolutionary Paths and Future Advances’, pp. 1179–1208. doi: 10.1177/0170840606066312.

Nonaka, I. and Takeuchi, H. (1995) *The knowledge-Creating Company : How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.

Nunamaker, J. F., Romano, N. C. and Briggs, R. O. (2001) ‘A Framework for Collaboration and Knowledge Management’, in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE, pp. 1–12.

O’Reilly, T. (2009) *What is Web 2.0*. ‘ O’Reilly Media, Inc.’

Oppitz, M. and Tomsu, P. (2018) ‘Creating Innovation’, in Oppitz, M. and Tomsu, P. (eds) *Inventing the Cloud Century: How Cloudiness Keeps Changing Our Life, Economy and Technology*. Cham: Springer International Publishing, pp. 349–375. doi: 10.1007/978-3-319-61161-7_13.

de Pablos, P. O. and Lytras, M. (2018) ‘Knowledge Management, Innovation and Big Data: Implications for Sustainability, Policy Making and Competitiveness’, *Sustainability (Switzerland)*, 10(6). doi: 10.3390/su10062073.

Paoloni, M. *et al.* (2020) ‘Knowledge Management, Intellectual Capital and Entrepreneurship: A Structured Literature Review’, *Journal of Knowledge Management*. Kempston: Emerald Group Publishing Limited, 24(8), pp. 1797–1818. doi: http://dx.doi.org/10.1108/JKM-01-2020-0052.

Paradice, D. B. and Courtney, J. F. (1989) ‘Organizational Knowledge Management’, *Information Resources Management Journal (IRMJ)*, 2(3), pp. 1–14.

Parsons, M. (2004) *Effective Knowledge Management for Law Firms*. Oxford University Press.

Pasquale, D. V. *et al.* (2020) ‘Detecting Customers Knowledge from Social Media Big Data: Toward an Integrated Methodological Framework Based on Netnography and Business Analytics’, *Journal of Knowledge Management*. Emerald Publishing Limited, 24(4), pp. 799–821. doi: 10.1108/JKM-11-2019-0637.

Peter, H. (2009) ‘Harmonisation of Knowledge Management – Comparing 160 KM Frameworks Around the Globe’, *Journal of Knowledge Management*. Emerald Group Publishing Limited, 13(4), pp. 4–31. doi: 10.1108/13673270910971798.

Piraquive, F. N. D., García, V. H. M. and Aguilar, L. J. (2013) ‘Technological Tools Virtual Collaborative to Support Knowledge Management in Project Management’, in Uden, L. et al. (eds) *7th International Conference on Knowledge*

Management in Organizations: Service and Cloud Computing. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 163–174.

Polanyi, M. (1967) *The Tacit Dimension.* New York: Anchor.

Politis, J. D. (no date) ‘The Relationship of Various Leadership Styles to Knowledge Management’, *Leadership and Organization Development Journal*, 22(8), pp. 354–364.

Prusak, L. (2001) ‘Where Did Knowledge Management Come From?’, *IBM systems Journal*, 40(4), pp. 1002–1007.

Quintas, P. (2005) ‘The Nature and Dimensions of Knowledge Management’, in Anumba;, C. J., Egbu;, C. O., and Carrillo, P. M. (eds) *Knowledge Management in Construction*. Blackwell Publishing Ltd. Oxford, UK, p. 30.

Ravi, A. (2002) ‘Implementing KM – A Balanced Score card Approach’, *Journal of Knowledge Management*. MCB UP Ltd, 6(3), pp. 240–249. doi: 10.1108/13673270210434340.

Razmerita, L., Phillips-Wren, G. and Jain, Lakhmi C. (2016) ‘Advances in Knowledge Management: An Overview’, in Razmerita, L., Phillips-Wren, G., and Jain, L. C. (eds) *Innovations in Knowledge Management: The Impact of Social Media, Semantic Web and Cloud Computing.* 95th edn. Berlin, Heidelberg: Springer, pp. 3–18. doi: DOI 10.1007/978-3-662-47827-1.

Razmerita, L., Phillips-Wren, G. and Jain, Lakhmi C (2016) ‘Advances in Knowledge Management: An Overview’, in Razmerita, L., Phillips-Wren, G., and Jain, L. C. (eds) *Innovations in Knowledge Management: The Impact of Social Media, Semantic Web and Cloud Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3–18. doi: 10.1007/978-3-662-47827-1_1.

Rennen, W. and Martens, P. (2003) ‘The Globalisation Timeline’, *Integrated Assessment*, 4(3), pp. 137–144. doi: 10.1076/iaij.4.3.137.23768.

Rubenstein-Montano, B. *et al.* (2001) ‘A Systems Thinking Framework for Knowledge Management’, *Decision support systems*. Elsevier, 31(1), pp. 5–16.

Sahoo, J., Pati, B. and Mohanty, B. (2017) ‘Knowledge Management as An Academic Discipline: An Assessment’, in *Managing Knowledge and Scholarly Assets in Academic Libraries*. IGI Global, pp. 99–126.

Sandars, J. (2004) ‘Knowledge Management: Information Technology’, *Work Based Learning in Primary Care*, 2(3), pp. 202–207.

Schmoldt, D. L. (1989) ‘Knowledge Management: An Application to Wildfire Prevention Planning’, *Fire Technology*. Kluwer Academic Publishers, 25(2), pp. 150–164.

Schultze, U. (2006) ‘The Career of Knowledge Management’, *Journal of Asian Business*. Ann Arbor: Regents of the University of Michigan, 22/23(2/3,1), p. 213.

Seleim, A. A. S. and Khalil, O. E. M. (2011) ‘Understanding the Knowledge

Management-Intellectual Capital Relationship: A Two-Way Analysis', *Journal of Intellectual Capital*. Bradford: Emerald Group Publishing Limited, 12(4), pp. 586–614. doi: http://dx.doi.org/10.1108/14691931111181742.

Senge, P. (1990) *The Fifth Discipline: The Art & Practice of Learning Organization.* New York: Doupleday Currence.

Serenko, A. *et al.* (2010) 'A Scientometric Analysis of Knowledge Management and Intellectual Capital Academic Literature (1994-2008)', *Journal of Knowledge Management*. Kempston: Emerald Group Publishing Limited, 14(1), pp. 3–23. doi: http://dx.doi.org/10.1108/13673271011015534.

Serenko, A. (2021) 'A Structured Literature Review of Scientometric Research of the Knowledge Management Discipline: A 2021 Update', *Journal of Knowledge Management*, ahead-of-p(ahead-of-print). doi: 10.1108/JKM-09-2020-0730.

Serenko, A. and Bontis, N. (2004) 'Meta-Review of Knowledge Management and Intellectual Capital Literature: Citation Impact and Research Productivity Rankings', *Knowledge and Process Management*. Chichester: Wiley Periodicals Inc., 11(3), pp. 185–198.

Serenko, A. and Bontis, N. (2009) 'Global Ranking of Knowledge Management and Intellectual Capital Academic Journals', *Journal of Knowledge Management*. Kempston: Emerald Group Publishing Limited, 13(1), pp. 4–15. doi: http://dx.doi.org/10.1108/13673270910931125.

Serenko, A. and Bontis, N. (2017) 'Global Ranking of Knowledge Management and Intellectual Capital Academic Journals: 2017 Update', *Journal of Knowledge Management*. Emerald Group Publishing Limited, 21(3), pp. 675–692. doi: http://dx.doi.org/10.1108/JKM-11-2016-0490.

Shamim, S., Cang, S. and Yu, H. (2019) 'Impact of Knowledge Oriented Leadership on Knowledge Management Behaviour Through Employee Work Attitudes', *The International Journal of Human Resource Management*. Routledge, 30(16), pp. 2387–2417. doi: 10.1080/09585192.2017.1323772.

Shen, S. (1987) 'Knowledge Management in Decision Support Systems', *Decision Support Systems*, 3(1), pp. 1–11. doi: https://doi.org/10.1016/0167-9236(87)90031-5.

Shu-Mei, T. (2010) 'The Correlation Between Organizational Culture and Knowledge Conversion on Corporate Performance', *Journal of Knowledge Management*. Emerald Group Publishing Limited, 14(2), pp. 269–284. doi: 10.1108/13673271011032409.

Spender, J. C. (2015) 'Knowledge Management: Origins, History, and Development', in Bolisani, E. and Handzic, M. (eds) *Advances in Knowledge Management: Celebrating Twenty Years of Research and Practice*. Cham: Springer International Publishing, pp. 3–23. doi: 10.1007/978-3-319-09501-1_1.

Spicer, D. (2016) 'Raymond Tomlinson: Email Pioneer, Part 1', *IEEE Annals of the History of Computing*, 38(2), pp. 72–79. doi: 10.1109/MAHC.2016.25.

Stewart, T. A. (1997) *Intellectual Capital: The New Wealth of Organizations*. USA: Doubleday.

Suhwan, J., Young-Gul, K. and Joon, K. (2011) ‘An Integrative Model for Knowledge Sharing in Communities-of-Practice’, *Journal of Knowledge Management*. Emerald Group Publishing Limited, 15(2), pp. 251–269. doi: 10.1108/13673271111119682.

Sveiby, K.-E. and Rislings, A. (1986) *Kunskapsföretaget - Seklets Viktigaste Ledarutmaning?* Liber Förlag.

Sveiby, K. (2001) *Knowledge Management – Lessons from the Pioneers*, *Knowledge Management – Lessons from the Pioneers*. doi: 10.1021/jp209773s.

Sveiby, K. E. (1990) *Knowledge Management: 101 Tips for Leaders in knowledge-Intensive Organizations*. Stockholm: Suecia: Affarsvarlden.

Syed, J. *et al.* (2018) *The Palgrave Handbook of Knowledge Management*. Edited by J. Syed et al. Springer.

Topping, K. and Ehly, S. (1998) *Peer-Assisted learning*. Routledge.

Tuomi, I. (1986) ‘The Future of Knowledge Management’, *Lifelong Learning in Europe*, VII(2), pp. 1–15.

Vanham, P. (2019) *A Brief History of Globalization*, *https://www.weforum.org/*. https://www.weforum.org/.

Wallace, D. P. (2007) *Knowledge Management: Historical and Cross-Disciplinary Themes*. Libraries Unlimited.

Wenger, E., McDermott, R. A. and Snyder, W. (2002) *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business Press.

Wenger, E. and Snyder, W. (2000) ‘Communities of Practice: The Organizational Frontier’, *Harvard Business Review*. Harvard Business Review, pp. 139–145.

Whinston, B. A., Holsapple, W. C. and Bonczek, H. R. (1981) *Foundations of Decision Support Systems*. New York: Academic Press, Inc.

Wiig, K. M. (1986) ‘Management of Knowledge: Perspectives of a New Opportunity.’, in Bernold, T. (ed.) *User interfaces : gateway or bottleneck? : proceedings of the Technology Assessment and Management Conference of the Gottlieb Duttweiler Institute, Rüschlikon/Zürich, Switzerland, 20-21 October*. Elsevier Science Ltd, pp. 101–116.

Wiig, Karl M (1997a) ‘Integrating Intellectual Capital and Knowledge Management’, *Long Range Planning*. London: Elsevier Ltd, 30(3), pp. 399–405.

Wiig, Karl M. (1997) ‘Knowledge Management: An Introduction and Perspective’, *Journal of Knowledge Management*, 1(1), pp. 6–14. doi: 10.1108/13673279710800682.

Wiig, Karl M (1997b) ‘Knowledge Management: Where Did It Come From and

Where Will It Go?’, *Expert Systems With Applications*, 13(I), pp. 1–14.

Wiig, K. M. (1999) ‘What Future Knowledge Management Users May Expect’, *Journal of Knowledge Management*. Emerald, 3(2), pp. 155–166. doi: 10.1108/13673279910275611.

Wiig, K. M. (2000) ‘Chapter 1 - Knowledge Management: An Emerging Discipline Rooted in a Long History’, in Despres, C. and Chauvel, D. B. T.-K. H. (eds) *Knowledge Horizons: The Present and the Promise of Knowledge Management*. Boston: Butterworth-Heinemann, pp. 3–26. doi: https://doi.org/10.1016/B978-0-7506-7247-4.50004-5.

Wilson, T. D. (2002) ‘The Nonsense of “Knowledge Management”’, *Information research*, 8(1), pp. 1–8

CHAPTER 4: SCOPING, SYSTEMATIC REVIEW, DESIGN AND VALIDATION: A METHODOLOGICAL APPROACH.

Abstract: In this paper, we present the methodology for reviewing, designing and validating non-mathematical and non-statistical conceptual models. A systematic review protocol is presented and used as the basis for conducting the review. A multi methodology approach is presented, and the link between standards and the models are discussed.

4.1 Introduction

This research develops a data life cycle model using Integrated Vehicle Health Management (IVHM) engineering as an applied case study. According to Yin (2018, p. 50), "*a case study is an empirical method that investigates a contemporary phenomenon (the "case") in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident.*" The design and application of data life cycle models is certainly a modern phenomenon that is fundamentally theoretical in nature and can have both a theoretical and functional implementation. This practice is varied and universal in both academia and industry, making any meaningful research on this topic most appropriately achieved through a case study. The rationale for this practice-based method is that, furthermore, it will enhance data- and knowledge management in IVHM Engineering and research management. As a result, the outputs from this research will be applied towards the development and enhancement of IVHM data- and knowledge management processes.

Though there was significant use of organic literature review as a result of a scoping study, the systematic literature review approach was adopted for this study as the primary approach. The systematic review method is based on guidelines advocated by (Arksey and O'Malley, 2005; Kitchenham and Charters, 2007; Bown and Sutton, 2010; Brown and Jones-diette, 2014; Adams, Smart and Huff, 2017; Higgins JPT Chandler J, Cumpston M, Li T, Page MJ, Welch VA, 2019) for using systematic reviews in research interventions. It differs from the organic approach in that it deploys a pre-defined, detailed and rigorous protocol for reviewing the subject specific literature. In contrast to organic reviews, systematic reviews aim to offer a near exhaustive list of published and unpublished articles relating to a particular area under investigation. Whereas organic reviews attempt to provide a narrative critical evaluation of results of several studies, systematic reviews use clear and detailed standards to identify, critically evaluate and synthesise almost all of the literature on the topic of research.

This systematic review aims to review the literature on data life cycle management and metadata standards employed and their suitability for IVHM knowledge management requirements. This systematic review seeks to answer the following questions:

- Which data life cycle model is suitable for integrated vehicle health management (IVHM) data and knowledge management?
- How can IVHM data management requirements for medium to long-term data availability be addressed?
- What can be done to guarantee, the security, reliability and quality of IVHM data?

It is hoped that this review will provide guidance for improving data and knowledge management for IVHM research, engineering and prognostic health management.

4.2 Systematic Reviews

This approach is of high significance, especially as it offers a comprehensive summary of the subject area to the present date. A research protocol detailing the activities to be carried out was defined before commencing the review. The protocol is a plan describing how the systematic review shall be conducted. The objective is to make sure that the process is fair and replicable (Kitchenham and Charters, 2007; Staples and Niazi, 2007; Kitchenham and Brereton, 2013; Kitchenham, Budgen and Brereton, 2015; Rutter *et al.*, 2015; Booth, 2016).

A systematic literature review is defined as a " *summary of the research literature that is focused on a single question* (Bettany-Saltikov, 2012, p. 5)"; and "*a form* *of secondary study that uses a well-defined methodology to identify, analyse and interpret all available evidence related to a specific research question in a way that is unbiased and (to a degree) repeatable* (Bettany-Saltikov, 2012)." This approach is based on an ordered, structured, repeatable and thorough protocol that also facilitates the collection of evidence in the field and synthesised in an impartial manner (Kitchenham, Budgen and Brereton, 2015). Pettigrew and Roberts in Bettany (2012) emphasise that systematic reviews *'adhere closely to a set of scientific methods that explicitly aim to limit systematic error (bias), mainly attempting to identify, appraise and synthesise all relevant studies (of whatever design) in order to answer a particular question (or set of questions)'*." This method presents a rigorous and transparent mapping of areas of research (Arksey and O'Malley, 2005). This implies there is a resulting increase in the validity and reliability of finding as a result of adopting this approach. Risk of data misinterpretation is reduced and therefore increasing the credibility of the results. This is strengthened by the repeatability of the methodology. The method increases the uniformity of findings and provides justification for the extension of results. The main downside of this method is that it takes a lot of time – about six to Eight months subject to the quantity of obtainable relevant literature (Ten Ham-Baloyi and Jordan, 2016).

4.3 Research Protocol

As a first step, a scoping study was carried out to map the breadth and depth of data management in knowledge management implementations. This was followed by a research protocol detailing the objective, research question, inclusion and exclusion criteria, search strategy and data synthesis; developed in line with the guidelines defined in Rutter (2015), in Kitchenham (2015) and in Kitchenham and Charters (2007) and also detailed by Ten Ham-Baloyi and Jordan (2016). The preliminary studies yielded the keywords outlined in Table 4.1 below on page 79.

Table 4.1 Keywords Used in the Search

“Algorithm”	“Laboratory data”
“Metadata”	"Experimental data management"
"Data life cycle"	"Data management life cycle model"
"Data life cycle model"	"Managing experimental data"
"Scientific data life cycle"	“Data curation lifecycle model”
“Scientific metadata model”	“Experimental data”
"Data life cycle management"	"Curation lifecycle model"
"Research data management"	"Knowledge Management"
"Research data life cycle"	"Knowledge Management System"
"Data management life cycle"	"Laboratory information management system"
"Data management system"	"Laboratory data management system"

The selected keywords were then used to formulate search strings for Scopus, EBSCO and ABI databases, respectively.

4.4 Eligibility Criteria

4.4.1 Inclusion Criteria

All types of data management life cycle studies are eligible for inclusion regardless of the methodology. Articles must be from 2010 to date. An article must be about the data life cycle or data life cycle management, metadata standards and data management standards. Papers focusing on Knowledge Management Systems and data life cycle and metadata and data management standards were included, including papers not included in Scopus, EBSCO and ABI but cited in the relevant literature on knowledge management and data life cycle management. Only articles with full-text availability were included.

4.4.2 Exclusion Criteria

Articles older than 2010, editorials, blogs, opinions, in-progress articles and lecture notes were excluded. The details about studies excluded and the justification were captured at all stages of the process. Studies are excluded if the focus of a paper is undoubtedly not on data life cycle management, metadata and data management standards. Duplicated studies by the same authors, including articles not written in the English language, were excluded. Semantics and database management papers were not included.

4.5 Search Strategy

The research undertook a search using all identified keywords and index terms across all included databases. The databases used included SCOPUS, EBSCO and ABI/INFORM to ensure a comprehensive coverage of scholarly journals, dissertations and books. This search strategy was supported by the 12 step guidelines provided (Kable, Pich and Maslin-Prothero, 2012) (Box 1)

1. Provide a purpose statement
2. Document the databases or search engines used
3. Specify the limits applied to the search
4. List the inclusion criteria and exclusion criteria
5. List the search terms used
6. Document the search process
7. Assess retrieved articles for relevance
8. Document a summary table of included articles
9. Provide a statement specifying the number of retrieved articles
10. Conduct quality appraisal of retrieved literature
11. Critical review of literature
12. Check the reference list for accuracy

Figure 4-1 Box 1: 12 Step Guidelines by Kable et al. (2012).

SCOPUS, EBSCO and ProQuest ABI/INFORM electronic databases were searched for relevant studies.

Using the keywords outlined in (Table 4.1) the following search string (Box 2) was formulated and used across the three databases.

(algorithm OR Data OR "Experimental data" OR “laboratory Data” OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata)[within five words of] (model*) AND (Knowledge OR information) [within five words of] (manag* OR System*)

Figure 4-2 Box 2: Search String Composed

Furthermore, the reference lists of all the identified relevant reports and studies were searched for additional articles.

In tables 4.2, 4.3 and 4.4 the implementation of the search string in SCOPUS, EBSCO and ProQuest ABI/INFORM electronic databases respectively. The details of the combined search are included in Table 4.5 below. It gives a breakdown of the document count hit and how they are eventually narrowed down to a manageable output.

Table 4.2 Search String Applied to SCOPUS

STEPS	Filters	Query Approach - Scopus	Hit Count
Step 1	None	TITLE-ABS-KEY ((algorithm OR data OR "Experimental data" OR "laboratory Data" OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata) W/5 (model*) AND (knowledge OR information) W/5 (manag* OR system*))	1,855
Step 2	**<u>Limit to:</u>** Publication year 2010- Date English language publications Conference proceedings Journal Articles Chapters and books **Exclude:** Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	TITLE-ABS-KEY ((algorithm OR data OR "Experimental data" OR "laboratory Data" OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata) W/5 (model*) AND (knowledge OR information) W/5 (manag* OR system*)) AND (LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011) OR LIMIT-TO (PUBYEAR , 2010)) AND (LIMIT-TO (LANGUAGE , "English")) AND (EXCLUDE (EXACTSRCTITLE , "Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics")) AND (EXCLUDE (SRCTYPE , "d"))	845
Step 3	Screen relevance: Exclude titles and abstracts that are not relevant for the study.	Based on the Researcher’s evaluation	129
Step 4	Screen the full-text articles applying the exclusion inclusion criteria.	Based on the Researcher’s assessment	119
Step 5	Merge with other database Step 4 results	Calculate total including duplicates	119
Step 6	Calculate total excluding duplicates from Step 5	Review these for gaps (making use of relevant works referenced as well)	115

The following table shows the screening process for items queried from EBSCO database

Table 4.3 Search String Applied to EBSCO

Step	Filters	Query Approach -EBSCO	Hit Count
STEP 1	None	(Algorithm OR Data OR "Experimental data" OR “laboratory Data” OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata) n5 (model*) AND (Knowledge OR information) n5 (manag* OR System*)	**1098**
STEP 2	Linked to full-Text Publication Year 2010-2017 English	((algorithm OR Data OR "Experimental data" OR “laboratory Data” OR "lab data" OR "raw data" OR "data management")) AND (("life cycle" OR lifecycle OR curation OR storage OR metadata) n5 (model*)) AND ((Knowledge OR information) n5 (manag* OR System*))	**138**
STEP 3	Screen relevance: Exclude titles and abstracts that are not relevant for the study.	Based on the Researcher’s evaluation	**39**
STEP 4	Screen the full-text articles applying the exclusion inclusion criteria.	Based on the Researcher’s assessment	**31**
STEP 5	Merge with other database Step 4 results	Calculate total including duplicates	**31**
STEP 6	Calculate total excluding duplicates from Step 5	Review these for gaps (making use of relevant works referenced as well)	**29**

The following table shows the screening process for items queried from ProQuest ABI/INFORM database.

Table 4.4 Search String Applied to ProQuest ABI/INFORM Collection

STEPS	Filters	Query Approach - ProQuest ABI/INFORM	Hit Count
Step 1	None	(Algorithm OR Data OR "Experimental data" OR “laboratory Data” OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata) n/5 (model*) AND (Knowledge OR information) n/5 (manag* OR System*)	306
Step 2	Limit to: Full-text and peer reviewed Publication Year 2010-2017 English Exclude trade journals	(all(Algorithm) OR all(Data) OR all("Experimental data") OR all(“laboratory Data”) OR all("lab data") OR all("raw data") OR all("data management")) AND (all("life cycle") OR all(lifecycle) OR all(curation) OR all(storage) OR all(metadata)) n/5 (all(model*)) AND (all(Knowledge) OR all(information)) n/5 (all(manag*) OR all(System*))	42
Step 3	Screen for relevance	Based on the Researcher’s evaluation	10
Step 4	Screen the full-text articles applying the exclusion inclusion criteria.	Based on the Researcher’s assessment	09
Step 5	Merge with other database Step 4 results	Calculate total including duplicates	09
Step 6	Calculate total excluding duplicates from Step 5	Review these for gaps (making use of relevant works referenced as well)	07

Table 4.5 below shows the process for selecting the relevant articles for the systematic review. It also shows the search strings used in the respective individual databases and the combined results.

Table 4.5 Combine Study Selection Process

Phases	Database	SCOPUS	EBSCO	ProQuest ABI/INFORM	Combined Hit Count
Phase 1	Apply Search String	(algorithm OR Data OR "Experimental data" OR “laboratory Data” OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata) w/5 (model*) AND (Knowledge OR information) w/5 (manag* OR System*)	(algorithm OR Data OR "Experimental data" OR “laboratory Data” OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata) n5 (model*) AND (Knowledge OR information) n5 (manag* OR System*)	(algorithm OR Data OR "Experimental data" OR “laboratory Data” OR "lab data" OR "raw data" OR "data management") AND ("life cycle" OR lifecycle OR curation OR storage OR metadata) n/5 (model*) AND (Knowledge OR information) n/5 (manag* OR System*)	
Phase 2	Records Identified	**1,855**	**1,098**	**306**	**3,259**
	Filters and Limiters Applied	**845** (1010 records excluded)	**138** (960 records excluded)	**42** (264 records excluded)	**1,025** (2448 excluded)
Phase 3	Screened for Relevant Titles and Abstracts	**129** (716 records excluded)	**39** (99 records excluded)	**10** (32 records excluded)	**178** (847 excluded)
Phase 4	Full-text Articles (Inclusion Exclusion Criteria Applied)	**119** (10 records excluded)	**31** (08 records excluded)	**09** (01 records excluded)	**159** (19 excluded)
Phase 5	Total (Including duplicates)	**159**			**159** (7 excluded)
	Total (Duplicates removed)	**152**			**152**

4.6 Study Selection Process

After implementing the search strings and identifying the records and applying filters and limiters, a two-phase approach was adopted for selecting relevant articles for review. In phase one, the researcher reviewed all the titles and abstracts of the citations for relevance. The researcher then retained those citations whose titles and abstracts met the criteria for admissibility. In phase two, the researcher assessed the full-text of articles retained in phase one for relevance. This technique is based on recommendations by Bettany-Saltikov (2012). Some of the full-text articles were evaluated twice to ascertain their significance.

4.7 Data Extraction Process

In this process, some of the selected articles, especially those that were more than 20 pages long, were read at least twice in other to have an unambiguous understanding of the context and content. An excel database was created and used store the data extracted from the articles. It collected data in the following categories:

- Bibliographic information
- Type of study
- Subject area, Aim, Objectives and scope of study
- Methods and techniques
- Challenges or issues, arguments, propositions and conclusions
- Models, standards and theories discussed
- Sector/Industry

The articles cited in closely related studies that relevant to this research were also selected for review and data extraction. Some of the standards cited in the literature were further examined for relevant material and data extraction.

4.8 Risk of Bias/Quality Assessment in Included Studies

A bias is a systematic error, or deviation from the truth, in results or inferences. Several different terms are used to talk about the assessment of studies underpinning a guideline — critical appraisal, quality assessment, internal validity —, but in this module, we use the concept of ‘risk of bias’. Bias refers to factors that can systematically affect the observations and conclusions of the study and cause them to be different from the truth (Higgins, Altman et al. 2011).

Sampling variations and confidence interval parameters were not part of the studies reviewed, and therefore issues of imprecision were avoided. The main risks in this context were largely related to methodological features and reporting. In addition, the data life cycle models reviewed are predominantly context-specific and by implication, paradoxically subjective and therefore, free of other sources of risks. The scoping review resulted in the protocol document that pre-specified elements of interest to the review reported in a way that can be replicated. Umbrella reviews were also considered in the assessment of the results, especially in cases where they had informed the creation of new models to avoid re-invention of the wheel in that context. “*A key difference between scoping reviews and systematic reviews is that the former are generally conducted to provide an overview of the existing evidence regardless of methodological quality or risk of bias (4, 5). Therefore, the included sources of evidence are typically not critically appraised for scoping reviews*”(Higgins and Green, 2008; Sterne, Egger and Moher, 2008).

Risk of bias assessment (sometimes called "quality assessment" or "critical appraisal") helps to establish transparency of evidence synthesis results and findings. A risk of bias assessment is often performed for each included study in the review. Risk of bias assessment generally is not required with evidence synthesis methods outside of systematic reviews. However, this may depend on the evidence synthesis method that you are utilising (Ballard and Montgomery, 2017). For this review, only peer-reviewed articles were included, which excluded the need for a further quality assessment.

4.9 Narrative-Synthesis (Meta-Synthesis)

A narrative-synthesis is the non-statistical method used in the integration, assessment and analysis of results from research studies that are predominantly qualitative. As a consequence of the diverse and qualitative nature of the studies reviewed, a narrative synthesis was conducted. A narrative synthesis is "*an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarise and explain the findings of the synthesis.*" (Popay *et al.*, 2019, p. 5). The narrative synthesis technique is suitable for combining and evaluating outcomes from qualitative studies. In this method, shared varieties of thematic areas are sort across discrete qualitative studies (Evans, 2002). There are four principles in the conduct of narrative synthesis applied to this research proposed by Rogers et al. (2009), and supported by Grant and Booth (2009), Higginbottom et al. (2012), The Joanna Briggs Institute (2014) and Young and Waddell (2016):

(One) Develop a theory.

(Two) Develop an initial synthesis mindful of qualitative or quantitative nature of the evidence.

(Three) Explore relationships between;

- (a) Various features of separate studies and their conclusions.
- (b) the conclusions of different studies.

(Four) Evaluate the sturdiness of the synthesis.

Contrary to meta-analysis used in quantitative reviews, this method has a rather descriptive than a combining aim. Meta-analysis involves the use of significant quantitative results from numerous studies to carry out statistical analysis using standardised procedures. In other words, it queries existing literature or research for statistical data for analysis (Lipsey and Wilson, 2001). This facilitates computing, detect relationships, assess the dispersion of effects, detect patterns, and the correlation between outcomes and studies (Lipsey and Wilson, 2001; Rosenthal and DiMatteo, 2002; Hedges and Olkin, 2014).

4.10 Results

The implementation of the search strings in the SCOPUS, EBSCO and ABI databases resulted in combined 3259 citations. The following filters and limiters were applied to these citations: limiting to - publication year 2010- date, English only language publications and conference proceedings, full-text and peer-reviewed Journal articles, chapters and books, and excluding blogs, trade journals, lecture notes on computer science, subseries lecture notes on artificial intelligence and lecture notes on bioinformatics. At this first stage of screening, of the 3259 citations, after applying filters and limiters 1025 citations were retained and 2448 citations removed. Subsequently, the titles and abstracts were screened for relevance leading to the exclusion of 847 citations that failed the eligibility test and inclusion of 178 articles for full-text assessment. Furthermore, 159 articles, including duplicates, were retained after the full-text assessment. Finally, of these 159 articles, 7 were duplicates leaving the total number of full-text articles to be reviewed at 152.

In the next phase, the review of these models is seeking to answer the following questions:

1. How can data- and knowledge management be improved for IVHM?
2. What is the overall weakness in the current application of data life cycle management and related standards on IVHM?
3. What experiences can we draw valuable lessons from?

4.10.1 Data Life Cycle Model

Five phase data life cycle models: Three five-phase models were identified; CRUD (Create, Read, Update and Delete), Enterprise Data life cycle model and Michigan State University (MSU) Records life cycle (Table 4.6). These models make use of the classical elements of the data life cycle. The CRUD model provides a flexible data life cycle; it only obliges creation, storage and destruction while allowing use, sharing and archiving as optional. The Enterprise Data life cycle model is a closed life cycle, and its emphasis on destruction limits the availability of data in the long-term (Arass, Tikito and Souissi, 2017).

Table 4.6 Five Phase Data Life Cycle Models

MODELS		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
1.	CRUD Data life Cycle (Yu and Wen, 2010)	Create	Store	Destruct	Use / Share	Archive
2.	Enterprise Data life Cycle(Chaki, 2015)	Creation / Receipt	Distribution	Consumption	Disposition /Archival	Destruction Retire
3.	MSU Records Life Cycle (Faundeen and Hutchison, 2017)	Create or Receive	Use and file	Transfer and store	Dispose	Archive / Destroy

Table 4.74 Six Phase Data Life Cycle Models (cf. Alexslis Maindze, Skaf, and Jennions 2019)

MODELS		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
4.	USGS) Science Data Lifecycle Model (Faundeen and Hutchison, 2017)	Plan	Acquire	Process	Analyse	Preserve	Publish/Share
		Describe(Metadata, Documentation)					
		Manage Quality					
		Backup & Secure					
5.	University of Virginia. Steps in the Data Life Cycle (University of Virginia, 2017)	Proposal Planning & Writing	Project Start-Up	Data Collection (access, security, storage)	Data Analysis (versioning)	Data Sharing	End of Project (archive /storage)
6.	International Leader in Data Stewardship (ICPSR) Data life cycle (ICPSR (Inter-university Consortium for Political and Social Research), 2012)	Proposal Development and Data Management Plans	Project Start-up	Data Collection and File Creation	Data Analysis	Preparing Data for Sharing	Depositing Data
7.	UCSD Libraries Data life Cycle(Cushman, 2018)	Propose (Define question, Design study, Write DMP)	Collect /Create (discover, clean-up)	Describe (Document, Create Metadata)	Analyse (process, visualise, interpret data)	Publish (Report, publish findings, present results)	Share / Preserve (Deposit data, assign identifiers, preserve long-term)
8.	Generic Life Cycle Model (INCOSE, 2015a)	Concept Stage	Development	Production	Utilisation	Support	Retirement
9.	UK Data Archive Data life cycle	Creating Data	Processing	Analysing Data	Preserving Data	Giving Access to Data	Re-using Data

Seven phase data life cycle models: The key features of these models (Table 4.8) are the inclusion of discovery, knowledge repository & re-use. There is a noticeable absence of access control & security element. The Geospatial Data Lifecycle is a flexible non-linear and non-sequential comprising the classical Define, Inventory/Evaluate, Obtain, Access, Maintain, Use/Evaluate and Archive stages but lacks the discovery and reusability elements. These may not be weaknesses in themselves as each life cycle model has been designed to cater to the needs of their respective organisations. As noted in [48], not all elements of the Data Life cycle are applicable to all contexts and scenarios.

Table 4.8 Seven Phase Data Life Cycle Models

MODELS		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7
10	FGDC Stages of the Geospatial Data Lifecycle (Faundeen and Hutchison, 2017)(Ball, 2012)(Federal Geographic Data Committee - FGDC, 2010)	Define (user requirements)	Inventory/ Evaluate	Obtain	Access	Maintain	Use / Evaluate	Archive
11	Information Life Cycle	Data Generation	Data Transmission	Data storage	Data Access	Data Re-use	Data Archiving	Data Disposal
12	JISC: Supporting the Research Data Lifecycle (Arass, Tikito and Souissi, 2017)	Plan	Create (Store /Annotate)	Use	Appraise (Select / Discard)	Publish (Describe / Identify)	Discover (Access)	Re-use

Eight phase data life cycle models: The DataOne data life cycle model is one of eight phase data life cycle models (Table 4.9) designed by the National Science Foundation (NSF). It exclusively focuses on the phases field or laboratory data goes through rather than the role of person acting on the data (Allard, 2014). The model is developed as a foundation for the development of tools and services for data-intensive sciences and for encouraging best practices (Pouchard, 2015; Arass, Tikito and Souissi, 2017; DataONE, 2017; Plale and Kouper, 2017). The experimental geomorphology data life cycle model addresses the complexity in analysis, storage and search and retrieval posed by increasing volumes of laboratory data (Hsu *et al.*, 2015).

Table 4.5 Eight Phase Data Life Cycle Models

MODELS		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8
13	The Data Observation Network for Earth (DataONE) data life cycle (Pouchard, 2015) (Arass, Tikito and Souissi, 2017)(Plale and Kouper, 2017)(DataONE, 2017).	Plan	Collect	Assure (quality)	Describe (metadata standards)	Preserve (archive)	Discover (retrieval)	Integrate (Synthesis)	Analyse
14	DSD Laboratories lifecycle management of data (Laboratories, 2014)	Plan	Collect	Assure	Describe	Discover	Integrate	Analyse	Archive
15	Experimental Geomorphology (Hsu *et al.*, 2015)	Conception	Discovery (preliminary investigation)	Project Start-up (Create workflows)	Data collection and file creation	Data processing & Analysis	Preparation for Data Sharing	Publication of results	Discovery & Renewing the cycle
16	Data Documentation Initiative(DDI) (Arass, Tikito and Souissi,	Study Concept	Data Collection	Data Processing	Data Archiving	Data Distribution	Data Discovery	Data Analysis	Repurposing

<table>
<tr><th colspan="2">MODELS</th><th>Phase 1</th><th>Phase 2</th><th>Phase 3</th><th>Phase 4</th><th>Phase 5</th><th>Phase 6</th><th>Phase 7</th><th>Phase 8</th></tr>
<tr><td></td><td>2017) (Ball, 2012)
(Gregory, 2011)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="5">17</td><td rowspan="5">Library and Archives Canada (LAC)
Data Life Cycle
(Canadian Space Agency, 2017)</td><td>Planning</td><td>Collection, Creation or Reception</td><td>Storage and accessibility</td><td>Transform, analysis and use</td><td>Organisation</td><td>Distribution</td><td>Preservation & Disposal</td><td>Evaluation</td></tr>
<tr><td colspan="2"></td><td colspan="2">Use / Re-Use as part of the initial project</td><td colspan="3"></td><td rowspan="3"></td></tr>
<tr><td colspan="4"></td><td></td><td colspan="2">Parallel Actions</td></tr>
<tr><td colspan="7">Unexpected re-use</td></tr>
<tr><td colspan="8">Complete life cycle</td></tr>
</table>

Hsu et al. (2015) review data management practices and challenges for experimental geomorphology, and established that the lack of rules for metadata integration, documentation of workflow, data storage, motivation and training were impeding significant amounts of data sharing and re-use. These challenges mean that accessibility or availability of experimental or laboratory data is limited to the research or project life cycles, and therefore undermining reproducibility and quality control. They suggested that efficient and effective data management and sharing should consider the entire data life cycle, including metadata set information, disciplinary information, Quality information and readiness for re-use.

The Data Documentation Initiative (DDI) version 3.0 mostly linear, combined a life cycle model, integrates data application perspective and social science research data. Metadata requirements are grouped into five comprising conception of the study, collecting of data, logical data encoding structure, physical data encoding structure and archiving (Ball, 2012).

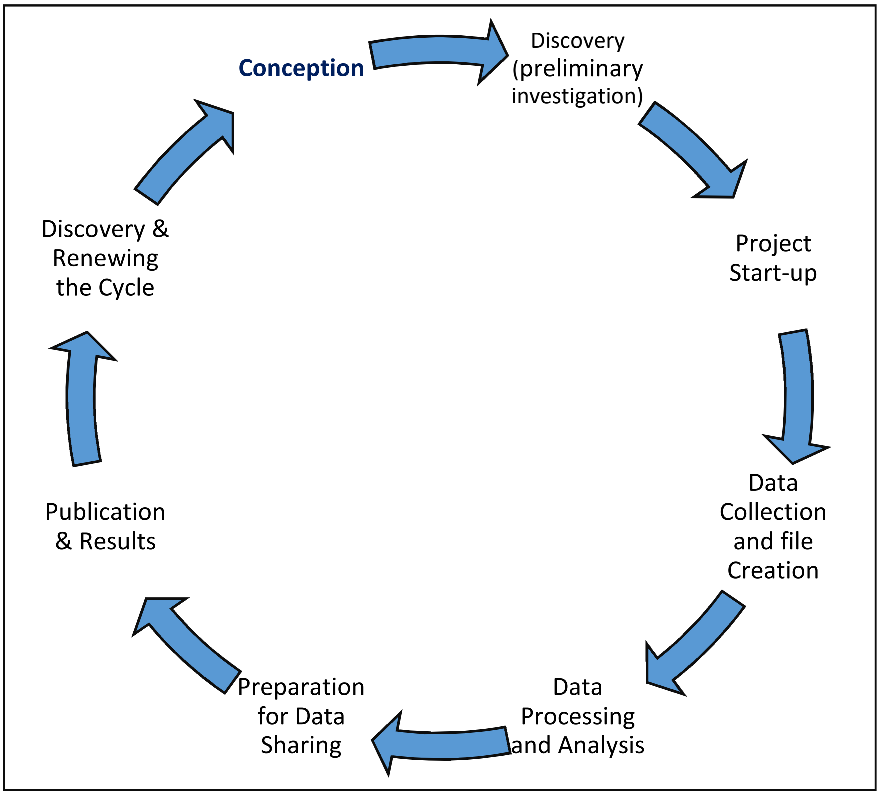


Figure 4-3 The Experimental Geomorphology Lifecycle (Hsu et al., 2015)

Table 4.10 Nine Phase Data Life Cycle Model

MODELS		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9
18	Digital Curation Centre (DCC) Curation Lifecycle Model (Faundeen and Hutchison, 2017)(Ball, 2012)	Conceptualise (Conceive and plan)	Create or Receive (Metadata created)	Appraise & Select (Quality+ Governance)	Ingest (Transfer to storage or archive)	Preservation Action	Store	Access, Use & Re-use	Transform (Migrate)	Occasional Actions Dispose Re-appraise Migrate

The DCC curation life cycle model (Figure 4.4, cf. J. Faundeen and Hutchison 2017; Ball 2012; DCC 2008) is the model with the highest number of phases with emphasis on preservation, archiving and management of data and publications for long-term availability and re-use. It concentrates on data curation and serves as a planning tool for data creators, curators and users. The model features a complete lifecycle of actions, sequential actions and occasional situational actions (Plale and Kouper, 2017). The model presents indispensable phases in the life cycle by using circles with a common center. The phases include: (i) Conceptualize (Conceive & Plan); (ii) Create or Receive (Generate metadata); (iii) Appraise and select (Quality and Governance); (iv) Ingest (Move to Storage or Archive); (v) Preservation action (Apply retention schedules); (vi) Store (Secure Storage); (vii) Access, use and reuse (Access policy and control); (viii) Transform (Migration), plus (ix) Occasional Actions: Migrate, Dispose, Reappraise and

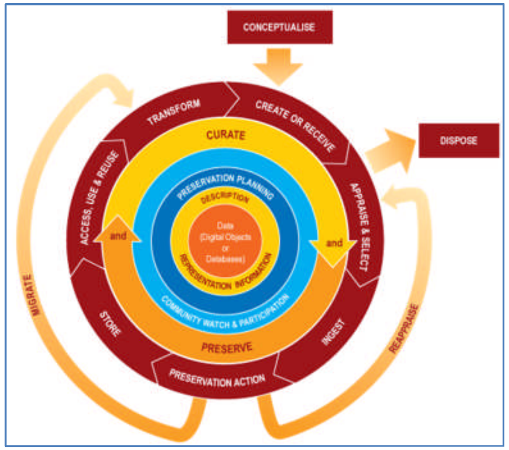


Figure 4-4 DCC Life cycle Model

(DCC, 2008)

Dispose. The DCC life cycle model provides an applied framework that can be loosely categorised into different levels and areas of curation such as technology level operations, bit-level preservation routines and metadata curation. The technology level operations involve migration, backup, indexing, and system upgrades. Bit-Level preservation routines include data-recording, checksum reporting, error-correcting and character replacement and file format registries. The metadata curation level defines the content and context of digital elements. This level fulfils the technical, descriptive, structural and preservation metadata requirements (Sabharwal, 2015; Parry, 2016).

One of many technical standards for metadata is the Dublin core. Activities that generate rich footnotes and significance to images in Digital humanities scholarship and teaching practices are given three levels of Curation.

Level one (L1) focuses digital files and technology used in their preservation whilst levels two and three denotes the exercise of intellectual control and scholarly processing, respectively (Figure 4.5). Furthermore, underlying these levels are aspects of metadata schema such as Dublin Core, interoperability standards such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), file format and data encoding, network characteristics and reliable hardware and software systems(Sabharwal, 2015)

.

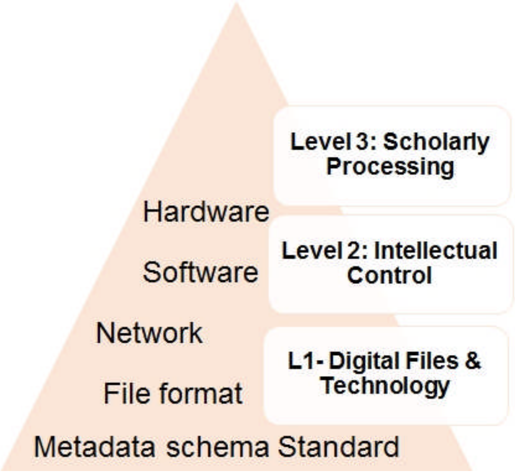


Figure 4-5 Levels of Data Curation

In addition to some of the standard mentioned above, Table 4.11 below shows four main standards commonly used to support data management through the data life cycle. The design of data life cycle models takes into consideration relevant and appropriate standards/frameworks to support respective domain data. Some relevant standards identified in the literature include the Open System Architecture for Condition Based Maintenance (OSA-CBM), the Core Scientific Metadata Model (CSMD), INCOSE systems engineering Management process (the V-model) and the National Oceanic and Atmospheric Administration (NOAA) Environmental Data Management (EDM) Framework. Each plays an important role in the data management life cycle as discussed below.

Table 4.61 Standards and Framework Identified

Standard / Framework	**Layer 1**	**Layer 2**	**Layer 3**	**Layer 4**	**Layer 5**	**Layer 6**	**Layer 7**
Open System Architecture for Condition Based Maintenance (OSA-CBM)(Jantunen *et al.*, 2017) (Löhr and Buderath, 2014)	Data Acquisition (DA) Using sensors or transducers to convert physical phenomenon into readable digital signals.	Data Manipulation (DM) Transformation and storage using algorithms.	State Detection (SD) Compare data from (DA) and (DM) against expected values	Health Assessment (HA) Analyse & determine the health level of a monitored system or device.	Prognostics Assessment (PA) model-based, data-driven or hybrid approach to project the remaining useful life (RUL) of asset	Advisory Generation (AG) Provide a report stating recommended optimal maintenance actions.	Presentation Presentation of information to system users (Xu and Xu, 2017)
Core Scientific Metadata Model (CSMD)(Yang, Matthews and Wilson, 2013)	Core Data file, Dataset Investigation, Investigator, SoftwareVersion, Parameter,Study, Study Manager	Security User_role ICAT_Authorisation ICAT_role Application	Communication Publication	Miscellaneous This_ICAT	Search Keyword Topic	Facility Shift, Instrument, Facility_instrument_scientist, Facility_cycle, Facility_user	Auxiliary information Status Sample Format Type
INCOSE systems engineering Management process. The V-Model (INCOSE, 2015a)	**S**tating the problem (concept studies)	**I**nvestigating alternatives (Concept development)	**M**odeling the system (preliminary design)	**I**ntegrating the system (final design)	**L**aunching the system (Fabrication)	**A**ssessing performance (verify components /performance)	**R**e-evaluation (demonstration & Validation)
National Oceanic and Atmospheric Administration (NOAA) Environmental Data Management (EDM) Framework (Beaujardière, 2016)**.**	**Planning and Production**	**Data Management** Processing, Verifying, documenting, advertising, distributing, preservation.	**Usage** **Discovery** **Analysis** **Tagging** **Gap assessment**				

The National Oceanic and Atmospheric Administration (NOAA) Environmental Data Management (EDM) Framework (Figure 4.6) (Beaujardière, 2016) was developed in 2012 to guide their data management plans and related activities for improved data management. The NOAA Framework has three main purposes (Beaujardière, 2016): One - to encourage a shared understanding of data management policies and activities across NOAA, Two - to make the most of the probability that environmental data are discoverable, accessible, well documented, and preserved for future use, and Three - to promote the development and use of homogeneous tools and practices across NOAA for handling environmental data. The core components (Beaujardière, 2016) of the NOAA EDM guidelines applicable to different classes of data and individual data life cycles comprises the following:

- **Assessment:** Assessing the planning, preservation, accessibility levels, metadata quality, valuing the current state, measuring progress, and getting feedback from users and implementers.
- **Architecture:** In the EDM framework, this includes observation platforms and systems, data collection and data processing systems, archival data canters and related data ingestion, storage and stewardship systems. Adoption of interoperability standards vital for support and streamline information sharing between internal as well as external data providers.
- **Standards:** Application of different data quality standards, metadata standards, interface standards, data models and format standards, common vocabularies that specify content and structure data, dataset documentation and how services are invoked.
- **Resources:** Competent and motivated personnel, cost of creating observations and data sharing, data access and data citation including pilot projects, teams, conferences, documentation and software.
- **Governance:** Provision of policies and strategic documents, monitoring and enforcement of data management procedures.

- **Principles:** Provides guidelines for case-by-case; full and open access to datasets, long-term data preservation, and quality assurance with ease of online discoverability and accessibility underscored by interoperability.

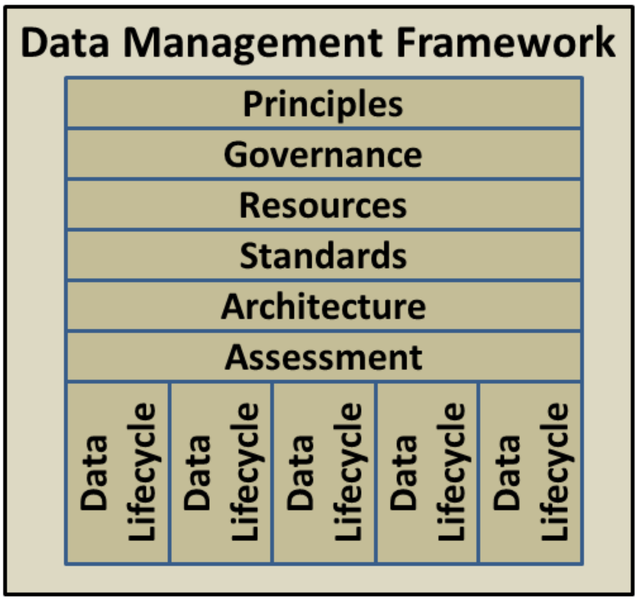


Figure 4-6 The Environmental Data Management Framework (Beaujardière, 2016)

The data life cycle is split into three activity groupings: (1) Planning and Production – complete system observations and data collection activities; (2) Data Management – Comprehensive data processing, verification, documentation, advertising, distribution and preservation activities; and (3) Usage – this comprises all end-user activities on the data.

4.10.1.1. Data Life Cycle Management Case Studies

The literature reveals a diversity of data life cycle approaches, standards, reference models by different organisations, each tailoring the classical data life cycle model or creating an entirely new one to suit their individual organisation requirements.

4.10.1.2 Addressing interoperability, reproducibility and re-use

In their reflective case study on the development and implementation of common data engineering architectures in the U.S. Army Test and Evaluation Command

(ATEC), Andreas Tolk and Robert Aaron (2010) proposed a continuous interoperability solution in a bid to highlight the significance of effective and efficient data engineering in data transfer. The main objective was to define a requirements approach that would guarantee the unambiguous interpretation and structure of data. They proposed the use of metadata as a mitigation major for the risk of transferring the same data twice caused by unreliable and bad data. They argued that regardless of the standards used, orchestrating the phases and steps was ultimate. They recommended that a common reference model in which four key steps; data administration, data management, data alignment and data transformation are performed should be made more explicit for the purposes of engineering management. They concluded that supporting a common data architecture comparatively in domains would facilitate understanding, interoperability, reproducibility and re-use.

Furthermore, the perspective is supported by Porcal-Gonzalo (2015) who analysed metadata schema ISO 19115 to define a framework for data preservation and information re-use for the EU and Spain, proposed the creation of SDIs-based networked system for the preservation of geographical information underpinned by metadata. The "SDI is a virtual networked structure of geographical, and therefore georeferenced data and geographical information interoperable services"*[75, P. 291]*. This provides the management and maintenance protocols that guarantee the efficient and effective consumption of the short and long-term. Despite the challenges posed by frailty and volume of data which decreases stability and durability to this effect, careful planning and strict regulation of accessibility and use ensure the mitigation effect required. Given the significance of metadata and interoperability in assuring reliability, authenticity, completeness, accountability, usability and accessibility, the author suggested mandating metadata related to preservation as a way of achieving balance across the data life cycle from planning to re-use. The author introduces the element of undertaking "*periodic cost-benefit analysis*" to forecast the cost of data preservation, and the significance of creating metadata from the planning to the re-use phase. This proposition is adopted by the United States Geological Survey (USGS) (Faundeen and Hutchison, 2017) in the design of their Data Lifecycle model. They conclude by recommending a system that can is complex, networked enough to adapt changing models, standards, formats and technological developments.

In another study, Ofner et al. (2013) proposed an all-inclusive reference framework for master Data life cycle Management (MDLM) by applying the Component Business Model (CBM) method to resolve the uncoordinated processing of shared master data resources and the inconsistency and inaccuracies of master data. They reviewed a number of case studies and reference models and concluded that the MDLM model was grounded in the existing literature and demonstrated the relationship between master data life cycle and strategic decisions and the impact of the latter on it. This reference standard can also help reveal gaps and stress points to fine granularity. They finally acknowledged the need for more research to extend and detail the reference standard.

With objective "*to provide a formal and infrastructure agnostic model to describe data life cycles in distributed systems and a programming model which facilitates data life cycle management for developers*" [p.26] based on Petri networks, Simonet et al. (2015) deployed four uses cases:

- (i) **Storage cache:** ability and ease with which to program distributed applications with Active Data.
- (ii) **Collaborative sensor network:**
    - a. the flexibility to bequest data management systems;
    - b. the capability to build heterogeneous applications that can sustain independent data life cycles scattered all through many local systems;
    - c. the simplicity of execution and organisation amongst dispersed nodes with Active Data.
- (iii) **Incremental MapReduce:** How can a current system be enhanced by strengthening an Active Data's capacity to handle variable data?

(iv) **Data provenance:** the entire history of spin-offs and coverage throughout the data life cycle, fundamental to the preservation of the quality of scientific data asset over time.

The above scenarios were run using the Grid500 grid test environment by integration into five data management systems; Bitdew, Inotify, iRODS, Globus Online, and HDFS & Hadoop. The Active Data executes user code through data life cycle from planning, creation, replication, and transfer through to deletion. In conclusion, the alleviation of both the user and application imposed complexity of Data life Cycle Data - transfer, archiving, replication, processing, deletion, can be achieved through Active Data - "*a formal model* [*"…Transition-based programming model" for distributed data life cycles and a programming model to allow code execution at each stage of the data life cycle*"] (Simonet, Fedak and Ripeanu, 2015, p. 26).

4.11 Consolidated Research Questions

The address objective, a number of questions were identified through the systematic review.

i. How can IVHM data management requirements for medium to long-term data availability be addressed?
ii. What can be done to guarantee, the security, reliability and quality of IVHM data

4.12 Discussion (Maindze, Skaf and Jennions, 2019)

Data management through its entire life cycle still presents a number of complex challenges relating to interoperability, volume, storage, data citation, and metadata standards and data provenance (Yang, Matthews and Wilson, 2013; Porcal-Gonzalo, 2015; Beaujardière, 2016). This perhaps explains the proliferation of discipline or domain-specific data life cycle models. Various disciplines and organisations are creating standardised frameworks, data ontologies, standards and unique data life cycle models to suit their respective requirements. Metadata standards like Dublin core (Hsu *et al.*, 2015), Core Scientific Metadata (CSMD) provisions the basic metadata required to enhance the search functionalities over data portals and knowledge libraries (Matthews *et al.*, 2010) but however falls short of propagating the complete provenance data. The shortfall with the CSMD is that neither does it support for access to the derived data produced during analysis nor does it allow the provenance of data supporting the final publication to be traced through the stages of analysis to the raw data.

However, though metadata standards are relevant for all data and knowledge outputs and organisations, each organisation or project seems to have separate requirements for their own research or projects data. This is reflected in the 17 data life cycle models identified, which all highlight the significance of metadata and standards. The models seem to differ on the depth and breadth of applicability and priority of requirements. For instance, the USGS emphasised three critical cross-cutting activities namely; metadata description, quality assurance and protection from corruption or loss to be performed parallel to planning, acquisition, processing, analysis, preservation, publishing and sharing to achieve enhanced quality, understanding and long-term reuse (Plale and Kouper, 2017)(Faundeen and Hutchison, 2017). The Digital Curation Centre (DCC) model on the other hand, suggests metadata should comprise rules and formalised entities for automatable tools and services as well as the role of data managers and data curators in the improvement of knowledge (Plale and Kouper, 2017). Taking the DataOne data life cycle model to illustrate the variance, though it represents all the classical stages of the data management life cycle, all parts of the life cycle are not mandatory (Harrison, 2013; Allard, 2014; Pouchard, 2015; DataONE, 2017; Hidalga *et al.*, 2017; Plale and Kouper, 2017) This implies that the number of selected stages are dependent on the type of project or project requirements.

Furthermore, although the USGS addresses the weakness in most data life cycle models identified and can be used in diverse settings despite having been developed mainly for USGS science data. The requirements on which this is developed are not consistent with small research facilities like the IVHM centre Labs or the University as a whole. None of the models identified seems to satisfy the scenario of IVHM.

The core of IVHM is the capture data and analysis to establish an advance indication of a future fault, state or distinctive characteristics of current assets in the current state. The conduct of prognostics and diagnostics is reliant on high-quality, reliable data. This quality assurance can be determined by the data management life cycle. Data life cycle models are often integrated with software services and policies (Plale and Kouper, 2017). It is, therefore, fundamental to consider the associated process involved in the design and development of the software services and policies.

The INCOSE system engineering process [V-Model] is suitable for the development of any system, is vital in the case of IVHM Knowledge Management System and data lifecycle model. This V-Model is suitable for the IVHM data life cycle and system development because of the restricted nature of this project. The V-Model is best suited for these kinds of projects that have well-defined length and scope, consistent technology, and a clear and well documented technical and functional specifications (INCOSE, 2017).

4.13 Key Standards

Standards are documents that include principles, specifications, directives or features to ensure that virtual and hard entities, products, processes and services are suitable for their intended use (Ransom *et al.*, 2017). The use of standards facilitates the collection, collation, storage, exchange and incorporation of data by ensuring that there is a clear perception of how the data is interpreted and that the data is secure, high quality, interoperable, durable and has ensured integrity. Standards therefore, underpin the implementation of new technologies and transformations and ensure interoperability between entities, constituents and services provided by various organisations.

Furthermore, they spread awareness in industries where the goods and processes supplied by different suppliers who have to communicate with each other.

Standardisation is a consensual collaboration for the development of consensus-based technical specifications between industry, consumers, public authorities, researchers, and other interested parties (Brunsson, Rasche and Seidl, 2012) in a way that directs and normalises our activities.

For instance, the ability of the interfaces of different technological devices and systems to operate in conjunction with each other are controlled through the use of Information Technology standards. In manufacturing, the consistent maintaining, repairing and reproducing of diverse technologies and appliances are facilitated through standardisation. There are many organisations developing various standards globally, but the most popular ones are the International Organization for Standardization (ISO), and the International Electro-technical Commission (IEC) usually working in collaboration with other organisations.

The following five standards are extremely relevant in the conduct of IVHM operations, activities, diagnostics, prognostics, maintenance, and research, data- and knowledge management. They cover the development of both the IVHM data management and the IVHM Knowledge Management Systems.

4.13.1 BS ISO 30401:2018(E) Standard – Knowledge Management System

This standard was published in November 2018 in order to help organisations establish a knowledge-based management system that values and encourages value creation. Therefore, the aim is to conserve, disseminate, distribute and expand the tangible and intangible knowledge resources of an organisation. The framework sets out the principles and criteria for implementing an effective and efficient Knowledge Management System (KMS). The framework recommends a structured KMS so that varied entities can interact with this model, which is versatile and easy to understand. The standard covers the following areas for integration into the KMS life cycle requirements:

i. Leadership: Specify the necessary framework and precise commitment and management policy at the company level
ii. Planning: define objectives, calculate risks and opportunities to create time benchmarks
iii. Support: determine and provide the human, technological and documentary resources needed to implement change management
iv. Operations: Plan activities and governance strategy
v. Performance evaluation: implementation of KPIs, management reviews and KMS audits
vi. Improvement: define new measures and actions to overcome the difficulties targeted by the evaluation(Institution, 2018).

4.13.2 ISO 8000 – Master Data: Quality Management Framework

Data Governance is an organisation of decision rights and accountability for information-related procedures, implemented according to agreed-upon models that define roles and responsibilities for information management, including when and under what conditions and what method should be used (Rosenbaum, 2010; Reeves and Bowen, 2013). The ISO 8000 international standard for data quality defines the criteria, guidelines, specifications or features that can often be used to safeguard the expected value of provisioned materials, products, procedures and services. This framework aligns with ISO 9000 that mandates possession of explicit documented requirements to be compliant by obliging that requirements must be “stated” in a form that can easily be processed by a computer. The framework specifies the methodology depicted by the model (Figure 4.7) below:

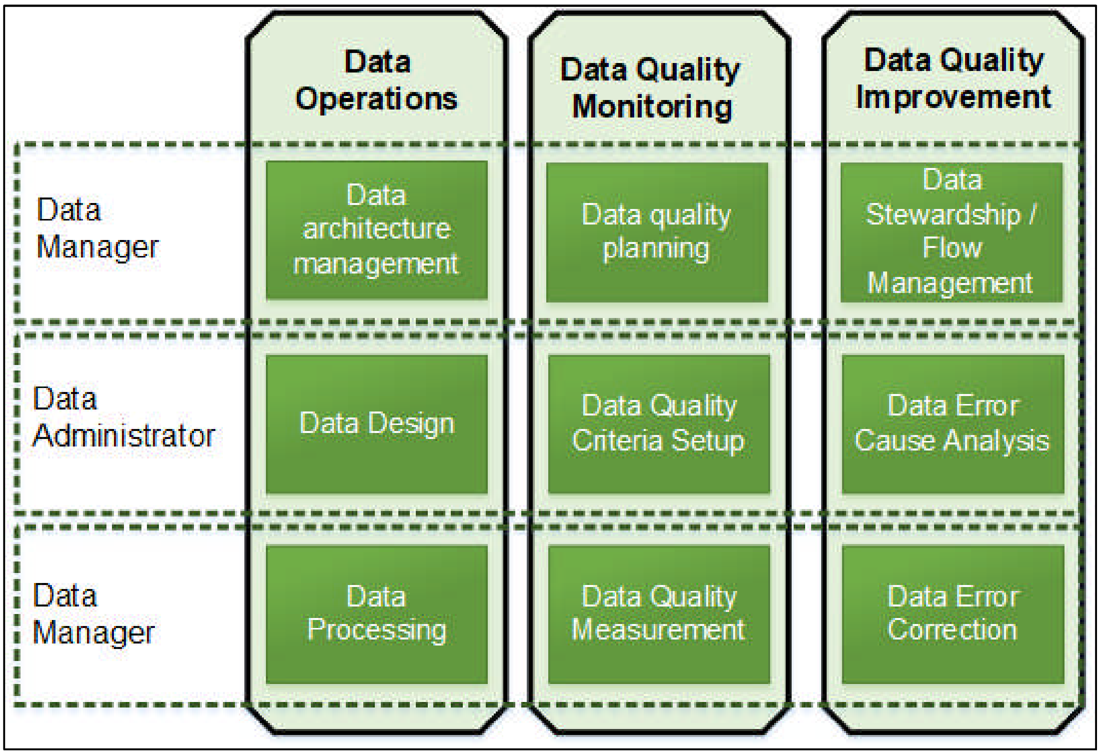


Figure 4-7 Master Data: Quality Management Framework (ISO8000)

There are three basic tenets that underlie the application of the standard, namely, people, process and continuous improvement.

- People – data quality management is rather grounded in the actions of individuals than a technological execution.
- Process – the effectiveness of data management depends on various fundamental processes.
- Continuous improvement – regular improvement should be applied to both the quality and the processes applied.

This framework is split into three upright 'processes' and three-level 'roles'. These fundamental processes include:

- **Data Operations processes:** These processes lay emphasis on the features that have an impact on both the quality of data and usage of data.
  - **Data Architecture Management** governs the data architecture for the entire establishment.

    - **Data Design** provides governance for data standards and definitions, database and system implementation and configuration.
    - **Data Processing** incorporates opportunities for generating, manipulating, updating and transmitting data.
- **Data Quality Monitoring** specifies a structured method for evaluating data quality levels.
    - **Data Quality Planning** helps set the data quality management goals and targets to align with organisational goals.
    - **Data Quality Criteria Setup** establishes measures and methods for assessing the quality of data.
    - **Data Quality Measurement:** is the procedure that actually uses those same data quality specifications in order to evaluate the levels of data quality.
- **Data Quality Improvement** is a process that fixes the observed data errors and removes the root causes of the data errors.
    - **Data Stewardship and Flow Management** refers to the system that evaluates data flows and duties and maintains the stewardship of data operations.
    - **Data Error Cause Analysis** is the systematic approach to identifying root causes of data errors in order to stop them from repeating.
    - **Data Error Correction** is the procedure that tries to fix data that does not fulfil standards or data quality requirements.

The framework further designates three broad functions: Data Manager, Data Administrator and Data Technician. These roles provide an overview of whether low-level procedures are strategic, tactical or functional.

In all, the standards provides for key parameters (Figure 4.8) for data management name quality and integrity, auditability, accountability and change management.

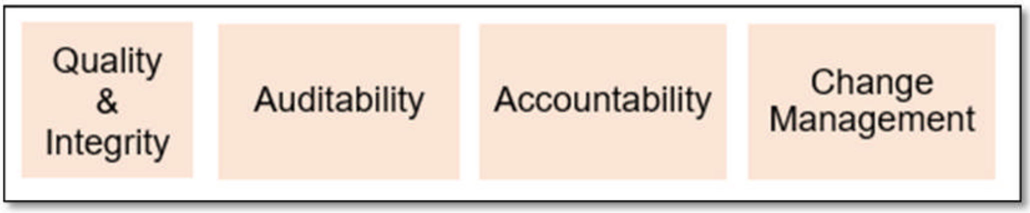


Figure 4-8 Four Fundamental Parameters for Data Management

The design of a data life cycle model for IVHM requires a careful consideration and integration of these features.

4.13.3 ISO/IEC 38500: IT Governance

ISO / IEC 38500 is a global standard for all organisations' corporate governance practices of information technology, which sets out principles, definitions and a model to encourage organisations to evaluate, guide and monitor the efficient , effective and appropriate usage of Information Technology (IT) within the workplace (Calder, 2008b, 2008a). The ISO / IEC 38500:2015 exists to strengthen secure, effective and appropriate use of IT in all organisations with emphasis on a number of fundamental characteristics depicted in Figure 4.9 below. The standard

Figure 4-9 Key Elements of Data Management in ISO38500

Achieves its stated aim and objectives through the following:

- Helping to ensure that stakeholders will have trust in the IT governance of an entity if the standards and practices suggested by the standard are adhered to;
- Advising and directing governing bodies in their respective organisations about the use of IT;
- Informing the creating and maintaining of an Information Technology Governance Taxonomy.

4.13.4 Dublin Core Metadata Initiative (DCMI)

DCMI is an open organisation that supports metadata design innovation and best practises across the metadata ecology community. Dublin Core has its roots in the Dublin Core Metadata Initiative, a short DCMI, created in Chicago in 1994. The Dublin Core Schema is a small set of vocabulary terms which can be used to characterise electronic media, as well as physical resources such as books or CDs. Dublin Core Metadata Element Set (DCMES) is a standard metadata schema (Table 4.12 below) developed to explain electronic records, documents, and web resources. It provides metadata vocabularies, concept schemes and interoperability of datasets. Schemas are machine-processable specifications that describe the metadata specifications architecture and syntax in a formal language of a schema.

DCMI working groups have developed standards on other metadata-related issues, namely encoding syntaxes, guidelines for use as well as metadata models. The two main schemas of the Dublin core are the Extensible Mark-up Language (XML) schema and the Resource Description Framework (RDF) serving two main purposes accordingly:

- XML schemas provide a way to describe XML document structure, including metadata.
- RDF includes metadata terminology for use in Semantic Web applications and Linkedin data implementations.

Table 4.7 The Dublin Core Set

Generic Metadata Data Elements		
1. Creator 2. Contributor 3. Publisher 4. Title 5. Date	6. Language 7. Format 8. Subject 9. Description 10. Identifier	11. Relation 12. Source 13. Type 14. Coverage 15. Rights

Developers generally use Dublin Core set in relation to relational databases and repositories, a great number of them use xml to describe the details of metadata records as structured documents. Document format implementers prefer text standards, closed-world quality assurance, top-down compliance, and focus on well-understood, tried-and-true software solutions. The perception of interoperability across domains and applications happens in the context of conformity to fixed formats (DCMI, no date b, no date a).

4.13.5 ISO15489 Records Management

The ISO15489 is the first worldwide records management standard that was published in 2001. It has since been translated into 15 languages and adopted in over 50 countries globally. An updated edition of ISO 15489 Part 1 was released in 2016, following a three-year revision and consultation cycle. The standard spells out the key concepts and principles for records creation, capture, storage and management regardless of their structure and form. The standard applies to all types of organisations and business. This includes the basic concepts and values for:

- Records governance
- Procedures for records capture, creation, storage and management.
- Annals, their storage systems as well as their metadata.
- Procedures, Policies, assigned roles and responsibilities, supervision and training to promote efficient and effective record-keeping.

- Regular review of current and future business environment and the identification of the prerequisite for records.

This first section of ISO 15489 is at the centre of a range of global standards and Technical documents offering more guidelines and advice on principles, procedures and methods for records-creation, records-capture, storage and maintenance. In addition, Part/s of the standard provides implementation guidelines on managing records.

4.14 Designing and Validating the Data Life Cycle Model

McKenzie (2010, p. 148) describes conceptual models "*generally informal and typically graphic depictions of systems that quickly and easily convey the overall functionality of a system*". With regards to specificity and precision, conceptual models are usually very informal, high-level description of an application or process.

Conceptual models generally specify and describe the application's destination task-domain, purpose and high-level functionality; the application's concepts expose users, including the task-domain data objects created and manipulated by users, their user-visible names, their attributes and the operations that can be performed on them; the links between those concepts; mapping of task-domain concepts to application concepts, typically expressed as conceptual model-written task scenarios. They are not scenarios of task level, mental models, and metaphors for design, use cases, or architecture for implementation (Johnson and Henderson, 2011; Johnson, 2015). The conceptual model focuses on application requirements or process organisation, and it is in this frame that the Data life Cycle model fits. The DLCM model provides a conceptual framework for creating and formalising data management rules for various task domains. Microsoft Visio shall be used for diagramming the data Life Cycle model.

4.14.1 Model Validation Techniques

Validation has often been described as the "*process of determining the degree to which a model is an accurate representation of the real world*" (Coleman and Steele, 2018) "*from the perspective of the intended uses*" (Carson, 2002; Sornette *et al.*, 2007; Graziani, 2008). A variety of specific techniques have been suggested in the modelling and simulation literature. Balci (1994) proposes a variety of 45 techniques for verification, validation and testing, Grouped into formal, static, dynamic, symbolic, constraint and formal. This research adopted peer reviews, walkthroughs, and face validation techniques from the Informal category to validate the DLCM due to the non-statistical or qualitative character of the proposed model. In addition, Carson (2002) cautions that no model can be verified or validated by 100% due to the fact that the approval of a model is never absolute. In addition, as suggested above, ad-hoc validation techniques may be established for a particular model and system.

Technical accuracy and completeness can be reliably determined through structured walkthroughs. A structured walkthrough can be described as a standardised process of peer-reviewing technical aspects of systems development, outputs and maintenance. The goal of the walk-through is to find flaws and boost improvements. Usually, errors arise as omissions or discrepancies, logic defects or design inconsistencies. This makes use of the structuring principle, which is defined by Krogstie (2016a, p. 23) as "*rule or assumption concerning how a model should be structured.*" He further proposed the SEQUAL- Semiotic Quality Framework with the five quality categories (Krogstie, 2016b, 2016a, 2017) adopted in this research:

- Physical quality issues: Determines the fitness for purpose and familiarity with the model
- Empirical quality issues: evaluates the clarity of texts and abbreviations used in the model
- Semantic Quality issues: examines users ability to make improvement suggestions to the model

- Pragmatic and social quality issues: Determines the ease of understanding and the use of a model. Looks at the agreement in knowledge versus agreement in model interpretation of end-users. Is there agreement on the quality of the model?
- Deontic Quality: Contribution to System goals and to the management of Data

4.15 Conclusion

Integrated Vehicle Health Management (IVHM) is shaped by prognostics and diagnostics that rely predominantly on the availability of high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health. The data has to be accurate, complete, timely, context-relevant, reliable and explicit (Dibsdale, 2011). IVHM is data-centric and driven by the OSA-CBM data model. The centrality of data for IVHM in the short, medium and long-term diagnostics whether it be historical asset health trends, operational status, load history, or fault identification, necessitates a data life cycle model or a hybrid model consistent with OSA-CBM. Of all the existing data life cycle models, there is none that is consistent with the requirements of IVHM data and knowledge management requirements that is, that integrates OSA-CBM, which is absolutely imperative to IVHM. The OSA-CBM model is based on the concept of metadata and interoperability that requires persistent visibility and traceability of data (Choudhary, Perinpanayagam and Butans, 2016) within and across diverse platforms, systems and devices, and therefore making data provenance a fundamental requirement. This has not been explicitly covered in any of the data life cycle models. A new model is needed, and one that integrates data acquisition with signal reception as well as data entry in order to accommodate the role of the systems or device operators for IVHM.

The emerging research questions are:

How can we further adapt and specify a data life cycle for IVHM? How can our model accommodate shifts in artificial intelligence and automation of diagnostics and prognostics? How can the OSA-CBM standard and Data life cycle be integrated for sustainable real-time and long-term data availability for IVHM and knowledge Management? How can we preserve, share data and design algorithms for interoperability and provenance of data for diagnostics and prognostics for legacy systems and assets?

4.16 References

Adams, R. J., Smart, P. and Huff, A. S. (2017) ‘Shades of Grey: Guidelines for Working with the Grey Literature in Systematic Reviews for Management and Organizational Studies’, International Journal of Management Reviews. Wiley Online Library, 19(4), pp. 432–454.

Allard, S. (2014) ‘Evaluating a Complex Project: DataOne’, in Ray, J. M. (ed.) Research Data Management: Practical Strategies for Information Professionals. Purdue University Press (Charleston Insights in Library, Informat), p. 436.

Arass, M. E., Tikito, I. and Souissi, N. (2017) ‘Data lifecycles analysis: Towards intelligent cycle’, 2017 Intelligent Systems and Computer Vision, ISCV 2017. doi: 10.1109/ISACV.2017.8054938.

Arksey, H. and O’Malley, L. (2005) ‘Scoping Studies: Towards a Methodological Framework’, International Journal of Social Research Methodology. Routledge, 8(1), pp. 19–32. doi: 10.1080/1364557032000119616.

Balci, O. (1994) Validation, Verification, and Testing Techniques Throughout the life Cycle of a Simulation Study, Annals of Operations Research. doi: 10.1007/BF02136828.

Ball, A. (2012) ‘Review of Data Management Lifecycle Models.’ University of Bath, p. 15.

Ballard, M. and Montgomery, P. (2017) ‘Risk of Bias in Overviews of Reviews: a scoping review of methodological guidance and four-item checklist’, Research Synthesis Methods. John Wiley & Sons, Ltd, 8(1), pp. 92–108. doi: 10.1002/jrsm.1229.

Beaujardière, J. D. La (2016) ‘NOAA Environmental Data Management’, Journal of Map & Geography Libraries. Routledge, 12(1), pp. 5–27. doi: 10.1080/15420353.2015.1087446.

Bettany-Saltikov, J. A. (Josette) (2012) How to do a Systematic Literature Review in Nursing: A step-by-step guide. 1st edn. Open University Press.

Booth, A. (2016) ‘Searching for Qualitative Research for Inclusion in Systematic Reviews: A structured Methodological Review’, Systematic Reviews. Systematic Reviews, 5(1), pp. 1–23. doi: 10.1186/s13643-016-0249-x.

Bown, M. J. and Sutton, A. J. (2010) ‘Quality Control in Systematic Reviews and Meta-Analyses’, European Journal of Vascular and Endovascular Surgery. Elsevier Ltd, 40(5), pp. 669–677. doi: 10.1016/j.ejvs.2010.07.011.

Brown, J. and Jones-diette, J. (2014) ‘Developing the Review Question and the Protocol’, 114(4), pp. 1–16.

Brunsson, N., Rasche, A. and Seidl, D. (2012) ‘The Dynamics of Standardization: Three Perspectives on Standards in Organization Studies’, Organization studies. SAGE Publications Sage UK: London, England, 33(5–6), pp. 613–632.

Calder, A. (2008a) ISO/IEC 38500: The IT Governance Standard. IT Governance Ltd.

Calder, A. (2008b) IT Governance: A Pocket Guide. 2019th edn. IT Governance.

Canadian Space Agency (2017) Open Government Implementation Plan : Canadian Space Agency (CSA), Open Information.

Carson, J. S. (2002) ‘Model Verification and Validation’, In Proceedings of The Winter Simulation Conference. IEEE, pp. 52–58.

Chaki, S. (2015) ‘The Lifecycle of Enterprise Information Management’, in Chaki, S. (ed.) Enterprise Information Management in Practice: Managing Data and Leveraging Profits in Today’s Complex Business Environment. Berkeley, CA: Apress, pp. 7–14. doi: 10.1007/978-1-4842-1218-9_2.

Choudhary, R., Perinpanayagam, S. and Butans, E. (2016) ‘Design and Analysis of Communication Model for Implementation of CBM Systems Based on OSA-CBM Framework’, in 2016 IEEE Aerospace Conference, pp. 1–7. doi: 10.1109/AERO.2016.7500821.

Coleman, H. W. and Steele, W. G. (2018) Experimentation, Validation, and Uncertainty Analysis for Engineers. John Wiley & Sons.

DataONE (2017) Data Life Cycle, Best Practices.

DCC (2008) ‘The DCC Curation Life-Cycle Model’.

DCMI (no date a) DCMI: DCMI Metadata Expressed in RDF Schema Language.

DCMI (no date b) DCMI: XML Schemas to Support the Guidelines for Implementing Dublin CoreTM in XML.

Rutter, D. et al. (2015) SCIE systematic research reviews : guidelines. Social Care Institute for Excellence.

Dibsdale, C. (2011) ‘Integrated Vehicle Health Management Operations Rooms’, in Jennions, I. K. (ed.) Integrated Vehicle Health Management: Perspectives on an Emerging Field. SAE International, pp. 113–123.

Evans, D. (2002) ‘Systematic Reviews of Interpretive Research: Interpretive Data Synthesis of Processed Data’, Australian Journal of Advanced Nursing, The. Australian Nursing and Midwifery Federation, 20(2), p. 22.

Faundeen, J. and Hutchison, V. (2017) ‘The Evolution, Approval and Implementation of the U.S. Geological Survey Science Data Lifecycle Model’, Journal of eScience Librarianship, 6(2), p. e1117. doi: 10.7191/jeslib.2017.1117.

Federal Geographic Data Committee - FGDC (2010) Stages of the Geospatial Data Lifecycle pursuant to OMB Circular A–16, sections 8(e)(d), 8(e)(f), and 8(e)(g).

Grant, M. J. and Booth, A. (2009) ‘A typology of reviews: An analysis of 14 review types and associated methodologies’, Health Information and Libraries Journal, 26(2), pp. 91–108. doi: 10.1111/j.1471-1842.2009.00848.x.

Graziani, F. (2008) Computational Methods in Transport: Verification and Validation. Springer.

Gregory, A. (2011) ‘The Data Documentation Initiative (DDI): An Introduction for National Statistical Institutes’. Open Data Foundation, pp. 1–10.

Ten Ham-Baloyi, W. and Jordan, P. (2016) ‘Systematic review as a research method in post-graduate nursing education’, Health SA Gesondheid, 21, pp. 120–128. doi: https://doi.org/10.1016/j.hsag.2015.08.002.

Harrison, K. (2013) A Guide to Data Management in Ecology and Evolution. Edited by K. Harrison. British Ecological Society.

Hedges, L. V and Olkin, I. (2014) Statistical Methods for Meta-Analysis. Academic press.

Hidalga, A. N. de la et al. (2017) ENVRI Reference Model News.

Higginbottom, G. M. A. et al. (2012) ‘Immigrant women’s experiences of maternity-care services in Canada: a protocol for systematic review using a narrative synthesis’, Systematic Reviews, 1(1), p. 27. doi: 10.1186/2046-4053-1-27.

Higgins, J. P. T. and Green, S. (2008) ‘Guide to the Contents of a Cochrane Protocol and Review’, in Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, Ltd, pp. 51–79. doi: 10.1002/9780470712184.ch4.

Higgins JPT Chandler J, Cumpston M, Li T, Page MJ, Welch VA, T. J. (2019) Cochrane Handbook for Systematic Reviews of Interventions. 2nd edn. Edited by T. J. Higgins JPT Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Chichester: John Wiley & Sons.

Hsu, L. et al. (2015) ‘Data Management, Sharing, And Reuse in Experimental Geomorphology: Challenges, Strategies, and Scientific Opportunities’, Geomorphology, 244, pp. 180–189. doi: 10.1016/j.geomorph.2015.03.039.

Hubert Ofner, M. et al. (2013) ‘Management of the master data lifecycle a framework for analysis’, Journal of Enterprise Information Management, 26(4), pp. 472–491.

ICPSR (Inter-university Consortium for Political and Social Research) (2012) Guide to social science data preparation and archiving : best practice through the data life cycle. 5th edn. Michigan. doi: http://dx.doi.org/10.3886/GuideToSocialScienceDataPreparationAndArchiving.

INCOSE (2015) INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities. 4th edn. Edited by W.-E. DaVid d. et al. John Wiley and Sons, Inc.

INCOSE (2017) Guide to the Systems Engineering Body of Knowledge (SEBoK). v1.9 R.D. Edited by R. Adcock. Hoboken, NJ: The Trustees of the Stevens Institute of Technology.

Institution, T. B. S. (2018) BS ISO 30401:2018 Knowledge Management Systems. First. Geneva: BSI Standards Limited.

Jantunen, E. et al. (2017) ‘The Way Cyber Physical Systems Will Revolutionise Maintenance’, in 30th Conference on Condition Monitoring and Diagnostic Engineering Management (COMADEM 2017). 10 to 13, Jul, 2017. Preston, United Kingdom, pp. 145–148.

Johnson, J. A. (2015) ‘Conceptual Models: Core to Good Design’, in Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2493–2494.

Johnson, J. and Henderson, A. (2011) Conceptual Models: Core to Good Design, Synthesis Lectures on Human-Centered Informatics. Morgan & Claypool Publishers.

Kable, A. K., Pich, J. and Maslin-Prothero, S. E. (2012) ‘A structured Approach to Documenting a Search Strategy for Publication: A 12 Step Guideline for Authors’, Nurse Education Today, 32(8), pp. 878–886. doi: https://doi.org/10.1016/j.nedt.2012.02.022.

Kitchenham, B. A., Budgen, D. and Brereton, P. (2015) Evidence-Based Software Engineering and Systematic Reviews. Chapman & Hall/CRC.

Kitchenham, B. and Brereton, P. (2013) ‘A Systematic Review of Systematic Review Process Research in Software Engineering’, Information and Software Technology. Elsevier B.V., 55(12), pp. 2049–2075. doi: 10.1016/j.infsof.2013.07.010.

Kitchenham, B. and Charters, S. (2007) ‘Guidelines for Performing Systematic Literature Reviews in Software Engineering.’, EBSE Technical Report, p. 57. doi: 10.1145/1134285.1134500.

Krogstie, J. (2016a) ‘Introduction to Business Processes and Business Process Modeling’, in Krogstie, J. (ed.) Quality in Business Process Modeling. Cham: Springer International Publishing, pp. 1–51. doi: 10.1007/978-3-319-42512-2_1.

Krogstie, J. (2016b) ‘Quality of Business Process Models’, in Krogstie, J. (ed.) Quality in Business Process Modeling. Cham: Springer International Publishing, pp. 53–102. doi: 10.1007/978-3-319-42512-2_2.

Krogstie, J. (2017) ‘Quality of Conceptual Models in Model Driven Software Engineering’, in Cabot, J. et al. (eds) Conceptual Modeling Perspectives. Cham: Springer International Publishing, pp. 185–198. doi: 10.1007/978-3-319-67271-7_13.

Laboratories, D. (2014) ‘" The Job is Not Finished Until the Customer is Satisfied "’. DSD Laboratories, p. 2.

Lipsey, M. W. and Wilson, D. B. (2001) Practical Meta-Analysis. SAGE publications, Inc.

Löhr, A. and Buderath, M. (2014) ‘Evolving the Data Management Backbone : Binary OSA-CBM and Code Generation for OSA-EAI’, in European Conference of the Prognostics and Health Management Society 2014. PHM, pp. 1–10.

Maindze, A., Skaf, Z. and Jennions, I. (2019) ‘Towards an Enhanced Data- and Knowledge Management Capability : A Data Life Cycle Model Proposition for Integrated Vehicle Health Management.’ in Proceedings of the Annual Conference of the PHM Society, pp. 1–14. doi: 10.36001/phmconf.2019.v11i1.842.

Matthews, B. et al. (2010) ‘Using a Core Scientific Metadata Model in Large-Scale Facilities’, International Journal of Digital Curation, 5(1), pp. 106–118. doi: 10.2218/ijdc.v5i1.146.

McKenzie, F. D. (2010) ‘Systems Modeling: Analysis and Operations Research’, in Sokolowski, J. A. and Banks, C. M. (eds) Modeling and Simulation Fundamental: Theoretical Underpinnings and Practical Domains. John Wiley & Sons Hoboken, NJ, USA, pp. 147–180.

Parry, M. J. (2016) ‘Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections’, The Electronic Library. Emerald Group Publishing Limited, 34(6), p. 1055.

Plale, B. and Kouper, I. (2017) ‘The Centrality of Data: Data Lifecycle and Data Pipelines’, in Data Analytics for Intelligent Transportation Systems. Elsevier, pp. 91–111. doi: https://doi.org/10.1016/B978-0-12-809715-1.00004-3.

Popay, J. et al. (2019) ‘Guidance on the Conduct of Narrative Synthesis in Systematic Reviews’, A Product from the ESRC Methods Programme, pp. 211–219. doi: 10.1111/j.1523-536x.1995tb00261.x.

Porcal-Gonzalo, M. C. (2015) ‘A Strategy for the Management, Preservation, and Reutilization of Geographical Information Based on the Lifecycle of Geospatial Data: An Assessment and a Proposal Based on Experiences from Spain and Europe’, Journal of Map & Geography Libraries. Routledge, 11(3), pp. 289–329. doi: 10.1080/15420353.2015.1064054.

Pouchard, L. (2015) ‘Revisiting the Data Lifecycle with Big Data Curation’, International Journal of Digital Curation, 10(2), pp. 176–192. doi: 10.2218/ijdc.v10i2.342.

Ransom, E. et al. (2017) ‘Science and standards’, The Routledge Handbook of the Political Economy of Science. New York: Routledge, pp. 329–340.

Reeves, M. G. and Bowen, R. (2013) ‘Developing a data governance model in health care: although the term may be unfamiliar, data governance is a longstanding obligation of the healthcare industry’, Healthcare Financial Management. Healthcare Financial Management Association, 67(2), pp. 82–87.

Rodgers, M. et al. (2009) ‘Testing methodological guidance on the conduct of narrative synthesis in systematic reviews: Effectiveness of interventions to promote smoke alarm ownership and function’, Evaluation, 15(1), pp. 49–73. doi: 10.1177/1356389008097871.

Rosenbaum, S. (2010) ‘Data Governance and Stewardship: Designing Data Stewardship Entities and advancing Data Access’, Health services research. Wiley Online Library, 45(5p2), pp. 1442–1455.

Rosenthal, R. and DiMatteo, M. R. (2002) ‘Meta-analysis’, in Pashler, H. and Wixted, J. (eds) Stevens’ handbook of experimental psychology. 3rd edn. John Wiley & Sons, Inc., pp. 391–428.

Sabharwal, A. (2015) Digital curation in the digital humanities: Preserving and promoting archival and special collections. Chandos Publishing.

Simonet, A., Fedak, G. and Ripeanu, M. (2015) ‘Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures’, Future Generation Computer Systems, 53(Supplement C), pp. 25–42. doi: https://doi.org/10.1016/j.future.2015.05.015.

Sornette, D. et al. (2007) ‘Algorithm for model validation: Theory and applications’, in Proceedings of the National Academy of Sciences, pp. 6562 LP – 6567. doi: 10.1073/pnas.0611677104.

Staples, M. and Niazi, M. (2007) ‘Experiences using systematic review guidelines’, Journal of Systems and Software, 80(9), pp. 1425–1437. doi: 10.1016/j.jss.2006.09.046.

Sterne, J. A. C., Egger, M. and Moher, D. (2008) ‘Addressing Reporting Biases’, in Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, Ltd, pp. 297–333. doi: 10.1002/9780470712184.ch10.

The Joanna Briggs Institute (2014) Joanna Briggs Institute Reviewers’ Manual: 2014 edition. The Joanna Briggs Institute.

Tolk, A. and Aaron, R. D. (2010) ‘Addressing Challenges of Transferring Explicit Knowledge, Information, and Data in Large Heterogeneous Organizations: A Case Example from a Data-Rich Integration Project at the U.S. Army Test and Evaluation Command’, Engineering Management Journal, 22(2), pp. 44–55.

UCSD (University of San Diego) (2018) Data Management.

University of Virginia (2017) Steps in the Data Life Cycle.

Xu, J. and Xu, L. (2017) ‘Chapter One - ISHM for Complex Systems BT - Integrated System Health Management’, in Integrated System Health Management Perspectives on Systems Engineering Techniques. Academic Press, pp. 1–53. doi: https://doi.org/10.1016/B978-0-12-812207-5.00001-8.

Yang, E., Matthews, B. and Wilson, M. (2013) ‘Enhancing the core scientific metadata model to incorporate derived data’, Future Generation Computer Systems, 29(2), pp. 612–623. doi: 10.1016/j.future.2011.08.003.

Yin, R. K. (2018) Case Study Research and Applications: Design and Methods. 6th edn. Sage publications.

Young, I. and Waddell, L. (2016) ‘Barriers and Facilitators to Safe Food Handling among Consumers: A Systematic Review and Thematic Synthesis of Qualitative Research Studies’, PLoS ONE. Edited by G.-J. Nychas. San Francisco, CA USA: Public Library of Science, 11(12), p. e0167695. doi: 10.1371/journal.pone.0167695.

Yu, X. and Wen, Q. (2010) ‘A View about Cloud Data Security from Data Life Cycle’, 2010 International Conference on Computational Intelligence and Software Engineering, (4072020), pp. 1–4. doi: 10.1109/CISE.2010.5676895.

CHAPTER 5: TOWARDS AN ENHANCED DATA- AND KNOWLEDGE MANAGEMENT CAPABILITY: A DATA LIFE CYCLE MODEL PROPOSITION FOR INTEGRATED VEHICLE HEALTH MANAGEMENT.

(This chapter is also a reformatted paper published at “*Annual Conference of the Prognostic Health Management Society (PHM) Society, 11*”)

Abstract: The creation, capturing, using and sharing of knowledge is based on data. The rate of data creation, collection, and elicitation through wide-ranging experiments, simulations, observations and measurements is rapidly increasing within Integrated Vehicle Health Management (IVHM). In addition, Knowledge Management (KM), data abstraction, analyses, storage and accessibility challenges persist, resulting in loss of knowledge and increased costs. This growth in the creation of research data, algorithms, technical papers, reports and logs, requires both a strategy and tool to address these challenges. A Data Life Cycle Model (DLCM) ensures the efficient and effective abstraction and management of both data and knowledge outputs. IVHM is characterized by prognostics and diagnostics, which depend heavily on high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health. IVHM does not yet have a systematic and coherent approach to its data management. The absence of a DLCM means that valuable knowledge might be lost or is difficult to find. Data visualization is fragmented and done on a project by project basis leading to increased costs. There is insufficient algorithm documentation and communication for easy transition between subsequent researchers and personnel. A systematic review of DLCMs, frameworks, standards and process models pertaining to data- and KM in the context of IVHM, found that there is no DLCM that is consistent with IVHM data and knowledge management requirements. Specifically, there is a need to develop a DLCM based on Open System Architecture for Condition-Based Maintenance (OSA-CBM) framework.

5.1 Introduction

There has been a steady growth in both scope of research as well as data extraction activities in IVHM depicted in Figure 5.1 below. This growth has been matched by the complexity of management and organization. With this growth in the creation of research data, data automating algorithms, technical papers, reports and theses, IVHM needs both a data management model and a Knowledge Management System that facilitate the storing, organizing and sharing of its research and knowledge outputs. Such a model needs to be secure and scalable with a high level of cross-platform or domain transferability. The absence of a suitable data life cycle model means weak data- and knowledge management for IVHM. This leads to increased costs and loss of valuable knowledge, thereby creating long-term uncertainties for diagnostic and prognostic management. The primary value for designing and implementing a data life cycle model includes the following:

- Enhanced and integrated requirements gathering for IVHM and IVHM data and Knowledge Management Systems
- Increased efficiency and effectiveness of planning and handling the growing volumes, diversity and complexity of data and data management
- Facilitate the design and development systems for high operational efficiency
- Making raw and derived data accessible to IVHM researchers and engineering operations
- Ensures the provision of secure, high quality, accurate and consistent asset data throughout its entire life-cycle
- Facilitates the retention of provenance data
- Ensuring timely, comprehensive, and secure approaches to data curation (Faundeen *et al.*, 2013)

In this paper, we carry out a non-exhaustive integrated and systematic review. We finally propose an integration of relevant elements of the Open System Architecture for Condition-Based Maintenance (OSA-CBM) framework, the International Council on Systems Engineering (INCOSE) process, the Core Scientific Metadata (CSMD) and United States Geological Survey (USGS) data life cycle models. The aim is to create a Data life cycle model suitable for the management of IVHM data, knowledge outputs, depth and breadth IVHM operations, IVHM research and IVHM system requirements.

The basic foundation of Integrated Vehicle Health Management (IVHM) consists of sensing, instrumentation and signal processing. This leads to the extraction of data and features selection, paving the way for prognostics prediction algorithms. Precision and well-timed availability of the above instruments are fundamental to maintenance scheduling (Perinpanayagam, 2013). The first stage of the IVHM cycle is the collection of data using simulations, observations, derivation, experiments and referencing (SODER) methods (Figure 5.1) about an asset. IVHM delivers value to stakeholders and reduces the cost of delivery by monitoring the health of an asset and making decisions based on the data collected (Jennions, 2011). Consequently, IVHM relies predominantly on the availability of high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health.

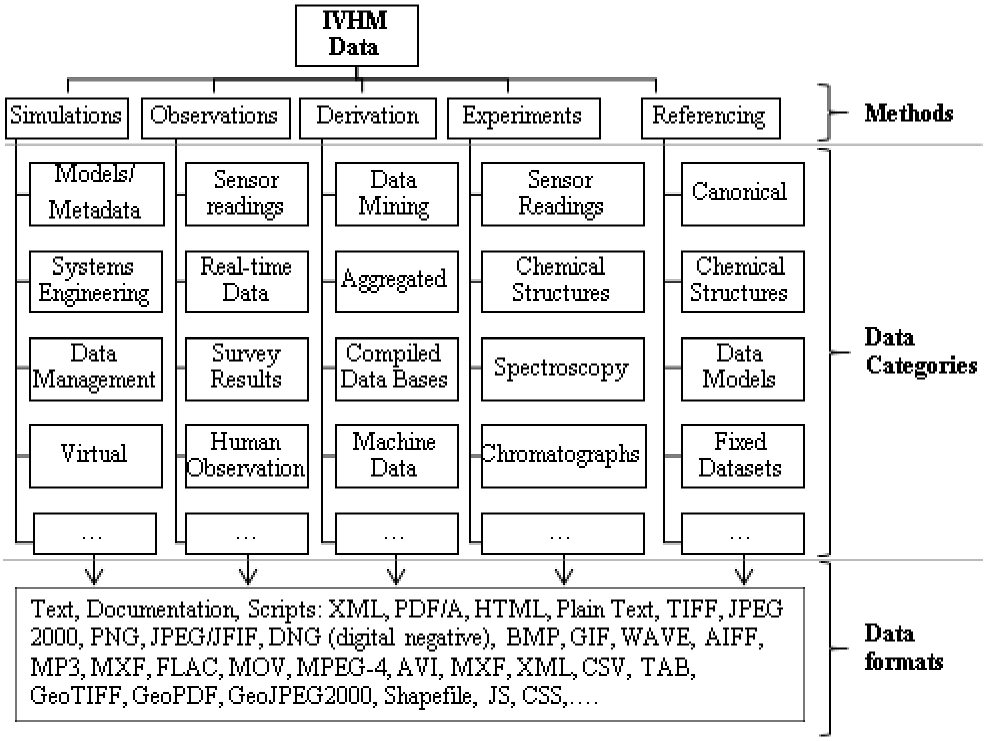


Figure 5-1 Data creation methods, data categories and file formats

The data has to be accurate, complete, timely, context-relevant, reliable and explicit (Dibsdale, 2011). The absence of any or all of these qualities in any dataset creates uncertainties and increases the probability of misdiagnosis, modelling errors (Arahchige and Perinpanayagam, 2017) and inaccurate predictions. This essentially indicates the malfunction of the IVHM system because it is "*the assembly of data related to the current and future activities of a critical system and transforms these data into the information and which is applied to make a functional decision*" (Prajapati, Roy and Prasad, 2018). A data life cycle model can be defined as a "*…a formal representation of all the possible states and all the valid state transitions of a data item, when handled by a particular system or by a user application, e.g. created, duplicated, deleted, backed-up"* (Simonet, Fedak and Ripeanu, 2015). It represents the requisite actions, operations, or processes to be taken at various stages of data creation and management (Faundeen *et al.*, 2013). Data management through its entire life cycle still presents a number of complex challenges relating to interoperability, volume, storage, data citation, and metadata standards and data provenance (Yang, Matthews and Wilson, 2013; Porcal-Gonzalo, 2015; Beaujardière, 2016) This perhaps explains the proliferation of discipline or domain-specific data life cycle models.

Various disciplines and organizations are creating standardized frameworks, data ontologies, standards and unique data life cycle models to suit their respective requirements. Metadata standards like Dublin core (Hsu *et al.*, 2015), Core Scientific Metadata (CSMD) provisions the basic metadata required to enhance the search functionalities over data portals and knowledge libraries (Matthews *et al.*, 2010) but falls short of propagating the complete provenance data. The shortfall with the CSMD is that neither does it support for "*access to the derived data produced during analysis, nor does it allow the provenance of data supporting the final publication to be traced through the stages of analysis to the raw data*" as pointed out by Yang et al.(2013).

However, though metadata standards are relevant for all data and knowledge outputs and organizations, each organization or project seems to have separate requirements for their own research or projects data. This is reflected in the 17 data life cycle models identified, which all highlight the significance of metadata and standards. The models seem to differ on the depth and breadth of applicability and priority of requirements. For instance, the USGS emphasized three critical cross-cutting activities namely; metadata description, quality assurance and protection from corruption or loss to be performed parallel to planning, acquisition, processing, analysis, preservation, publishing and sharing to achieve enhanced quality, understanding and long-term reuse (Faundeen and Hutchison, 2017; Plale and Kouper, 2017) The Digital Curation Centre (DCC) model, on the other, hand suggests metadata should comprise rules and formalized entities for automatable tools and services as well as the role of data managers and data curators in the improvement of knowledge (Plale and Kouper, 2017).

Taking the DataOne data life cycle model to illustrate the variance, though it represents all the classical stages of the data management life cycle, all parts of the life cycle are not mandatory (Harrison, 2013; Allard, 2014; Pouchard, 2015; DataONE, 2017; Hidalga *et al.*, 2017; Plale and Kouper, 2017). This implies that the number of selected stages are dependent on the type of project or project requirements. Furthermore, although the USGS addresses the weakness in most data life cycle models identified, it can be used in diverse settings despite having been developed mainly for USGS science data. The requirements on which this is developed are not consistent with small research facilities like the IVHM Centre Labs or the University as a whole. None of the models identified seems to satisfy the scenario of IVHM research, operations and systems.

The core of IVHM is the capture of data and analysis to establish an early detection of anomalies and an advance indication of future failures, state or distinctive characteristics of current assets based on their current state. The conduct of prognostics and diagnostics is reliant on high-quality reliable data. This quality assurance can be determined by the data management life cycle. Data life cycle models are often integrated with software services and policies (Plale and Kouper, 2017). It is therefore fundamental to consider associated processes involved in the planning, designing and developing of the software services and policies.

The INCOSE system engineering process [V-Model] is suitable for the development of any system. It is vital in the scenario of IVHM data life cycle model and Knowledge Management System. This is suitable for the IVHM data life cycle and system development because of the restricted duration and scope of this research project. The V-model is best suited for these kinds of projects that have well-defined length and scope, consistent technology, and a clear and well documented technical and functional specifications (INCOSE, 2017).

The next section below details the review and analyses leading to the creation of the IVHM Data Life Cycle Model.

5.2 Methodology: The Systematic Review

Following a scoping study, the search string depicted in Figure 5.2 was implemented. As a consequence of the diverse and qualitative studies reviewed, a narrative synthesis was applied. A narrative synthesis is "*an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarise and explain the findings of the synthesis.*" (Ten Ham-Baloyi and Jordan, 2016) This study identified 19 data life cycle models with varying degrees of complexity, composition and depth (Figure 5.3). According to Pouchard [2015, p.180] "*Data life cycle models present a structure for organizing the tasks and activities related to the management of data within a project or an organization.* " Data lifecycle models represent a description of "*data objects through a set of time-order"* (Plale and Kouper, 2017) The data life cycle models vary in steps or phases from one organization to another.

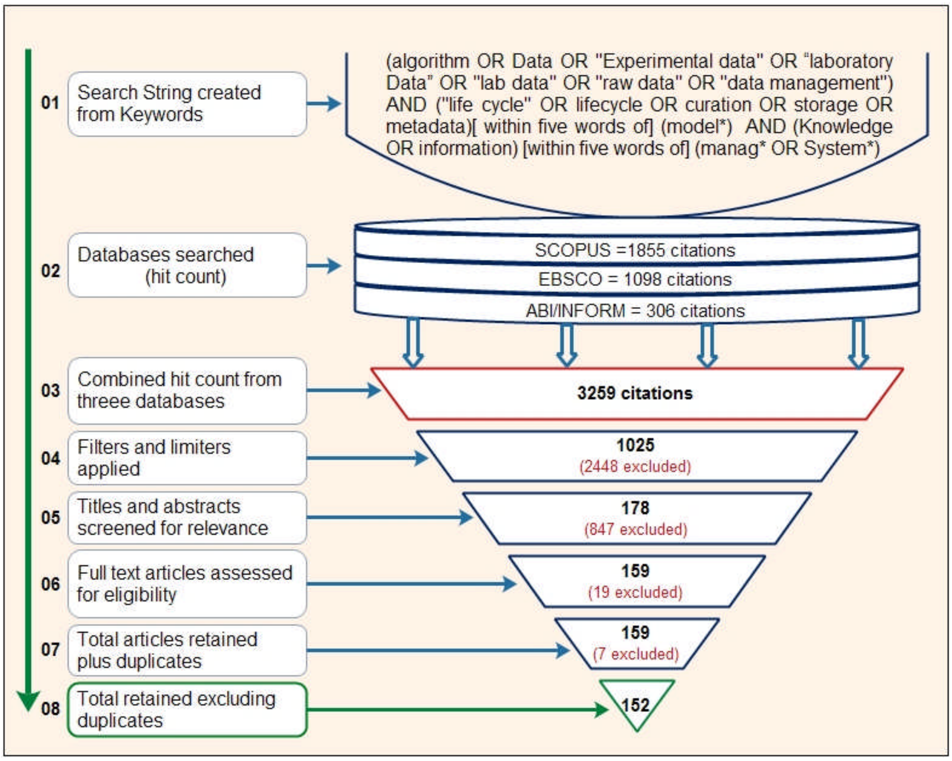


Figure 5-2 Scoping study and search string implementation

The longest model has nine phases, and the shortest has five phases. However, these data life cycles are not necessarily cyclical but rather functional as some phases run parallel across the entire life cycle.

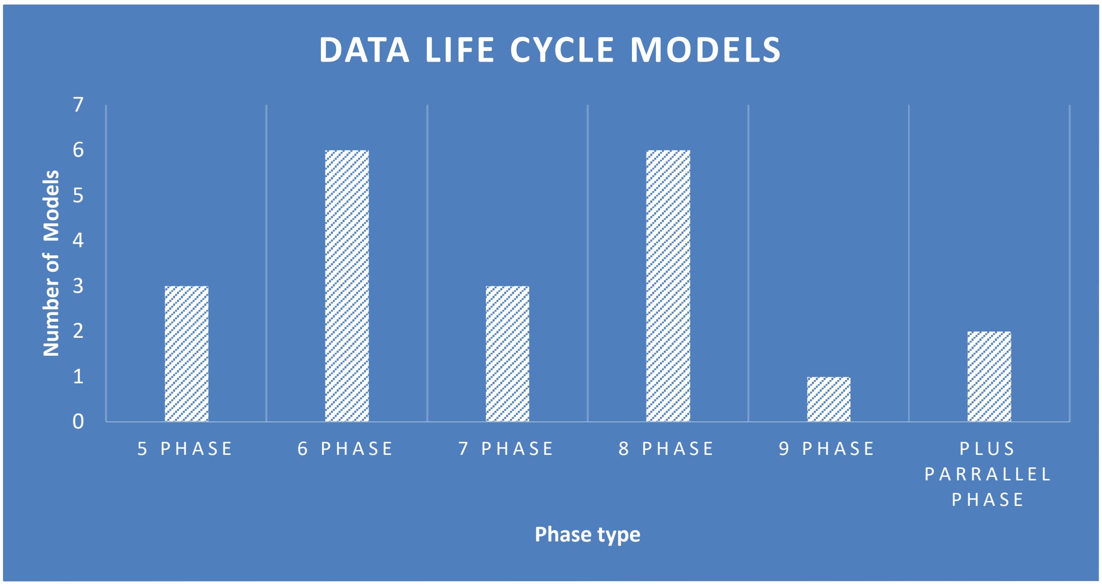


Figure 5-3 Data Life Cycle Models by Phases

The data life cycle models were then grouped by their respective numbers of data life cycles, i.e. five-phase, six-phase, seven-phase, eight-phase and nine-phase models. The review revealed nine standards and frameworks relating to data life cycles. In the reviewed studies, the data life cycle has been described as the set of activities that affect the short and the long-term preservation of datasets through a system from planning, creation, maintenance, re-use and purging (Simonet, Fedak and Ripeanu, 2015; Beaujardière, 2016)

5.2.1. Five-phase data lifecycle models

Three five-phase models were identified; CRUD (Create, Read, Update and Delete), Enterprise Data life cycle model and Michigan State University (MSU) Records life cycle. These models make use of the classical elements of the data life cycle. The

CRUD model provides a flexible data life cycle; it only obliges creation, storage and destruction while living use, sharing and archiving optional. The Enterprise Data life cycle model is a closed life cycle, and its emphasis on destruction limits the availability of data in the long-term (Arass, Tikito and Souissi, 2017). The Michigan State University (MSU) life does not feature sharing, classification and analysis.

5.2.2 Six-phase data lifecycle models

Six data life cycle models with six phases were identified; (i) United States Geological Survey (USGS) Science Data Lifecycle Model, (ii) University of Virginia, Steps in the Data Life Cycle, (iii) International Leader in Data Stewardship (ICPSR) Data Lifecycle, (iv) UCSD Libraries Data Lifecycle, (v) Generic Lifecycle Model, and UK Data Archive Data Lifecycle.

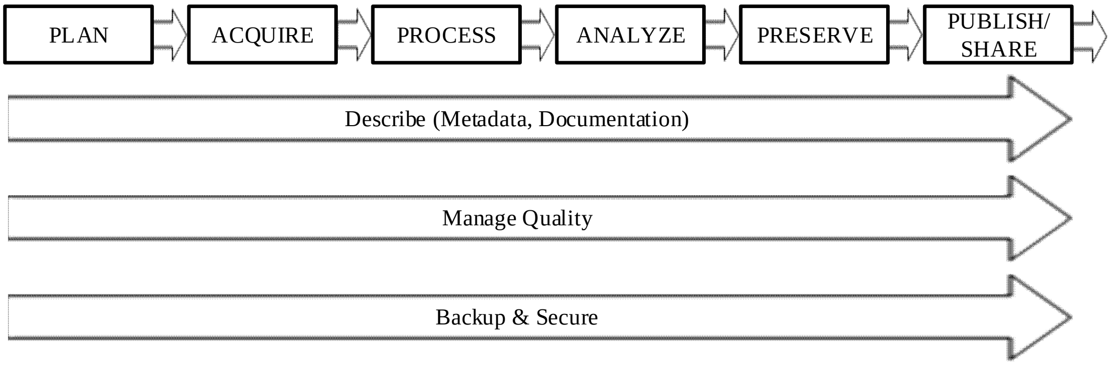


Figure 5-4 USGS Science Data Lifecycle Model (Faundeen *et al.*, 2013)

According to Faundeen and Hutchison (2017), data lifecycle models are fundamental to communication and data management and ensures adequate long-term preservation and accessibility. After reviewing more than 50 data lifecycle models [the latter four above inclusive], they came to the conclusion that none of the existing models was entirely consistent with the USGS data management requirements. It was imperative to USGS that their functional processes and workflows were adequately captured in a model.

Furthermore, like other organizations, the USGS developed its own data management lifecycle with the aim of reducing complexity and removing redundant or irrelevant steps or phases that were not in sync with their scientific workflows and processes. The USGS opted for a linear and easily operated illustration of their new model (Figure 5.4). The model (Faundeen *et al.*, 2013; Faundeen and Hutchison, 2017) included the basic classical data lifecycle phases and laid emphasis on three parallel phases; metadata, quality management, backup and Security.

i. Plan: The organization should identify the resources, methods, techniques, functional and technical system requirements and generate both plans for either data acquisition, data entry or signal reception and data management.
ii. Acquire: This is the data capture phase which can either data acquisition, data entry, signal reception or all three activities combined.
iii. Process: Raw as well as derived data verification, organization, transformation, integration, and extraction in appropriate format takes place in this step.
iv. Analyse: This encompasses demonstrable quality requirements fulfilment, data analytics, modelling and evaluation test results as well as methods and activities carried out to facilitate definitions of facts, identification of forms and trends, developing interpretations and testing hypotheses.
v. Preserve: Data storage for Long-term access and reuse. The purpose of this phase is the guarantee long-term preservation, ease of search and retrieval, accessibility and usability of the data. This step employs multi-copy/storage locations, long-term usefulness, accuracy and consistency, information security, metadata and file formats.
vi. Publish/Share: Put together quality-assured, metadata rich, platform or system-agnostic data, with relevant security safeguards and share with interested parties of stakeholders.
vii. [Parallel to phases 1-6] Describe (Metadata, Documentation): Establishes an obligation to create and upgrade metadata on any or all the stages of the lifecycle including the documentation of usage in specific systems, applications and settings.

viii. [Parallel to phases 1-6] Manage Quality: Mandatory to accurately undertake data collection, handling, processing, usage, and maintenance across all the phases of the scientific data lifecycle, is the use of protocols and methods. This implies effective and efficient quality assurance and quality control.

[Parallel to phases 1-6] Back-Up and Secure: Regularly create image backups of both files and databases on either onsite or offsite devices. Access control and other security measures must be taken to prevent accidental data loss and data corruption.

The USGS data lifecycle model encapsulates the activities and steps in the latter four models. However, different research activities or projects will use some or all elements of the data lifecycle in dissimilar ways. The data lifecycle management is influenced by the requirements of a particular project or organization (Faundeen and Hutchison, 2017) .

5.2.3 Seven-phase data lifecycle models

The key feature of these models is the inclusion of discovery, knowledge repository and reuse. There is a noticeable absence of access control and security element. The Geospatial Data Lifecycle is a flexible non-linear and non-sequential comprising the classical Define, Evaluate, Obtain, Access, Maintain, Use/Evaluate and Archive stages but lacks the discovery and reusability elements. These may not be weaknesses in themselves as each life cycle model has been designed to cater for to the needs of their respective organizations. As noted in Yu and Wen (2010) not all elements of the data lifecycle are applicable to all contexts and scenarios.

5.2.4 Eight-phase data lifecycle models

The DataOne data lifecycle model is one of eight phase data lifecycle models designed by the National Science Foundation (NSF) exclusively to focus on the phases field, or laboratory data goes through rather than the role of the person on data (Allard, 2014). The model is developed as a foundation for the development of tools and services for data-intensive sciences and for encouraging best practices

(Pouchard, 2015; Arass, Tikito and Souissi, 2017; DataONE, 2017; Plale and Kouper, 2017).

The experimental geomorphology data life cycle model addresses the complexity in analysis, storage and search and retrieval posed by increasing volumes of laboratory data (Hsu *et al.*, 2015). Hsu et al., (2015) review data management practices and challenges for experimental geomorphology, established that the lack of; rules for metadata integration, documentation of workflow, data storage, motivation and training were impeding significant amounts of data sharing and re-use. These challenges mean that accessibility or availability of experimental or laboratory data is limited to the research or project lifecycles, and therefore undermining reproducibility and quality control. They suggested that efficient and effective data management and sharing should consider the entire data lifecycle, including metadata set information, disciplinary information, Quality information and readiness for reuse.

The Data Documentation Initiative (DDI) version 3.0 mostly linear combined lifecycle model integrates data application perspective and social science research data. Metadata requirements are grouped into five comprising conception of the study, collecting of data, logical data encoding structure, physical data encoding structure and archiving (Ball, 2012).

The DCC curation lifecycle model (DCC, 2008; Ball, 2012; Faundeen and Hutchison, 2017) is the model with the highest number of phases with emphasis on preservation, archiving and management of data and publications for long-term availability and reuse. It concentrates on data curation and serves as a planning tool for data creators, curators and users. The model features a complete lifecycle action, sequential actions and occasional situational actions (Plale and Kouper, 2017). The model presents indispensable phases in the life cycle by using circles with a common centre. The phases include: These are: (i) Conceptualize (Conceive & Plan); (ii) Create or Receive (Generate metadata); (iii) Appraise and select (Quality and

Governance); (iv) Ingest (Move to Storage or Archive); (v) Preservation action (Apply retention schedules); (vi) Store (Secure Storage); (vii) Access, use and reuse (Access policy and control); (viii) Transform (Migration), plus (ix) Sporadic Actions: Migrate, Dispose, Reappraise and Dispose.

The DCC lifecycle model provides an applied framework that can be loosely categorized into different levels and areas of curation such as technology level operations, bit-level preservation routines and metadata curation. The technology level operations involve migration, backup, indexing, and system upgrades. Bit-Level preservation routines include data-recording, checksum reporting, error-correcting and character replacement and file format registries. The metadata curation level defines the content and context of digital elements. This level fulfils the technical, descriptive, structural and preservation metadata requirements (Sabharwal, 2015; Parry, 2016). One of many technical standards for metadata is the Dublin Core. Activities that generate rich footnotes and significance to images in Digital humanities scholarship and teaching practices are given three levels of Curation.

Level one (L1) focuses on digital files and technology used in their preservation while levels two and three denotes the exercise of intellectual control and scholarly processing respectively. Furthermore, underlying these levels are aspects of metadata schema such as Dublin Core, interoperability standards like the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), file format and data encoding, network characteristics and reliable hardware and software systems (Sabharwal, 2015).

5.2.5 Standards and frameworks

In addition to some of the standard mentioned above, Figure 5.5 below shows four main standards commonly used to support data management through the data life cycle. The design of data lifecycle models takes into consideration relevant and appropriate standards/frameworks to support respective domain data. Some relevant standards identified in the literature included the Open System Architecture for Condition Based Maintenance (OSA-CBM), the Core Scientific Metadata Model (CSMD), INCOSE systems engineering Management process (the V-model) and the National Oceanic and Atmospheric Administration (NOAA) Environmental Data Management (EDM) Framework. Each plays an important role in the data management lifecycle.

The Open System Architecture for Condition Based Maintenance (OSA-CBM) is an ISO-13374 compliant (Felke *et al.*, 2010; Jantunen *et al.*, 2017) "*standard architecture for moving information in a condition-based maintenance system*"(MIMOSA, 2018). The ISO 13374 provides the standards for condition monitoring and diagnostics of machines and machine systems – Data processing, communication and presentation. The OSA-CBM framework (Figure 5.5) facilitates the integration different software and hardware components in order to decrease costs, enhance interoperability, boost competition, merge design changes, and stimulate collaboration in condition-based maintenance (Sreenuch, Tsourdos and Jennions, 2013; Löhr and Buderath, 2014). According to (Choudhary, Perinpanayagam and Butans, 2016) OSA-CBM is fundamental in the creation data identities ['data CVs'] in the form of metadata that includes attributes such as id, site, time, alert, algorithm type, name, description and others.

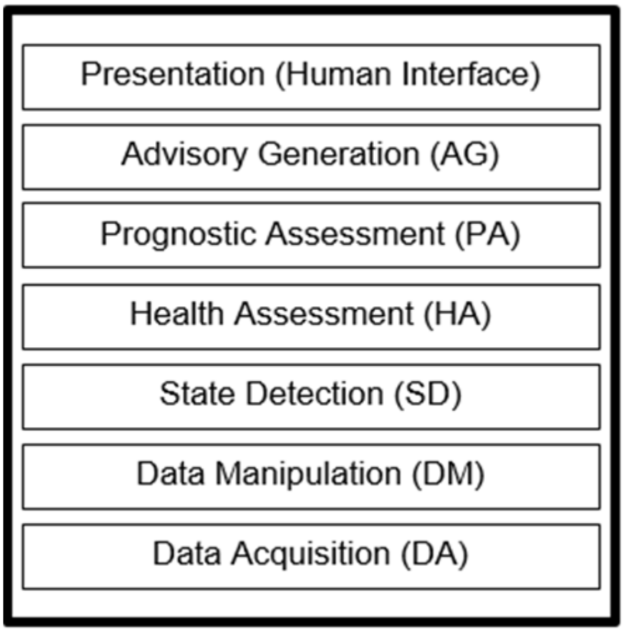


Figure 5-5 OSA-CBM Functional Blocks (Lebold *et al.*, 2002; Redding, 2011; Jantunen, Junnola and Gorostegui, 2017; MIMOSA, 2018)

Its functional capability includes human-machine interface provisioning information accessibility. Its focus on interoperability and metadata makes its relevance in the data lifecycle models extremely pertinent. Figure 5.5 above displays the basic Architecture of OSA-CBM and the following paragraphs explain the data flow between the layers;

Data Acquisition (DA): This block uses either transducers or sensors to pick up physical manifestations and convert to clean digital signals for computerization to extract relevant information. The DA block is essentially a server of cleaned digital signal data.

Data Manipulation (DM): The DM block processes raw data from the DA block by means of mathematical algorithms, generating computed and the virtual sensor readings which are stored in a database.

State Detection (SD): In this block, the resulting data from the DA and DM blocks are compared to known profiles for any discrepancies, and if so, identify the profile associated with the data.

Health Assessment (HA): While taking possible faults into consideration, historical health trends, functional status and load history are analysed to diagnose faults and current health situation.

Prognostic Assessment (PA): Using current data and projected usage, current health conditions of assets and the remaining useful life (RUL) are forecast using either model-based (physical phenomena of degradation), data-driven (pattern recognition and machine learning algorithms) or hybrid approach (a combination of statistical data and physical phenomena) to get the best outcome.

Advisory Generation (AG): This is the decision support layer that provides recommendations on steps and movements that need to be undertaken to optimize the useful health of the system in consideration.

Presentation Block (Human Interface): The presentation layer displays health valuations, prognostic valuations, or decision support recommendations and alerts.

The OSA-CBM framework provides the parameters for systems architecture required for the successful application of Integrated Vehicle Health Management (IVHM) to a product (Redding, 2011). Thus, the structure of OSA-CBM architecture provisions the foundation for evaluating important IVHM technologies and database standards (Goebel, 2011). This is significant because IVHM is a data-driven and data acquired from transducers and sensors directs much of the thinking surrounding it. In addition data management, data integrity, data quality are imperative for features or faulty conditions extraction-fundamental in condition-based maintenance (CBM) and Prognostic health management (Dibsdale, 2011; Goebel, 2011) IVHM systems require the capability to organize and manage small as well as large data sets in linked tables to promote the easeful appreciation and deliver a comprehensive language for data definition, retrieval, and update. Therefore, making data management an imperative competence in the operations room in particular and IVHM in general (Dibsdale, 2011).

Unlike OSA-CBM the Core Scientific Metadata (CSMD) model (Figure 5.6) is designed for use in large scale laboratory-centered scientific facilities to represent data acquired from scientific experiments or structural sciences (Yang, Matthews and Wilson, 2013). The Science and Technology Facilities Council (STFC) generated the model. It is designed around the hierarchical concept of scientific studies and investigation which are usually characterized by experiments, measurements, simulations, modelling and observations. The outcomes typically include three phases – Phase One being the acquisition of raw data which is analysed in Phase Two to create derived data which is eventually published in Phase Three. The Core Scientific Metadata (CSMD) model mainly involves experimental data acquisition and partially automated creation of metadata. The CSMD model supports interoperability of data management and accessibility, and facilitates cataloguing, data curation and data reuse for medium and to long-term (Matthews *et al.*, 2010).

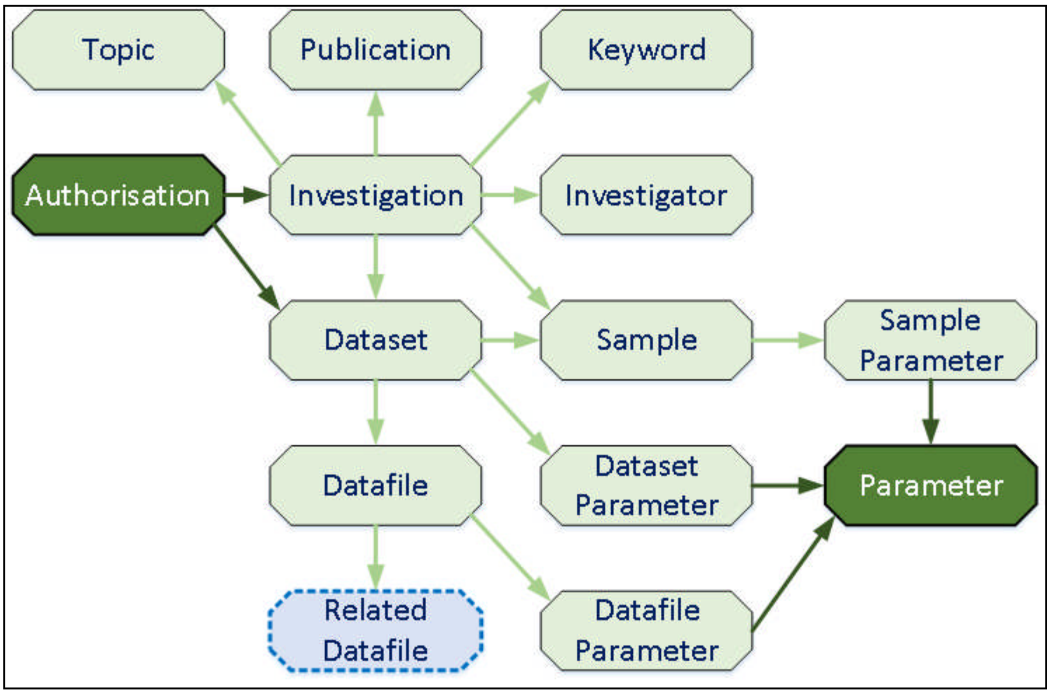


Figure 5-6 The CSMD Model (Matthews *et al.*, 2010)

The main elements of the CSMD model (Matthews *et al.*, 2010) include:

Investigation: This is the most important entity of the study that specifies title, abstract, dates, data collection tools, facility and the unique identifiers referencing the particular model.

Investigator: Stakeholders of the study; Main researcher, support researcher, sponsors, institutions and their roles.

Topic and keyword: This include managed and unmanaged terms used in interpreting and cataloguing the research.

Publication: Assigns references to publications linked with the research.

Sample: Detailed data on study research sample. Unique details such as precautions on the toxicity of elements, and attributes relating to tagging of samples, substance annotation are captured in the model.

Dataset: Research projects can include more than one dataset on which diverse and multiple samples are analysed. Research activity can include raw datasets on which analysed datasets are subsequently inserted.

Datafiles: In the CSMD, this takes the form of physical data objects stored on physical storage disk drives (Yang, Matthews and Wilson, 2013). Its metadata includes name, version, location, data format, time created, modified by and time modified, and other details such as Check-sum.

Parameter: Defines explicit aggregates like pressure, temperature, volume or scattering angle connected to the research. These aggregates can be used to describe sample parameters, study environment or the variables being measured. Parameters can be linked at various levels with datasets and metadata elements.

Authorization: The CSMD model can specify access controls on investigation and datasets.

In a study of the problem of managing provenance of derived data in scientific research, Yang et al (2013) found that although the initial CSMD model provisioned accessibility, usage and reuse of experimental raw data, it did not "*support access* *to the derived data produced during analysis, nor does it allow the provenance of data supporting the final publication to be traced through the stages of analysis to the raw data*" (Parry, 2016). In other words, the original CSMD model recognizes the sources' provenance of the derivative data but fails to describe the transformation provenance. They emphasized the significance of keeping track of previous work and the need for a resilient data management tool and computational workflows that would capture the flow of data, raw data to derived data through to final publication.

They recommended that data trails generated during analysis should be captured to ensure reliable reproducibility research outcomes – an essential element in valid scientific research. Because much of the data in scientific facilities is generated in large volumes and at significant costs, repetition of data collection is not a viable or preferred option and therefore any bid to replicate results. Therefore, replication of test results would be best achieved through the re-analysis of already collected raw data.

To demonstrate the validity of their proposition, they extended the CSMD model to account for derived data and to record the data analysis process enough for each of their use case. They extended the CSMD model to a software agnostic one that contains resultant data product to include a description of transformation provenance which is not covered in the existing model. They carried a pilot implementation with experimental scientists, with annotations employing the ICAT data catalogue scheme. They identified five fundamental factors for capturing provenance data.

They include the following:

- the *data objects* involved
- the *programs* that produce or consume data objects
- the *ordering* of the programs
- the *parameters* to the programs; and

- the *people*: This refers to those who drive the programs and therefore fundamental for accountability, security, attribution and archival processes. This element is excluded from the extended model

They proposed six items as an extension to the CSMD model as follows:

1. Adding a *Software Execution* subclass of investigation: The subclass is for modelling the executions of one data analysis task in the process.
2. Linking a program to a *software execution*: This is a runtime notion associated a single software program, one or more data files and zero to any number of parameters that drive the program and the outputs from the data files and parameters. The persistent and catalogued aspects of derived data provenance trail are often determined by researchers.
3. Linking software executions with datasets: Software executions are linked to input datasets and output datasets from an execution of a program and associating multiple software executions to input datasets extended to include many to many relationships between investigator and dataset, to capture their context in the provenance process.
4. Associating parameters with a software execution: Parameters must be linked with no less than one of many possible programs [can take zero or more parameters] executions corresponding to unique datasets but with assorted output datasets and runtime parameters.
5. Re-introducing the study: A study represents an amalgamation of associated investigations and a channel for relating Software Executions to each other and other types of investigations.
6. Introducing study nesting: Nesting an investigation inside one or more larger ones.

In conclusion, Yang et al., (2013) all observed that their proposed extended model was domain agnostic though developed for solving structural science data management problems from large scale facilities, it can be used to resolve issues of derived data management in many disciplines as well as in limited size scenarios like university research laboratories [such as the IVHM lab].

The extended model captures both the data source and its provenance – the transformation it has gone through in the lifecycle. The ICAT prototype does not yet allow for the propagation of the complete provenance of output data without unpacking the datasets to facilitate the querying of used transformations for researchers and neither does the extended proposed CSMD model. In addition, there is still scope for improving the software and hardware environment which have equally not been covered in this proposition.

International Council on Systems Engineering (INCOSE) Systems Engineering V-Model [Verification and Validation model] (Figure 5.7) is a process for ensuring effective and efficient satisfaction, high quality, trust worthy, cost efficient and schedule customer as well as stakeholder requirements throughout a systems lifecycle. It illustrates the product lifecycle from foundation to obsolescence or destruction (INCOSE, 2015a). The V-model lifecycle is a waterfall-like step-by-step process implementation that ensures predictability, stability, reproducibility, and substantial surety.

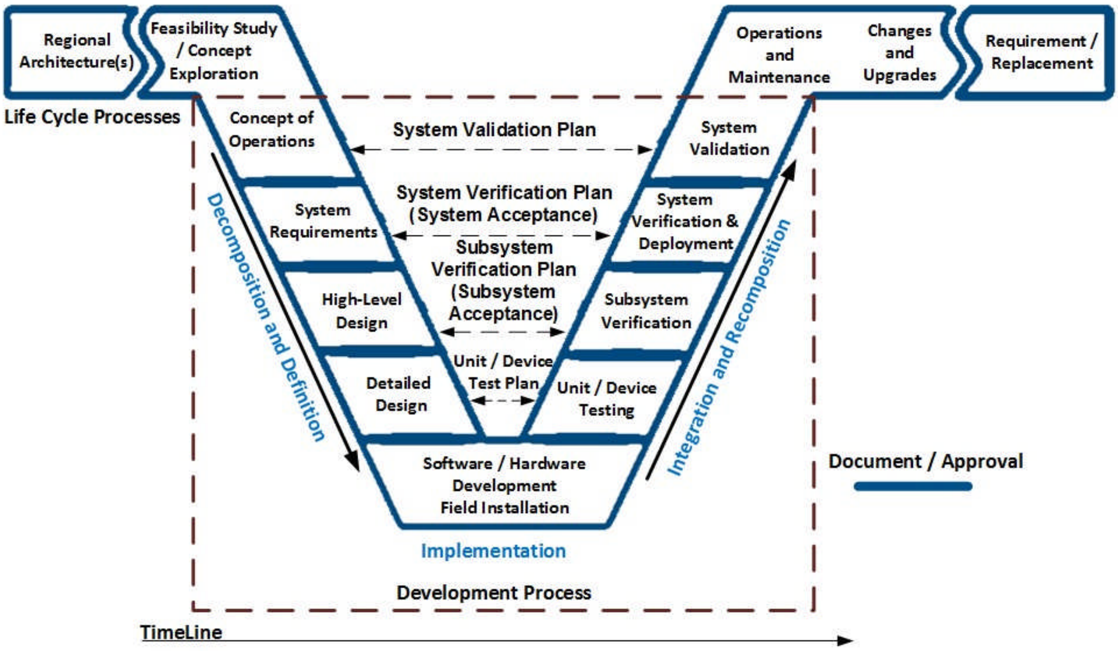


Figure 5-7 Systems Engineering Process - V-Model

"Systems engineering is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, then proceeding with design synthesis and system validation while considering the complete problem" INCOSE (2015b).

This definition covers the data lifecycle from planning, creation to retirement disposal. A key strength of the systems engineering process is the consideration of the complete lifecycle of a project during the project development phase. It depicts the ten basic steps involved in the conception, planning, functional and technical specifications, and the implementation of a system. The implementation can vary from system to system in terms of timescales, costs and predictable lifespan. The essential of the V-Model illustrates the gradual advancement from requirements specification, system/product implementation, to verification processes.

The left part of the V-model characterizes problem identification. It depicts the process of functional and technical requirements gathering with evolving granularity, for product systems and subsystems, including components and the relationship amongst them. The left part of the V-Model comprises artefact abstraction, selection, and the design description of the product. The right part of the V-Model (Figure 5.7) depicts activities relating to the running and preserving, changing and enhancing, and eventual obsolescence or replacement of the system (INCOSE, 2015a, 2017)

The Systems Engineering process can be used in the development of any product or system. The V-Model has thus guided the development of the proposed Data Life Cycle Model for Integrated Vehicle Health Management. The process can be summarized into the following seven tasks: State the problem (concept studies), Investigate alternatives (concept development), Model the system (preliminary design), Integrating the system (final design), Launch the system (fabrication), Assess performance (verify components/performance), and Re-evaluate (demonstration & Validation)(Jacobs, 2015).

5.3 Commonalities Between Data Lifecycle Models, OSA-CBM, Standards

Data management through its entire lifecycle still presents a number of complex challenges relating to interoperability, volume, storage, data citation, and metadata standards and data provenance (Yang, Matthews and Wilson, 2013; Porcal-Gonzalo, 2015; Beaujardière, 2016). This perhaps explains the proliferation of discipline or domain-specific data lifecycle models. Various disciplines and organizations are creating standardized frameworks, data ontologies, standards and unique data lifecycle models to suit their respective requirements. Metadata standards like Dublin Core (Hsu *et al.*, 2015), Core Scientific Metadata (CSMD) provisions the basic metadata required to enhance the search functionalities over data portals and knowledge libraries (Matthews *et al.*, 2010), but falls short of propagating the complete provenance data. The shortfall with the CSMD is that neither does it support for “*access to the derived data produced during analysis, nor does it allow the provenance of data supporting the final publication to be traced through the stages of analysis to the raw data*” as pointed out by Yang (2013).

However, though metadata standards are relevant for all data and knowledge outputs and organizations, each organization or project seems to have separate requirements for their own research or projects data. This is reflected in the 17 data lifecycle models identified, which all highlight the significance of metadata and standards.

The models seem to differ on the depth and breadth of applicability and priority of requirements. For instance the USGS emphasized three critical cross-cutting activities namely; metadata description, quality assurance and protection from corruption or loss to be performed parallel to planning, acquisition, processing, analysis, preservation, publishing and sharing to achieve enhanced quality, understanding and long-term reuse (Faundeen and Hutchison, 2017; Plale and Kouper, 2017)

The Digital Curation Centre (DCC) model suggests metadata should comprise rules and formalized entities for automatable tools and services as well as the role of data managers and data curators in the improvement of knowledge (Plale and Kouper, 2017). Taking the DataOne data lifecycle model to illustrate the variance, though it represents all the classical stages of the data management lifecycle, all parts of the lifecycle are not mandatory (Harrison, 2013; Allard, 2014; Pouchard, 2015; DataONE, 2017; Hidalga *et al.*, 2017; Plale and Kouper, 2017). This implies that the number of selected stages are dependent on the type of project or project requirements. Furthermore, although the USGS addresses the weakness in most data lifecycle models identified and can be used in diverse settings despite having been developed mainly for USGS science data. The requirements on which this is developed are not consistent with small research facilities like the IVHM centre Labs or the University as a whole. None of the models identified seems to satisfy the scenario of IVHM.

The core of IVHM is the capture and analysis of data to establish advance indication of a future failure, state or distinctive characteristics of current assets in current the state. The conduct of prognostics and diagnostics is reliant on high-quality, reliable data. This quality assurance can be determined by the data management lifecycle. Data lifecycle models are often integrated with software services and policies (Plale and Kouper, 2017). It is therefore fundamental to consider the associated process involved in the planning, designing and developing of the software services and policies.

The INCOSE system engineering process [V-Model] suitable for the development of any system, is vital in the scenario of IVHM Knowledge Management System and data lifecycle model. This is suitable for the IVHM data lifecycle and system development because of the restricted nature of this project. The V-Model is best suited for these kinds of projects that have well-defined length and scope, consistent technology, and a clear and well documented technical and functional specifications (INCOSE, 2017).

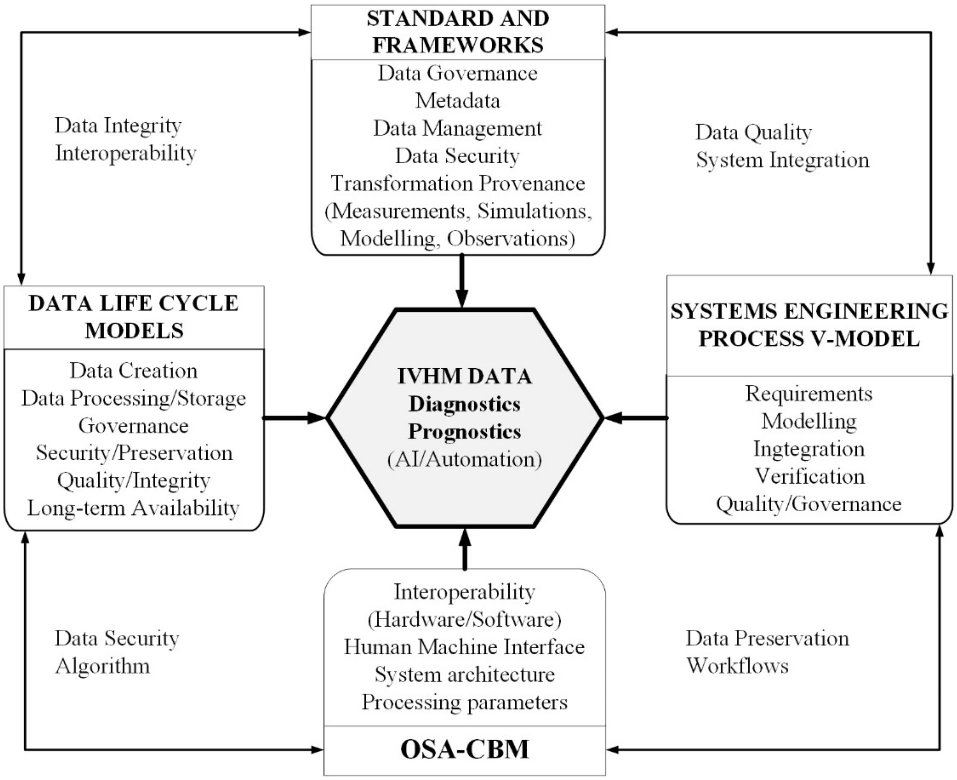


Figure 5-8 Integration Parameters

The conceptual model above (Figure 5-8) sheds light on the relationship between standards, data lifecycle models, Open System Architecture for Condition-Based Maintenance (OSA-CBM) and the INCOSE systems engineering process model. It depicts the commonalities (integration parameters) that are required for the IVHM Data lifecycle model.

Integrated Vehicle Health Management (IVHM) is shaped by prognostics and diagnostics that rely predominantly on the availability of high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health. The data has to be accurate, complete, timely, context-relevant, reliable and explicit (Dibsdale, 2011). IVHM is data-centric and driven by the OSA-CBM data model. The centrality of data for IVHM in the short, medium and long-term diagnostics whether it be historical asset health trends, operational status, load history, or fault identification, necessitates a data lifecycle model or a hybrid model consistent with OSA-CBM.

Of all the existing data lifecycle models, there is none that is consistent with the requirements of IVHM data and knowledge management requirements; that integrates OSA-CBM, which is absolutely imperative to IVHM. The OSA-CBM model is based on the concept of metadata and interoperability that requires persistent visibility and traceability of data (Choudhary, Perinpanayagam and Butans, 2016) within and across diverse platforms, systems and devices, and therefore making data provenance a fundamental requirement. This has not been explicitly covered in any of the data lifecycle models. A new model is needed, and one that integrates data acquisition with signal reception as well as data entry in order to accommodate the role of the systems or device operators for IVHM. In the next section, we describe the various phases of the proposed IVHM data lifecycle model.

5.4 The Integrated Vehicle Health Management Data Lifecycle Model

The proposed IVHM Data lifecycle model (IVHM-DLCM) (Figure 5.9) is a hybrid that integrates relevant standards, frameworks and models that fit the profile of IHVM research and engineering activities. The IVHM-IVHM Data lifecycle model is scalable and can be used in diverse IVHM activities of all depths and breadths. It takes into consideration interoperability, integrity, quality, security, provenance and preservation of data throughout the lifecycle.

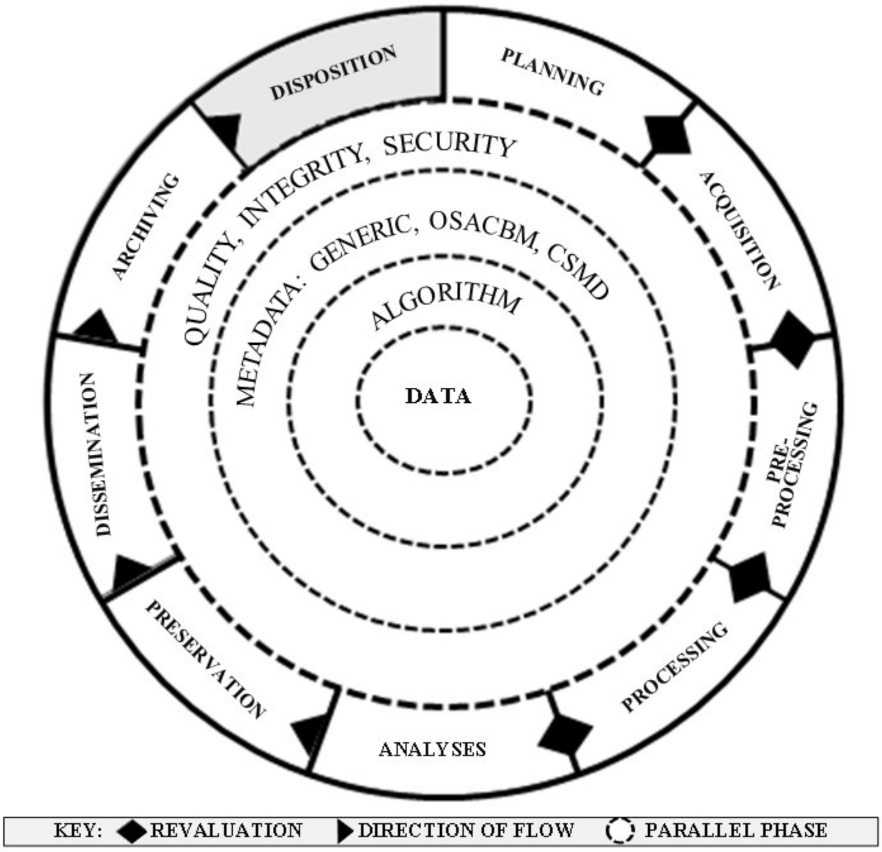


Figure 5-9 The IVHM Data Lifecycle Model

The IVHM-DLCM has nine discrete phases with three other phases running parallel. There also the revaluation process that runs parallel to phases One, Two, Three and Four.

5.4.1. Distinct Phases

5.4.1.1 Planning

The planning phase is the conception and beginning of simulations, observations, derivation, experiments and referencing (SODER) activity. In this phase, the research or engineering task is defined and planned – resources and planned deliverables for each phase are explored and explained. The requirements for success, quality, integrity and security are defined, including systems requirements. The metadata parameters both generic and standard are considered in this phase. The data management plan, retention schedules and requirements specification are some of the recommended deliverables in this phase. In this phase, we also do the selection of the sensor modules and other accessories like software tools required for the SODER activity. File formats, data storage and sharing plans are developed at this stage.

5.4.1.2 Acquisition

The data acquisition phase represents data creation from scratch or the extraction of existing raw or derived data for reuse. In this stage, observational, experimental, simulation data creation takes place as well as the retrieval of existing datasets, derived or reference data for reuse. Data governance standards and best practices ensuring integrity, security, quality and metadata are considered. The quality of the process is also vital as it has a direct impact on the quality of the data created of retrieved for reuse.

5.4.1.3 Pre-Processing

The pre-processing phase aims to flag out of range data values, missing values to mitigate the risk of making decisions based on misleading results. It is fundamental to IVHM machine learning activities. It is used to clean the original signal by eliminating noise and to improve object component condition. In other words, this phase represents the low-level computation of sensor data and constitute a key element of the Open System Architecture for Condition-Based Maintenance (OSA-

CBM) architecture. Sensor data is transformed into an understandable format. This is the stage where data cleansing –detecting and correcting mistakes, incomplete, inaccurate, irrelevant and incorrect records from datasets. This step employs best practices to ensure that data is free of inconsistencies, correct, usable and reliable.

5.4.1.4 Processing

In this phase, meaningful and relevant is extracted in suitable formats from raw datasets created after pre-processing for future use. This step involves activities such as validation, aggregation, summarization, sorting, classification and validation. It includes the conversation of data in usable and desired forms and formats. It can take the form mechanical, manual or digital processing.

5.4.1.5 Analyses

This represents activities like organization, interpretation and presentation of data. It involves statistical data analytics, simulations, modelling and other computation activities that reveal trends, facts, faults and tests theories and assumptions. In this step raw data is transformed into information and communicated to the stakeholders.

5.4.1.6 Revaluation: Phase 1 – Phase 5

Revaluation is one of the most important tools in The Systems Engineering Process. Re-evaluation observation of outputs and using the information to modify the system, the inputs, the product or the process (INCOSE, 2017). It takes place in the first five phases. Feedback is collected at each phase on a situational basis to continually improve the steps and eliminate problems. The loops are used specifically when issues are identified as the lifecycle moves from one phase to the next.

5.4.1.7 Preservation

Preservation includes steps and processes for active data storage for the duration that it might be needed. It also involves access control and backup for security. These are the security actions taken to reduce the chances of data corruption and data loss. They also include submission to reliable data repositories.

5.4.1.8 Dissemination

This step includes the preparation and dissemination of datasets, derived data as well as findings or outcomes to relevant stakeholder communities. It improves accessibility as well as being a recommended best practice.

5.4.1.9 Archive

This the last step before disposal. It represents the retraction of data and related outputs from active circulation. The retention schedule is applied to the data and only accessible on demand.

5.4.1.10 Dispose

The data has reached the end of its useful and potential useful life. It is purged at this stage. It is securely disposed of to ensure that there can be no unwarranted access.

5.4.2. Parallel Phases

5.4.2.1 Metadata: Generic, CSMD, OSA-CBM

The strategy for metadata generation and documentation is referenced throughout the data lifecycle. As the data transitions from one phase to the next, it might also change platforms and devices, making this phase particular important for addressing provenance issues. The metadata phase draws on the best generic metadata created by systems, the core scientific metadata model and the Open System Architecture for Condition-Based Maintenance (OSA-CBM) built-in metadata to create a scalable metadata model within the lifecycle. This allows for easy contraction and expansion research and engineering activities. It accommodates small to large scale data-generating activities.

5.4.2.2 Data Integrity, Quality and Security

Data integrity involves creating, processing and maintaining the assurance, accuracy, consistency and completeness of data throughout its lifecycle. The content and meaning of data is maintained throughout its lifecycle. This also includes compliance with statutory requirements. Quality represents the use of best practice protocols and methods of collecting and organizing data that ensures its accessibility, completeness, validity, accuracy, consistency, availability and timeliness. Security involves the protection of data from unauthorized access to modify, use, delete and disclose. It includes protection against theft, breach of agreements, data protection laws and unintended or hateful modification. The computer system security, physical security and file security are all part of this step.

5.4.2.3 Algorithms

Refers to the algorithms and agents that automate research and engineering tasks. They are maintained as datafiles and at the same time they represent agents within the data- and Knowledge Management System.

5.5 Conclusion

We have identified key data lifecycle models (DLCM) and frameworks and found that though they had some of the elements for IVHM data, they lacked some essential ones. Thus, sustaining the cycle of data and knowledge management issues – creation, quality, storage, security and provenance. In the review, we found that the USGS DLCM encapsulated most data lifecycle models. The USGS reviewed more than 50 DLCMs to develop their one, and therefore was chosen as the ideal Data Life Cycle for integration. The strength of the proposed data lifecycle lies in the integration of key elements of the OSA-CBM framework, the CSMD, the engineering process model to create a scalable model that fits the depth and breadth of IVHM Research and Engineering operations. This model supports the design and implementation of protocols for effective and efficient data management. It provides a foundation for Data- and Knowledge Management System requirements.

Metadata, in addition to guaranteeing data safety and security, provides long-term preservation and re-use of data and makes it simpler to discover relevant data. It also allows us to grasp what the texts are about. Using metadata reduces the amount of time it takes to find relevant information. It also makes it easier to discover text documents since metadata describes exactly what they include. When reusing existing data, metadata retention and documentation are critical. The Data Life Cycle Model and metadata have for values in common: they both enable classification of content, information security, search and retrieval, and improvement of Customer experience.

ACKNOWLEDGEMENT

This Project is sponsored by Boeing and hosted by Integrated Vehicle Health Management (IVHM) Centre at Cranfield University, United Kingdom.

5.6 References

Allard, S. (2014) ‘Evaluating a Complex Project: DataOne’, in Ray, J. M. (ed.) Research Data Management: Practical Strategies for Information Professionals. Purdue University Press (Charleston Insights in Library, Informat), p. 436.

Arahchige, B. and Perinpanayagam, S. (2017) ‘Uncertainty quantification in aircraft gas turbine engines’, in Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering. IMECHE, pp. 1628–1638. doi: 10.1177/0954410017699001.

Arass, M. E., Tikito, I. and Souissi, N. (2017) ‘Data lifecycles analysis: Towards intelligent cycle’, 2017 Intelligent Systems and Computer Vision, ISCV 2017. doi: 10.1109/ISACV.2017.8054938.

Ball, A. (2012) ‘Review of Data Management Lifecycle Models.’ University of Bath, p. 15.

Beaujardière, J. D. La (2016) ‘NOAA Environmental Data Management’, Journal of Map & Geography Libraries. Routledge, 12(1), pp. 5–27. doi: 10.1080/15420353.2015.1087446.

Choudhary, R., Perinpanayagam, S. and Butans, E. (2016) ‘Design and Analysis of Communication Model for Implementation of CBM Systems Based on OSA-CBM Framework’, in 2016 IEEE Aerospace Conference, pp. 1–7. doi: 10.1109/AERO.2016.7500821.

DataONE (2017) Data Life Cycle, Best Practices.

DCC (2008) ‘The DCC Curation Life-Cycle Model’.

Dibsdale, C. (2011) ‘Integrated Vehicle Health Management Operations Rooms’, in Jennions, I. K. (ed.) Integrated Vehicle Health Management: Perspectives on an Emerging Field. SAE International, pp. 113–123.

Faundeen, J. and Hutchison, V. (2017) ‘The Evolution, Approval and Implementation of the U.S. Geological Survey Science Data Lifecycle Model’, Journal of eScience Librarianship, 6(2), p. e1117. doi: 10.7191/jeslib.2017.1117.

Faundeen, J. L. et al. (2013) The United States Geological Survey Science Data Lifecycle Model, U.S. Geological Survey Open-File Report 2013–1265. South Dakota: U.S. Dept. of the Interior. doi: http://dx.doi.org/10.3133/ofr20131265.

Felke, T. et al. (2010) ‘Architectures For Integrated Vehicle Health Management’, AIAA Infotech@Aerospace 2010, (April). doi: 10.2514/6.2010-3433.

Goebel, K. (2011) ‘Basic Principles’, in Jennions, I. K. (ed.) Integrated Vehicle Health Management: Perspectives on an Emerging Field. SAE International, pp. 55–65.

Ten Ham-Baloyi, W. and Jordan, P. (2016) ‘Systematic Review as a Research Method in Postgraduate Nursing Education’, Health SA Gesondheid, 21, pp. 120–128. doi: https://doi.org/10.1016/j.hsag.2015.08.002.

Harrison, K. (2013) A Guide to Data Management in Ecology and Evolution. Edited by K. Harrison. British Ecological Society.

Hidalga, A. N. de la et al. (2017) ENVRI Reference Model News.

Hsu, L. et al. (2015) ‘Data Management, Sharing, and Reuse in Experimental Geomorphology: Challenges, Strategies, and Scientific Opportunities’, Geomorphology, 244, pp. 180–189. doi: 10.1016/j.geomorph.2015.03.039.

INCOSE (2015a) INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities. 4th edn. Edited by W.-E. DaVid d. et al. John Wiley and Sons, Inc.

INCOSE (2015b) Systems engineering handbook: A guide for system life cycle processes and activities. Fourth. Edited by W.-E. DaVid d. et al. Wiley. doi: INCOSE-TP-2003-002-03.2. 1.

INCOSE (2017) Guide to the Systems Engineering Body of Knowledge (SEBoK). v1.9 R.D. Edited by R. Adcock. Hoboken, NJ: The Trustees of the Stevens Institute of Technology.

Jacobs, S. (2015) ‘Systems Engineering’, in Jacobs, S. (ed.) Engineering Information Security. John Wiley & Sons, Inc., pp. 31–60. doi: 10.1002/9781119104728.ch02.

Jantunen, E. et al. (2017) ‘The Way Cyber Physical Systems Will Revolutionise Maintenance’, in 30th Conference on Condition Monitoring and Diagnostic

Engineering Management (COMADEM 2017). 10 to 13, Jul, 2017. Preston, United Kingdom., pp. 145–148.

Jantunen, E., Junnola, J. and Gorostegui, U. (2017) ‘Maintenance supported by cyber-physical systems and cloud technology’, in 2017 4th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 708–713. doi: 10.1109/CoDIT.2017.8102678.

Jennions, I. K. (2011) ‘Introduction’, in Jennions, I. K. (ed.) Integrated vehicle health management: perspectives on an emerging field. SAE International, pp. 1–7.

Lebold, M. et al. (2002) ‘OSA-CBM architecture development with emphasis on XML implementations’, Maintenance and Reliability Conference (MARCON), pp. 6–8.

Löhr, A. and Buderath, M. (2014) ‘Evolving the Data Management Backbone : Binary OSA-CBM and Code Generation for OSA-EAI’, in EUROPEAN CONFERENCE OF THE PROGNOSTICS AND HEALTH MANAGEMENT SOCIETY 2014. PHM, pp. 1–10.

Matthews, B. et al. (2010) ‘Using a Core Scientific Metadata Model in Large-Scale Facilities’, International Journal of Digital Curation, 5(1), pp. 106–118. doi: 10.2218/ijdc.v5i1.146.

MIMOSA (2018) Open System Architecture for Condition-Based Maintenance (OSA-CBM), OSA-CBM.

Parry, M. J. (2016) ‘Digital curation in the digital humanities: preserving and promoting archival and special collections’, The Electronic Library. Emerald Group Publishing Limited, 34(6), p. 1055.

Perinpanayagam, S. (2013) ‘Chapter 2 Sensors, Instrumentation, and Signal Processing’, in Jennions, I. (ed.) Integrated Vehicle Health Management: the Technology. Pennsylvania: SAE International, pp. 5–25.

Plale, B. and Kouper, I. (2017) ‘The Centrality of Data: Data Lifecycle and Data Pipelines’, in Data Analytics for Intelligent Transportation Systems. Elsevier, pp. 91–111. doi: https://doi.org/10.1016/B978-0-12-809715-1.00004-3.

Porcal-Gonzalo, M. C. (2015) ‘A Strategy for the Management, Preservation, and Reutilization of Geographical Information Based on the Lifecycle of Geospatial Data: An Assessment and a Proposal Based on Experiences from Spain and Europe’, Journal of Map & Geography Libraries. Routledge, 11(3), pp. 289–329. doi: 10.1080/15420353.2015.1064054.

Pouchard, L. (2015) ‘Revisiting the Data Lifecycle with Big Data Curation’, International Journal of Digital Curation, 10(2), pp. 176–192. doi: 10.2218/ijdc.v10i2.342.

Prajapati, A. K., Roy, B. K. and Prasad, R. (2018) ‘A State of Art Review of Integrated Vehicle Health Management System’, in 2018 3rd International Conference for Convergence in Technology (I2CT), pp. 1–5. doi: 10.1109/I2CT.2018.8529590.

Redding, L. (2011) ‘Integrated vehicle health management: perspectives on an emerging field’, in Jennions, I. K. (ed.) Integrated vehicle health management: perspectives on an emerging field. SAE International, pp. 17–26.

Sabharwal, A. (2015) Digital curation in the digital humanities: Preserving and promoting archival and special collections. Chandos Publishing.

Simonet, A., Fedak, G. and Ripeanu, M. (2015) ‘Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures’, Future Generation Computer Systems, 53(Supplement C), pp. 25–42. doi: https://doi.org/10.1016/j.future.2015.05.015.

Sreenuch, T., Tsourdos, A. and Jennions, I. K. (2013) ‘Distributed embedded condition monitoring systems based on OSA-CBM standard’, Computer Standards & Interfaces, 35(2), pp. 238–246. doi: 10.1016/j.csi.2012.10.002.

Yang, E., Matthews, B. and Wilson, M. (2013) ‘Enhancing the core scientific metadata model to incorporate derived data’, Future Generation Computer Systems, 29(2), pp. 612–623. doi: 10.1016/j.future.2011.08.003.

Yu, X. and Wen, Q. (2010) ‘A View about Cloud Data Security from Data Life Cycle’, 2010 International Conference on Computational Intelligence and Software Engineering, (4072020), pp. 1–4. doi: 10.1109/CISE.2010.5676895.

CHAPTER 6: THE IMPORTANCE OF DATA LIFE CYCLE MODEL IN IVHM RESEARCH AND ENGINEERING PRACTICE: A VALIDATION SURVEY.

Abstract: Data life cycle models are imperative to IVHM data- and knowledge management practices. They are increasingly being used in research and engineering activities. For data life cycle models to be accepted and used to support research, engineering practices and decision making, it is imperative that they are verified and validated. The verification and validation ensure that the specialist model performs as intended, and represents and correctly reproduces the behaviours of the real world. Peer review and structured walkthrough are undoubtedly some effective best practices used in the validation of models. This paper uses a purposive survey to query expert perception of the proposed IVHM Data life cycle model as an additional validation tool.

6.1 Introduction

Although the verification and validation of models are absolutely imperative, there are no particular tests that can be readily applied on them to help decide their correctness. The correctness of the model is important in alleviating fears of users and decision-maker about the accuracy and reliability of their outcomes. These fears are resolved through verification and validation (Sargent, 2011; Krogstie, 2012, 2016c; Nelson *et al.*, 2012). It is still a challenge to find an algorithm for selecting relevant validation techniques and procedures thereby making every new model a challenge (Sargent, 2011) with regards to determining its accuracy, completeness and correctness. Validation has often been described as the "*process of determining the degree to which a model is an accurate representation of the real world*" (Coleman and Steele, 2018) "*from the perspective of the intended uses*" (Carson, 2002; Sornette *et al.*, 2007; Graziani, 2008). Model validation is the use of both formal and informal quantitative and qualitative tools used to determine the accuracy and correctness of a model (Barlas, 1996; Sargent, 2013). Model verification is ensuring that it is fit for purpose “and its implementation are correct" (Sargent, 2015).

The objective of this section is to finalise the verification and validation of the developed data life cycle model through peer review or expert consultation. This is achieved by administering an online survey (Appendix B) using Qualtrics survey tool. The survey seeks to verify and validate the following characteristics: adequate accuracy, correctness of the logic, adequate level of detail, reasonableness of the relationships, structure and flow. The questionnaire survey was developed to support an evidence-based process, and further provide a practice guideline for the validation of non-statistical, non-mathematical, conceptual data life cycle models. Fundamentally, the survey of experts and professionals was to ascertain that the model addresses the right problem, provides accurate information about the system or process being modelled and that it can actually be used.

6.2 Survey Methodology

6.2.1 Sample Selection

For this study, a purposive sample of 126 participants from both industry and academia were surveyed via the online Qualtrics survey tool after receiving the ethics approval (Appendix A). Purposive sampling helps in the process of identifying and selecting information-rich cases related to the phenomenon under investigation (Palinkas *et al.*, 2015). A purposive sample has been described as a sort of non-probabilistic judgmental or expert sampling technique (Sharma, 2017) aimed at generating a sample assumed to be reasonably characteristic of the population. It can be achieved by applying professional experience of the demographic to select a sample of entities that reflect a cross-section of the population in a non-random manner (Lavrakas, 2011).

Such purposive sampling techniques include high sampling variance, homogeneous sampling and normal sampling of cases. In addition, it involves extreme sampling of cases, complete population sampling ad judgmental sampling (Etikan, Musa and

Alkassim, 2016; Sharma, 2017). The advantage of purposive sampling is that it provides a theoretical, analytical and logical foundations for generalisations. It also provides a variety of non-probabilistic sampling techniques, especially in multi-phase research designs. The disadvantage of the purposive sampling rests on the high risk of researcher bias resulting from judgemental subjectivity, especially where such judgements have been ill-conceived and lacking in criteria clarity or expert elicitation parameters. It has also been argued that the non-probabilistic nature makes it hard to defend the representativeness of the sample and to convince the reader of the works of theoretical, analytical and logical extrapolation (Etikan, Musa and Alkassim, 2016).

This survey is subject to possible sampling error; that is to say the findings could vary from those that would have been reached if the entire population taking selected for the research were surveyed. For the entire sample, the sampling error range is plus or minus 2.5 percentage points at confidence level of 95 per cent. The participants were selected mainly from Cranfield University, partner organisations and other academic institutions. At the beginning of the questionnaire, the context was set and the term “Data life Cycle model” was defined as “*…a formal representation of all the possible states and all the valid state transitions of a data item, when handled by a particular system or by an user application, e.g. created, duplicated, deleted, backed-up*” (Simonet, Fedak and Ripeanu, 2015).

6.3 Survey Results

6.3.1. Survey Distribution

The following (Table 6.1 and Figure 6.1) shows the responses returned from the different professional groups that took part in the survey. A large number of respondents were Researchers (41%) followed by Engineers (32%) as well as Managers (8%) and Analysts (8%). The largest number of respondents were from Industry with a combined 48% of responses followed closely by academia with a combined 47% responses. There were an additional 5% with non-specified occupation.

Table 6.8 Demographic Composition and Response Count

Participant Background	Count
Engineer	13
Manager	3
Researcher	15
Professor	1
Lecturer or Senior Lecturer	1
Analyst	3
Other	2
Total	38

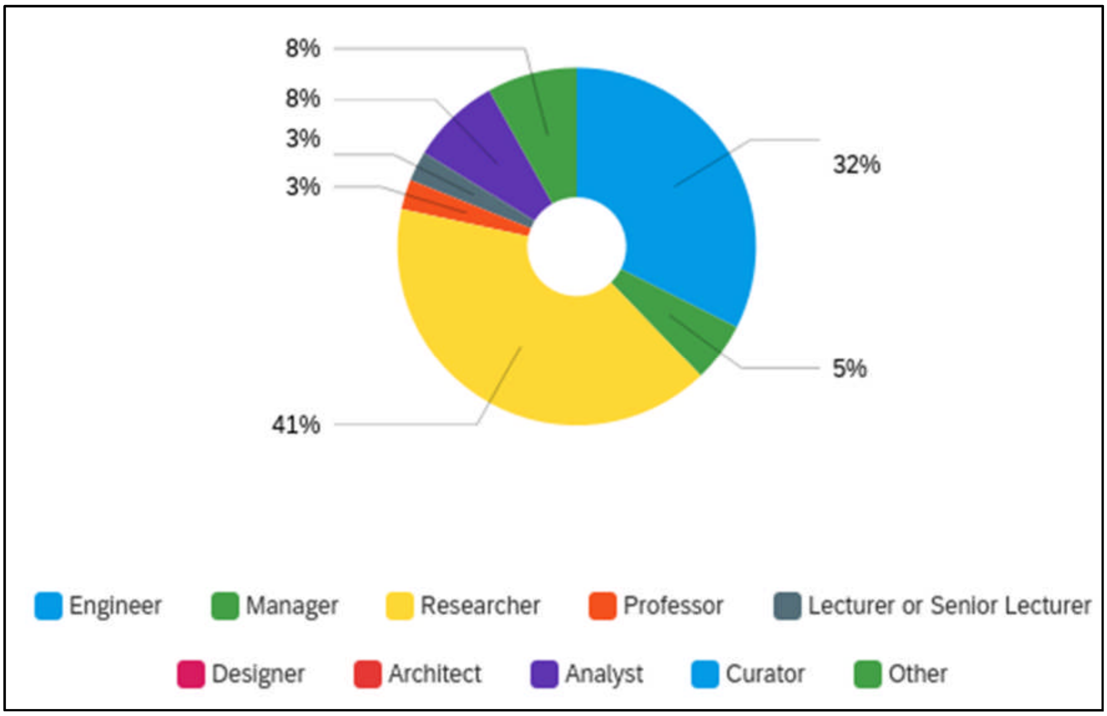


Figure 6-1 Proportion of Responses

6.3.2 Response Rate and Representativeness

Of the 126-participant surveyed, there were 64 survey entrants of which 38 successfully completed their responses. This level of participation gave a response rate (RR) of 30.2% and a completion rate (CR) of 59.4%. Response Rates represent the number of participants divided by the number of respondents surveyed. In order to avoid non-response bias researchers are expected to achieve an RR of about

60% as it has implications for representativeness. It studies where non-response bias occurs, it undermines the reliability and validity of the survey. A non-responsive bias of 70% occurs if a survey attains only a 30% RR. If a survey returns an RR of 20%, it gets a nonresponse bias of 80% (Kelley *et al.*, 2003; Baruch and Holtom, 2008; Fincham, 2008; Manfreda *et al.*, 2008; Nulty, 2008; Fowler Jr, 2013; Dillman, Smyth and Christian, 2014; Rea and Parker, 2014; Fink, 2017). However, the response rate is important if it bears on representativeness"(Cook, Heath and Thompson, 2000).

$$Response\ rate(RR) = \frac{Total\ Number\ of\ Responses}{Total\ number\ Surveyed} \times 100$$

CR is calculated by dividing the number of valid responses returned by the total number suitable in the chosen population

$$Completion\ rate(CR) = \frac{Completed\ Responses}{Survey\ entrants} \times 100$$

Representativeness: Representativeness refers to how well the sample drawn for the survey research compares with the target population. Lack of response to the questionnaire by potential respondents in a sample or population is referred to as non-response bias (Fincham, 2008; Dillman, 2011; Lavrakas, 2011; Fowler Jr, 2013; Fink, 2017; Lavrakas *et al.*, 2019).

Figure 6.2 below details the responses of the participants on the extent to which they agreed with various characteristics of the IVHM Data Life Cycle Model. About five in Six (84%) of the participants 'Strongly agreed' or 'agreed' with the statement 'This model accurately represent the data life cycle'. Over four in five (81%) 'Strongly agreed' or 'agreed' with the statement 'The model is clear and easy to understand'. Almost two thirds (68%) of respondent either 'Strongly agreed' or 'agreed' with the statement that 'The model layout is easy to read and pleasant to the eye' while 21% expressed indifference and 10% disagreed. On the subject of the model sequence, approximately four in five (79%) 'Strongly agreed' or 'Agreed' with the assertion that ‘the model design sequence is accurate’. On the improvements of requires specification, about two thirds (68%) ‘Strongly agreed’ or ‘Agreed’ with the statement that ‘The model enhances requirements specification’. Nine participants (24%) indifferent while about three more (8%) disagreed in the subject of requirements specification. Approximately five in six (84%) ‘Strongly agreed’ or ‘Agreed’ with the statement that ‘The terminology is clear and correct’. Finally, there was no strong disagreement on any of the assessed characteristics of the data life cycle model.

6.3.3 Characteristics of the Proposed Data Life Cycle Model

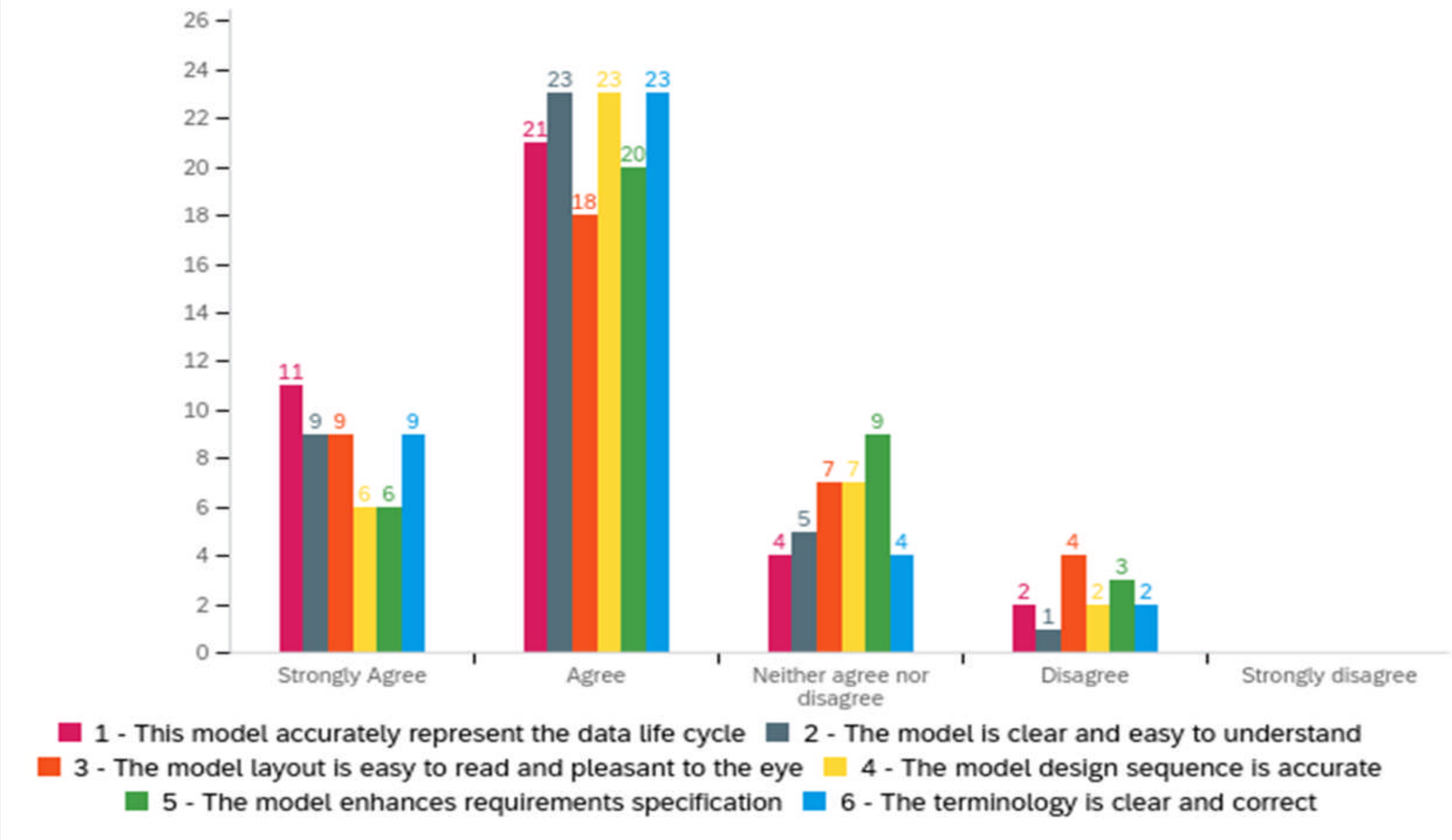


Figure 6-2 Q2 To what extent do you agree with the following statements?

As shown in Figure 6.3 below there is an overall agreement on most of the physical, empirical, semantic, and pragmatic and deontic quality issues assessed in this question.

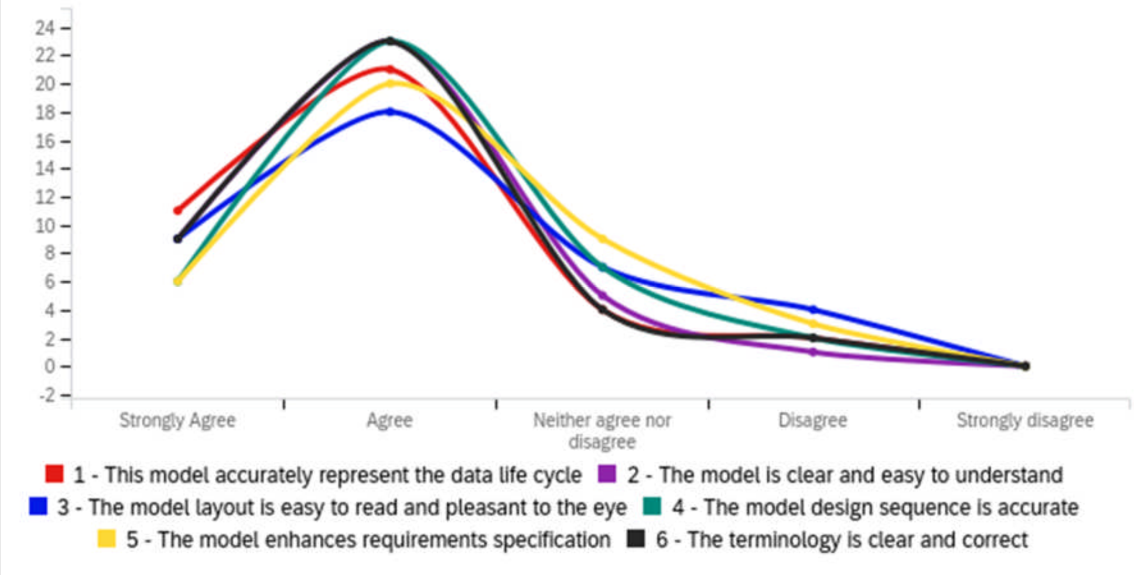


Figure 6-3 Overall Perception of the IVHM Data Life Cycle Model

6.4 Implementing the Model

The respondents were invited (Q3) (Appendix C) to 'Identify facilitators to the model's implementation. Examples of responses included:

- OEMs
- University research project funded by external organisations.
- Computational power and memory need to be accounted for. The cycle of refresh rates and archiving for the live data could be included.
- The facilitators will be part of a variety of stakeholders starting from operators, maintenance engineers, life cycle and cost engineers, strategy and innovation managers etc. On the academic side, it will be researchers working in the area of product life-cycle.
- IVHM staff and participants in the project.
- Industries, third parties, academics
- industries, Academics and Government organizations
- Qualified trained engineers, Software, Hardware, Time Management

- Require more depth and descriptions.

6.5 Recommending or Adopting the Model

Generally, more two thirds (73%) of the respondents (Figure 6.4) indicated a likely they would recommend the model to a friend, colleague or organisation. Almost a third (27%) specified an ‘Extreme’ likelihood they would do so whilst 46% expressed said they ‘likely’ to recommend the model.

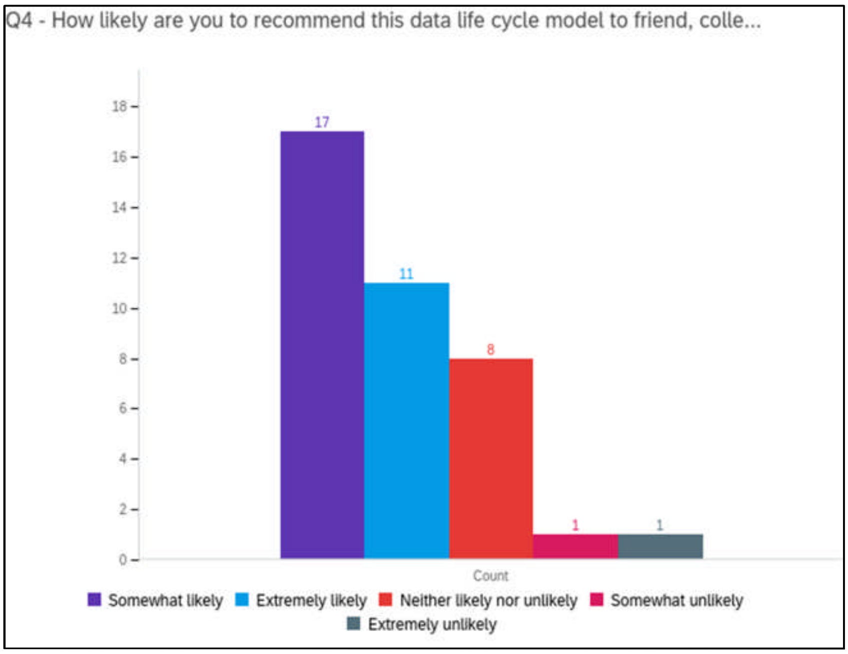


Figure 6-4 Q4. How likely are you to recommend this data life cycle model to friend, colleague, or organisation?

A further eight of the respondents (22%) indicated they would ‘Neither’ likely ‘Nor’ unlikely to make a recommendation of the model.

6.6 Difficulties of Implementation

The participants were requested (Q5) to ‘Identify barriers to the model's implementation’, and some of their views (Appendix D) included:

- Maybe will meet the challenge from some existing data management system. For a more professional perspective, it is unknown from my current level.

- Probably a key barrier is getting the model adopted by a key player or organisation. I think you need both a simplified version - of the sort that would be on presentation slides or a poster - because what you have just now is quite a lot to take in - and a development where you map what the implications of each phase are for organisations.
- Validation, data security and confidentiality
- It's very high level; there isn't sufficient detail to actually implement.
- A use case implementation may be helpful.
- Loss of information between the stages
- This is tricky as it is dependent on the organisation, and issues associated with data storing and sharing practices and the gate keepers who handle this.
- Missing information
- lack of completeness
- Poorly written Software, Inefficient Hardware.
- Apply this to machine learning approaches, how does training fit in to the model? How about simulation and generation of data for machine learning?
- lack the importance of an agile approach when applying model in enterprise environment, during continuous iteration of the lifecycle sometimes steps are skipped, need for a flexible approach throughout successive steps in cycle
- The similarities in its approach may deter uptake as it could be seen as an unnecessary replacement

6.7 Additional Observations

The survey invited respondents to provide additional comments and suggestions about the model. This request returned a diverse set of ideas (Appendix E) some outlined below:

- You are effectively asking us to do "peer review" of a model. I don't see the context, the empirical grounding, nor how this is related to existing theories and models. I therefore cannot meaningfully evaluate it.

- Every project has unique peculiarities. So it is extremely hard to correlate all different fields in a single model. However, it is a very good approach for a generic architecture.
- A flow chart description may help the reader easily understand the overall model.
- Contingencies for backing up or recovering lost data required […].
- By using this model academics [and] researchers will learn about the new contexts in knowledge management and a good framework for managing research data. In industries, the model is robust and suitable for long-time industrial information management.
- Overall, this model is impressive. It is correct, to the point and avoids un-necessary 'jargon'. The diagram could be re-designed slightly as the dotted lines that form the circumference of the inner circles can play 'mind-tricks', causing the reader to lose focus, so a solid line or some larger dashes may suit better.
- The headings of the development life cycle phase is good, but direction is lost in top-level descriptions.
- Using this approach in a smaller acquisition process to prove the concept would strengthen the case for its use

6.8 Conclusion

In summary, these preliminary research findings have not only helped to provide a basic understanding of the data life cycle model and its value, but have also highlighted certain topic areas for further investigation. It was generally accepted that the model is a valid one for address a wide range of IVHM research and engineering data management challenges.

The data collected from the 38 respondents has confirmed that the expressed characteristic of the Data Life Cycle Model are valid.

6.9 References

Barlas, Y. (1996) ‘Formal aspects of model validity and validation in system dynamics’, System Dynamics Review: The Journal of the System Dynamics Society. Wiley Online Library, 12(3), pp. 183–210.

Baruch, Y. and Holtom, B. C. (2008) ‘Survey response rate levels and trends in organizational research’, Human relations. Sage Publications Sage UK: London, England, 61(8), pp. 1139–1160.

Carson, J. S. (2002) ‘Model verification and validation’, in Proceedings of the winter simulation conference. IEEE, pp. 52–58.

Coleman, H. W. and Steele, W. G. (2018) Experimentation, validation, and uncertainty analysis for engineers. John Wiley & Sons.

Cook, C., Heath, F. and Thompson, R. L. (2000) ‘A meta-analysis of response rates in web-or internet-based surveys’, Educational and psychological measurement. Sage Publications Sage CA: Thousand Oaks, CA, 60(6), pp. 821–836.

Dillman, D. A. (2011) Mail and Internet surveys: The tailored design method--2007 Update with new Internet, visual, and mixed-mode guide. John Wiley & Sons.

Dillman, D. A., Smyth, J. D. and Christian, L. M. (2014) Internet, phone, mail, and mixed-mode surveys: the tailored design method. John Wiley & Sons.

Etikan, I., Musa, S. A. and Alkassim, R. S. (2016) ‘Comparison of convenience sampling and purposive sampling’, American journal of theoretical and applied statistics. New York, 5(1), pp. 1–4.

Fincham, J. E. (2008) ‘Response rates and responsiveness for surveys, standards, and the Journal’, American journal of pharmaceutical education. American Journal of Pharmaceutical Education, 72(2), p. 43. doi: 10.5688/aj720243.

Fink, A. (2017) How to Conduct Surveys: A Step-by-Step Guide. 6th edn. Sage Publications.

Fowler Jr, F. J. (2013) Survey research methods. Sage publications.

Graziani, F. (2008) Computational methods in transport: verification and validation. Springer.

Kelley, K. et al. (2003) ‘Good practice in the conduct and reporting of survey research’, International Journal for Quality in health care. Oxford University Press, 15(3), pp. 261–266.

Krogstie, J. (2012) ‘Quality of Business Process Models’, in Sandkuhl, K., Seigerroth, U., and Stirna, J. (eds) The Practice of Enterprise Modeling. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 76–90.

Krogstie, J. (2016) 'SEQUAL Specialized for Business Process Models', in Quality in Business Process Modelling. Springer International Publishing, pp. 103–138. doi: 10.1007/978-3-319-42512-2_3.

Lavrakas, P. (2011) 'Encyclopedia of Survey Research Methods'. Thousand Oaks, California. doi: 10.4135/9781412963947 NV - 0.

Lavrakas, P. J. et al. (2019) 'Probability Survey-Based Experimentation and the Balancing of Internal and External Validity Concerns', in Experimental Methods in Survey Research. (Wiley Online Books), pp. 1–18. doi: doi:10.1002/9781119083771.ch1.

Manfreda, K. L. et al. (2008) 'Web surveys versus other survey modes: A meta-analysis comparing response rates', International Journal of Market Research, 50(1), pp. 79–104. doi: 10.1177/147078530805000107.

Nelson, H. J. et al. (2012) 'A conceptual modeling quality framework', Software Quality Journal, 20(1), pp. 201–228. doi: 10.1007/s11219-011-9136-9.

Nulty, D. D. (2008) 'The adequacy of response rates to online and paper surveys: what can be done?', Assessment & evaluation in higher education. Routledge, 33(3), pp. 301–314.

Palinkas, L. A. et al. (2015) 'Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research', Administration and policy in mental health, 42(5), pp. 533–544. doi: 10.1007/s10488-013-0528-y.

Rea, L. M. and Parker, R. A. (2014) Designing and conducting survey research: A comprehensive guide. John Wiley & Sons.

Sargent, R. (2011) 'Verification and Validation of Simulation models', in Engineering Management Review, IEEE, pp. 166–183. doi: 10.1109/WSC.2010.5679166.

Sargent, R. G. (2013) 'Verification and validation of simulation models.', Journal of simulation. Springer, 7(1), pp. 12–24.

Sargent, R. G. (2015) 'An introductory tutorial on verification and validation of simulation models', in 2015 Winter Simulation Conference (WSC), pp. 1729–1740. doi: 10.1109/WSC.2015.7408291.

Sharma, G. (2017) 'Pros and cons of different sampling techniques', International journal of applied research, 3(7), pp. 749–752.

Simonet, A., Fedak, G. and Ripeanu, M. (2015) 'Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures', Future Generation Computer Systems, 53(Supplement C), pp. 25–42. doi: https://doi.org/10.1016/j.future.2015.05.015.

Sornette, D. et al. (2007) 'Algorithm for model validation: Theory and applications', in Proceedings of the National Academy of Sciences, pp. 6562 LP – 6567. doi: 10.1073/pnas.0611677104.

CHAPTER 7: DEVELOPING THE KNOWLEDGE MANAGEMENT SYSTEM USING A DATA LIFE CYCLE MODEL: THE IVHM CENTRE CASE.

(Parts of this paper are reformats of the extended abstract published at the *'Annual Conference of the Prognostic Health Management Society (PHM), 2018'.)*

Abstract: Data forms the foundation on which knowledge is created, captured, used and shared. The lack of an approach consistent with technological changes and needs can facilitate the loss of knowledge and increased costs. Integrated Vehicle Health Management (IVHM) is characterized by prognostics and diagnostics, which depend heavily on high-quality data to perform data-driven, model-based and hybrid computational analysis of asset health. As a result, managing data and knowledge for Integrated Vehicle Health Management (IVHM) requires a data life cycle model that adopts the OSA-CBM data model and integrates with other approaches. This project will propose such a model and use it to support the development of an IVHM Knowledge Management System.

7.1 Introduction

The IVHM Centre at Cranfield University has existed since 2008 and runs an active research program that generates high-value knowledge through theoretical and experimental work. The Centre produces a wide range of algorithms and processes to capture and analyse data from experiments carried out on rigs and other facilities. The Centre has worked on more than 40 projects, generating more than 120 technical papers, publishing 6 books, and generating lots of algorithms and experimental data, from 10 experimental rigs (Skaf and Jennions, 2017). With this growth in the creation of research data, data automating algorithms, technical papers, reports and theses, the Centre needs both a data management model and a Knowledge Management System that facilitate the storing, organizing and sharing of its research outputs and which is secure and scalable with a high level of cross-platform transferability.

However, the IVHM Centre does not yet have a systematic and coherent approach to IVHM knowledge and data management. A Knowledge Management System is therefore needed to improve the short and long-term transferability of the Centre's range of research projects and the outputs from these projects. A system that presents information to enable browsing by a range of categories and searching on specific terms is desired.

The OSA-CBM framework is based on the concept of metadata and interoperability that requires persistent visibility and traceability of data (Choudhary, Perinpanayagam and Butans, 2016) within and across diverse platforms, systems and devices, and therefore makes data provenance a fundamental requirement (Maindze, 2018).

7.2 Relationship DLCM, KM and KMS

The most common mechanisms for organisational knowledge management and implementing Knowledge Management System are company knowledge portals (Chua, 2004). Well-developed technology infrastructure and a systematic process of knowledge management are some of the fundamental criteria for success in organisational knowledge management (Kuan, 2005). The development of such infrastructure that is fit for purpose and ensures that long term availability of data requires the integration of the data life cycle model. The effective creation, managing and sharing of knowledge requires a Knowledge Management System. As has been revealed by the literature review, an effective creation, retention or increase in value in knowledge management requires an understanding of the relationship Data, information, knowledge and wisdom in Figure 7.1.

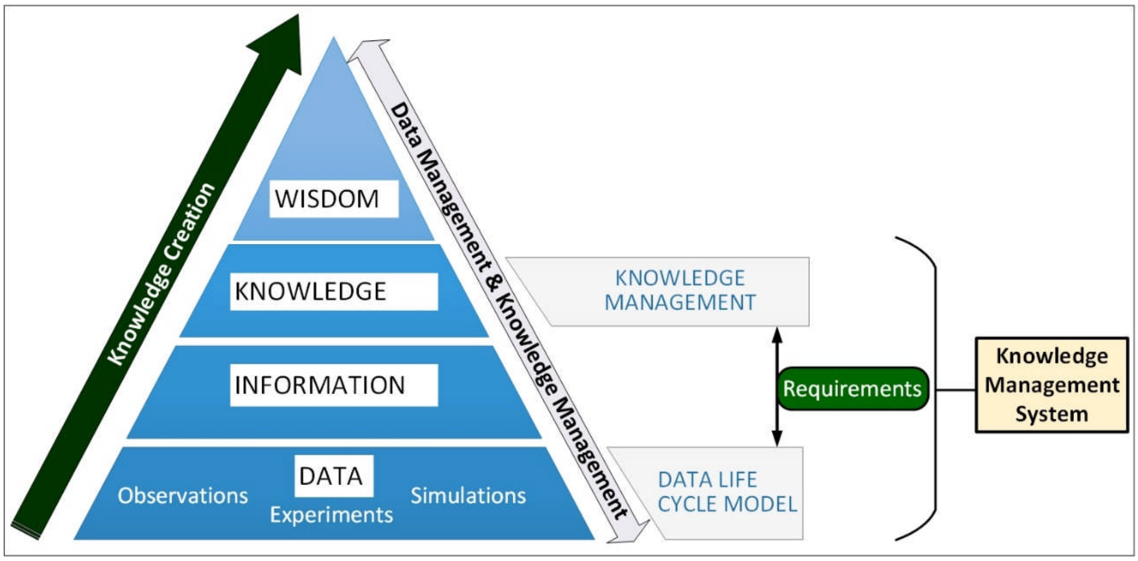


Figure 7-1 Relationship Knowledge Management KMS and the DLCM

The scoping literature review revealed that the foundations of all knowledge: Data usually acquired through observations, experiments or simulations requirements careful management throughout its life cycle. As the pyramid demonstrates, data is present and required at all stages of the knowledge management pyramid. This data can take on different forms depending on what level it is on the pyramid. Raw unprocessed data exists at the bottom of the pyramid. As it gets processed to create information and knowledge through the knowledge creation process, derived data, metadata and tacit observations are created. These observations which may be presented as knowledge artefacts in themselves can be seen as metadata which is transformed into further knowledge in the processes of organic and systematic reviews (Maindze, Skaf and Jennions, 2019).

As a result, two major requirements determinants for a Knowledge Management System then become the ‘data life cycle model’ and the ‘knowledge management’ layer. The data life cycle model considers data throughout its life and the various transformations that occur as it moves up the knowledge creation pyramid. The knowledge layer, on the other hand, caters for transformed data that is, knowledge which requires a different approach or requirement for its management. Therefore, data management requirements are then defined using the Data Life Cycle Model, which provides a comprehensive representation of process and data states. The knowledge layer informs the requirements for managing knowledge, derived data and metadata.

The integration of these two determinants further clarifies the distinction between two Knowledge Management System architectures. One before (as-is) the integration of the DLCM and the other derived (new) after integration.

7.3 KMS Architecture Before

At the start of the project, there was an R0 MSC KMS Requirements. The requirement analyses produced the functional specifications that were summarised at a high level as R1 MSc KMS Requirements (R1) (Appendix G). These requirements lead to the piloting of a Knowledge Management System (KMS) version 1.0 based on the SharePoint 2016 platform. The aim of this pilot was to understand the data and knowledge flows and the extent to which IVHM data and knowledge management requirements were met via the initial sets of requirements. The pilot was based on a 3-tiered Knowledge Management System architecture depicted in Figure 7.2 below, which comprises infrastructure services, knowledge services and presentation services (Chua, 2004).

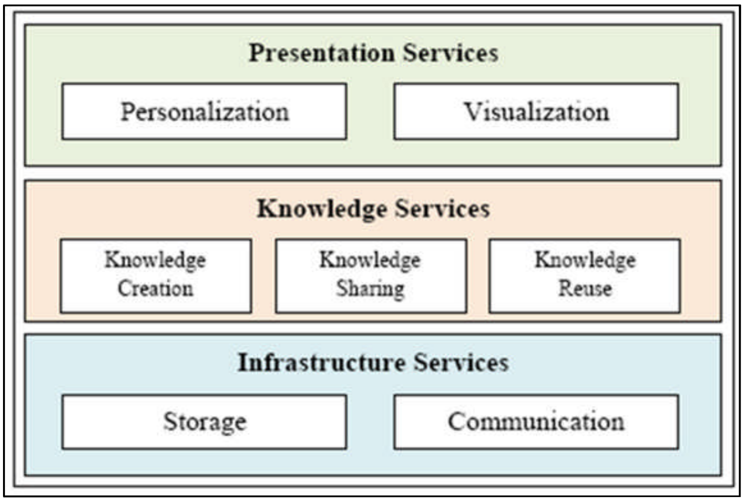


Figure 7-2 3-Tiered Architecture (Chua, 2004)

The infrastructure services layer focuses on knowledge repositories, communication, collaboration, workflows, technology platform and features (Chua, 2004). This KMS architecture model does not sufficiently or explicitly support that data life cycle model. The integration of data- and knowledge management requirements lead to requirements R3 and creation of an IVHM KMS hybrid architecture explained in the next section.

7.4 KMS Architecture After

A consideration of the data life cycle model in the process of requirements gathering and KMS development lead to the creation of the Hybrid KMS architecture Figure 7.3 below. The architecture provides a physical layer representing the data sources that feed into the technologies (applications) and knowledge services layer. These two layers are flanked by the data management layer and the data life cycle model. The data management and data life cycle model here represent both a physical and a theoretical implementation. This means they exist as processes and as procedures for data management and for supporting the creation and management of systems requirements.

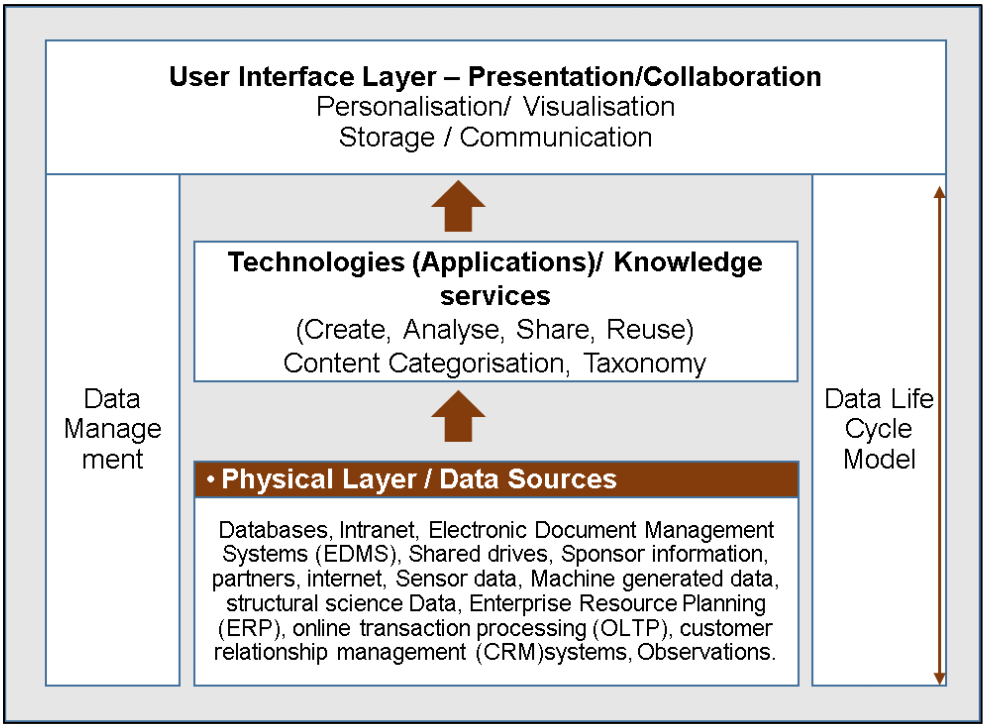


Figure 7-3 Hybrid KMS Architecture

This architecture results from the integration of data and knowledge management requirements R1 + R2 (Appendix H) through which emerging conflicts are resolved (Appendix I). The success of this integration was a function of the validation of the data life cycle model using various validation and verification techniques including peer reviews through paper publication, conference presentations and surveys.

7.5 KMS Version 2.0 - SharePoint 2016

The new version of the KMS is further created using SharePoint 2016. Any other platform or system could have been used because the choice of what tool to use has little or no impact on the application of the model. Using or trying to use any other tool would entail increased financial costs, high clerical overheads and require retraining of staff and researchers. The department has been using earlier versions of SharePoint since 2008, and the University had recently deployed SharePoint 2016 and therefore adopting it means reduced cost of implementation and training. SharePoint 2016 offered a familiar interface to staff and students and therefore reduces cost, time and level of training. The sections below depict some of the key features of the implementation. The design of the KMS took into consideration the diverse character of the Centre's research and engineering activities exemplified in Figure 7.4 below.

Figure 7-4 IVHM Research Facilities

The IVHM centre performs experimental, observation and simulation activities (Figure 7.5). These activities do include elements of social science research addressing the ‘human elements’. Datasets are published in various formats for medium to long term accessibility.

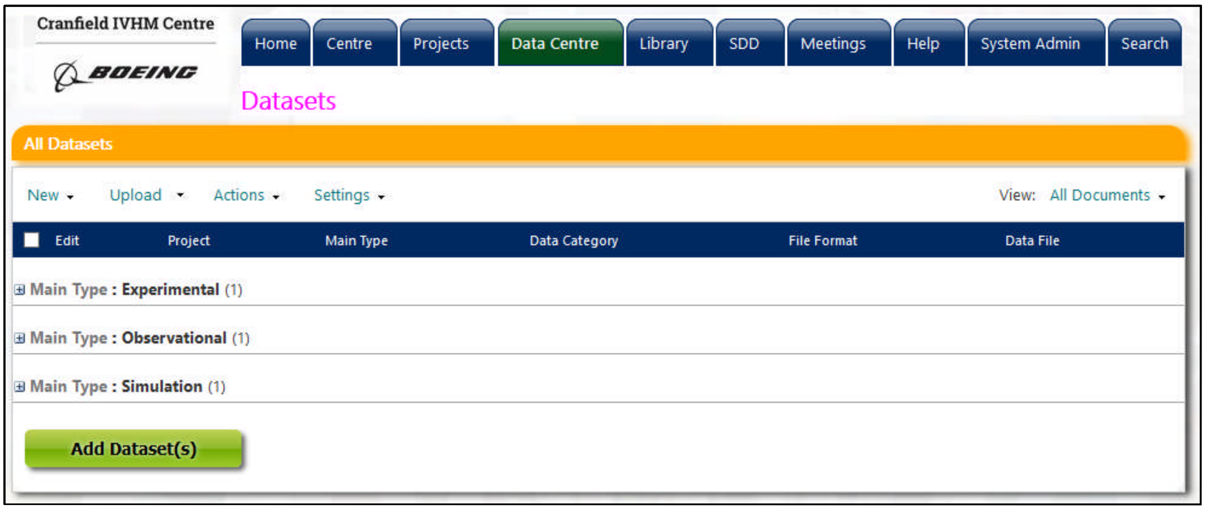


Figure 7-5 Data Capture and Management Library

The taxonomy built as a result of deploying the Data life cycle model as a starting point for devolving the KMS is shown below in Figure 7.6. It covers security classification, data category, datafile, document type and formats results from the metadata strategy (Appendix F).

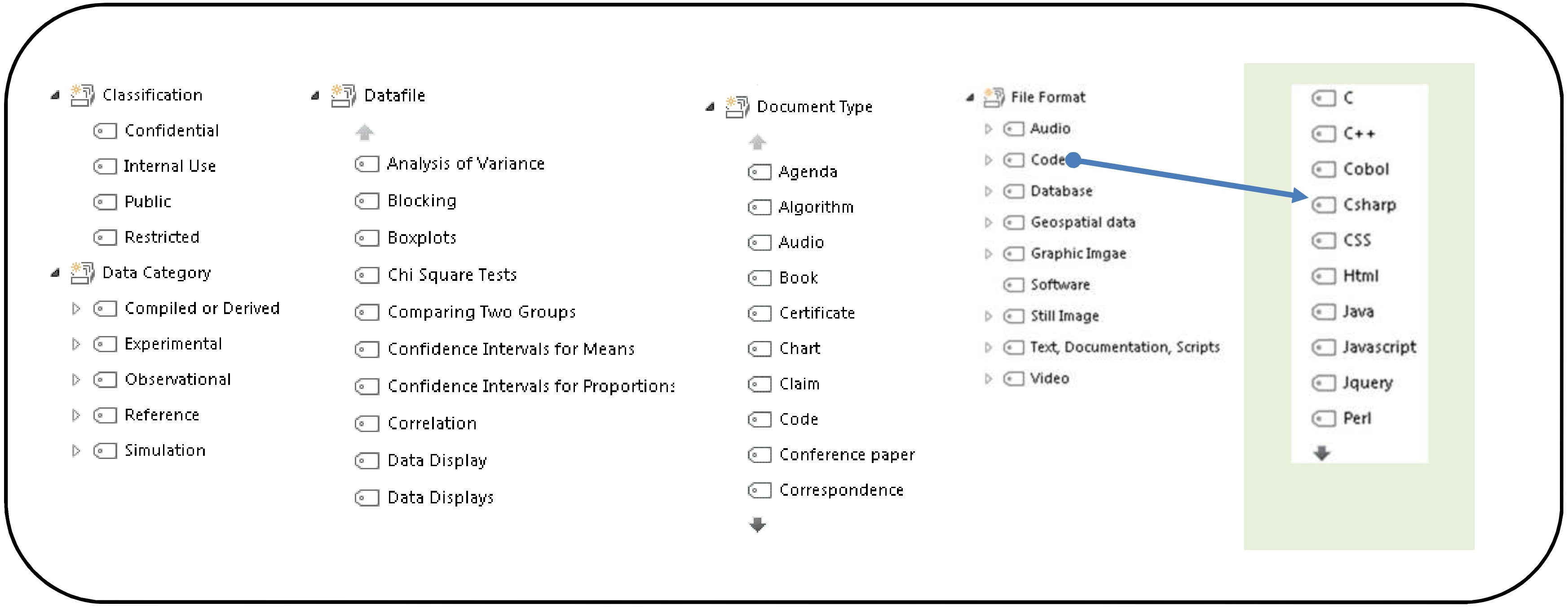
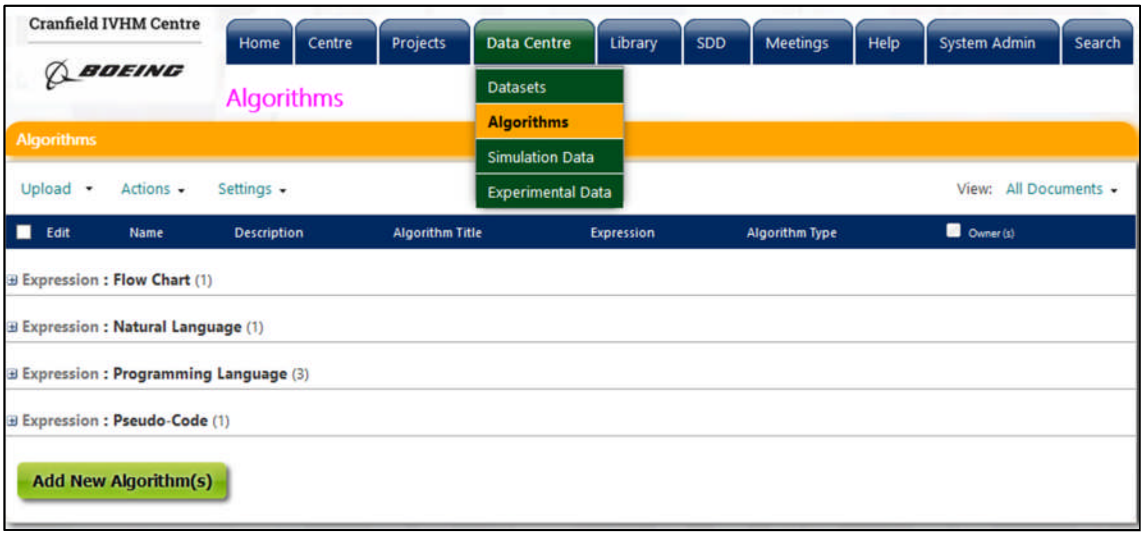


Figure 7-6 KMS Version 2.0 Data Taxonomy

The following shows an example (Figure 7.7) of algorithm capture library and some of the metadata that is captured. In addition, the elements specified in the taxonomy (Figure 7.6), information about the platform as well as specific tools used in the data creation process are captured.



Algorithms

Algorithms

Edit	Name	Description	Algorithm Title	Expression	Algorithm Type	Owner (s)

Expression : **Flow Chart** (1)

Expression : **Natural Language** (1)

Expression : **Programming Language** (3)

Expression : **Pseudo-Code** (1)

Figure 7-7 Data Capture and Management Library

Furthermore, the Knowledge Management System also captures details about the centre's publications (Figure 7.8).

Education and Publications

The World of Civil Aerospace, 2019

Hybrid Prognostic Methodology

AeroTech Seattle, September 2015

Published October 2014

Published October 2013

Published at AeroTech, Montreal, September 2013

Launched at AeroTech, Toulouse, October 2011

Published November 2012

Figure 7-8 Capture and Documentation of Publications

A search engine (Figure 7.9) is configured for search and retrieval in addition to other system navigation and functionality.

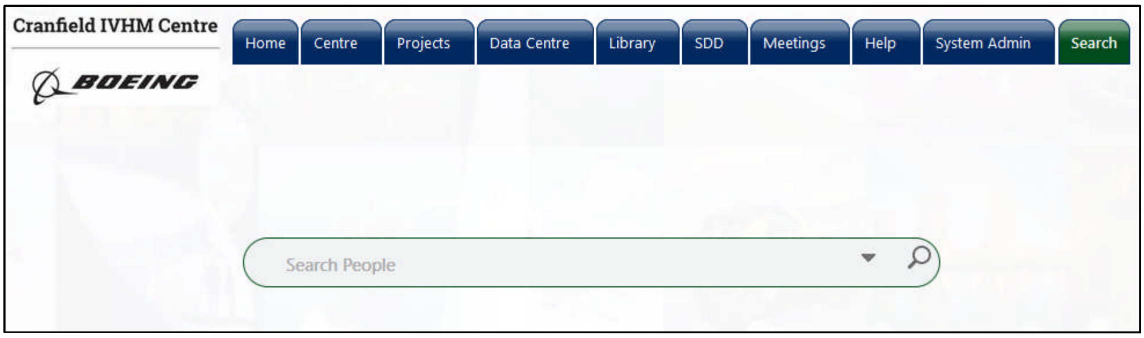


Figure 7-9 Search and Retrieval

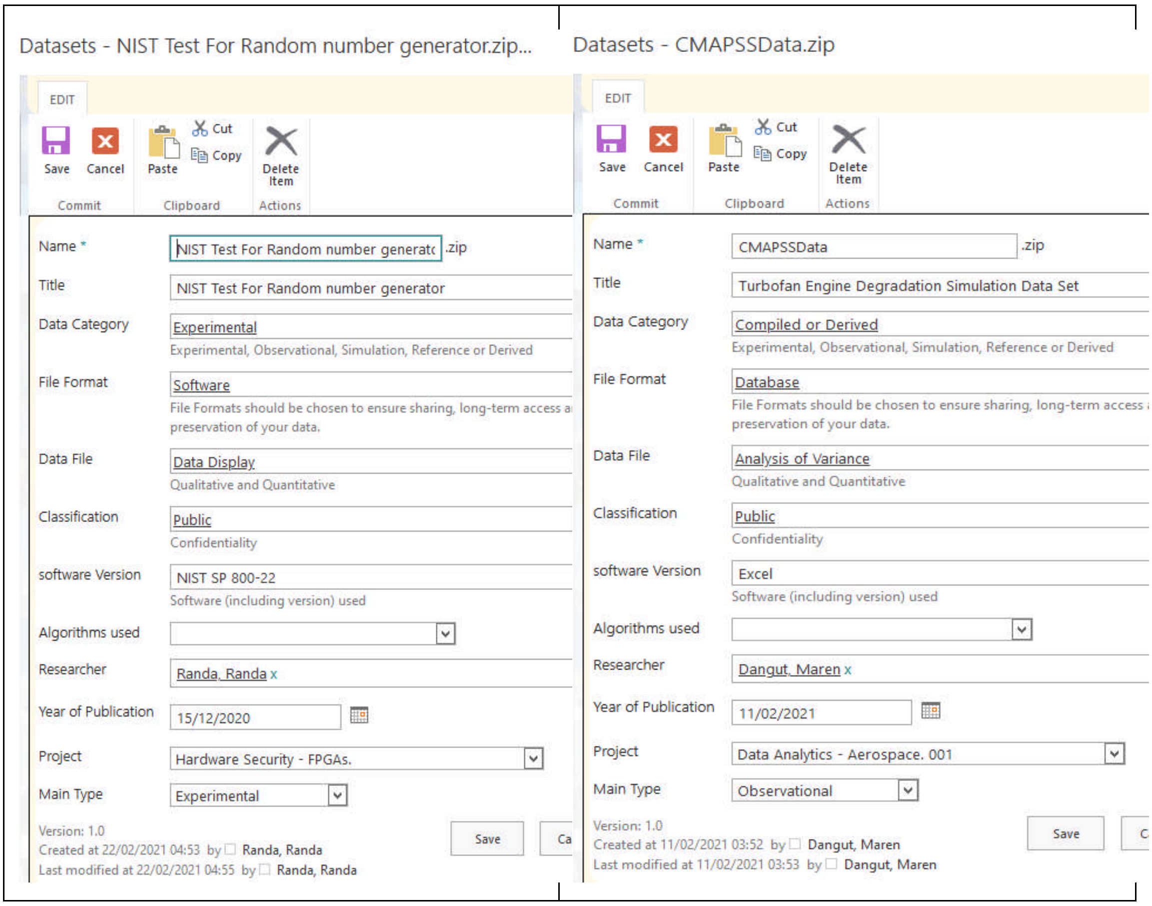
7.6 USE Cases

The following table list (Table 7.1) key actors and use cases of the IVHM Data- and Knowledge Management System (KMS).

Table 7.9 Use Case of IVHM Data- and KMS

Use Case	Data Sets Required	Category	Actor
Add content to KMS	N/A	All	All Stakeholders (managed)
Search Data/projects Store	All Data	All	Phd Students, Researchers Fellows, Lecturers, Engineers (managed)
Enterprise Data Lifecycle Management	All	All/data availability, capability	Data creators, Data Managers, Data Governors' Data curators
Enterprise Data Quality Management	All	Quality, Access, Costs & Security	Information Mangers/Data Curators
Improved maintenance planning and reduced maintenance costs	Maintenance Data + other Data	Costs	Field Engineers
Improved from combining maintenance schedules with other data	Maintenance Data + Assets + other	safety	Maintenance managers and Engineers
Making new generation of data professional interested in IVHM	Existing Data + Asset	Strategy	Data Scientists and Model Engineers

Video Games and IVHM simulation apps – Digital twin	Existing data + Asset	Commercial	Data Scientists and Model Engineers
Improved capacity planning using actual asset capacity data	Consist data	Strategy	Maintenance managers and Engineers
Improved and easier analytics (prognostics/diagnostics)	Historical data		Data Scientists and Model Engineers
Analytical Applications	Historical + Consistent data	Strategy & Costs	Application Developers, Model Engineers and Build Managers
Enterprise Master Data Management	Consistent Data	Accessibility/Quality	Enterprise/ Solution Architects, Data Curators
Enterprise Data Quality Management	Consistent data	Quality/Costs/Safety	Information Managers



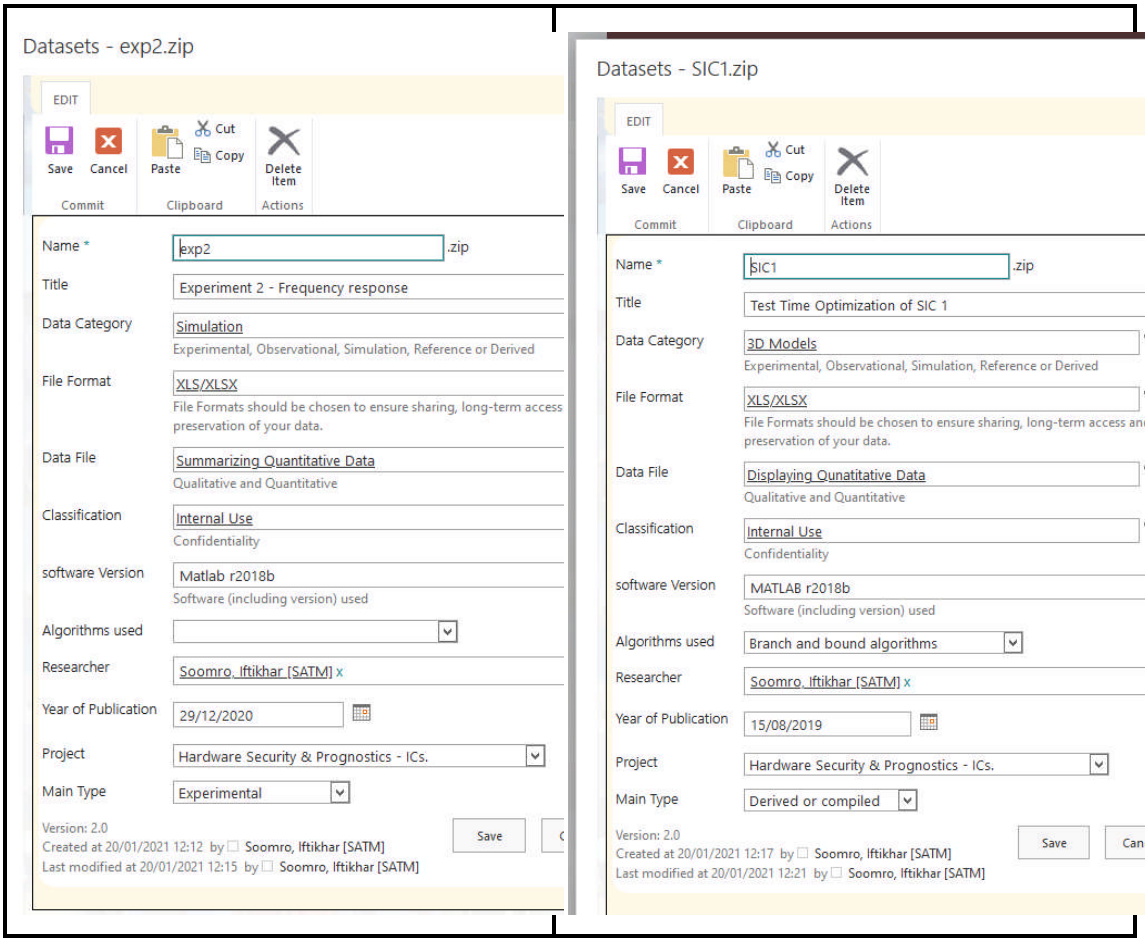
Datasets - NIST Test For Random number generator.zip...

Name *: NIST Test For Random number generatc .zip
Title: NIST Test For Random number generator
Data Category: Experimental
Experimental, Observational, Simulation, Reference or Derived
File Format: Software
File Formats should be chosen to ensure sharing, long-term access a
preservation of your data.
Data File: Data Display
Qualitative and Quantitative
Classification: Public
Confidentiality
software Version: NIST SP 800-22
Software (including version) used
Algorithms used:
Researcher: Randa, Randa x
Year of Publication: 15/12/2020
Project: Hardware Security - FPGAs.
Main Type: Experimental
Version: 1.0
Created at 22/02/2021 04:53 by ☐ Randa, Randa
Last modified at 22/02/2021 04:55 by ☐ Randa, Randa

Datasets - CMAPSSData.zip

Name *: CMAPSSData .zip
Title: Turbofan Engine Degradation Simulation Data Set
Data Category: Compiled or Derived
Experimental, Observational, Simulation, Reference or Derived
File Format: Database
File Formats should be chosen to ensure sharing, long-term access
preservation of your data.
Data File: Analysis of Variance
Qualitative and Quantitative
Classification: Public
Confidentiality
software Version: Excel
Software (including version) used
Algorithms used:
Researcher: Dangut, Maren x
Year of Publication: 11/02/2021
Project: Data Analytics - Aerospace. 001
Main Type: Observational
Version: 1.0
Created at 11/02/2021 03:52 by ☐ Dangut, Maren
Last modified at 11/02/2021 03:53 by ☐ Dangut, Maren

Figure 7-10: system use for compiled and experimental Data categories.



Datasets - exp2.zip

Name *	exp2 .zip
Title	Experiment 2 - Frequency response
Data Category	Simulation Experimental, Observational, Simulation, Reference or Derived
File Format	XLS/XLSX File Formats should be chosen to ensure sharing, long-term access preservation of your data.
Data File	Summarizing Quantitative Data Qualitative and Quantitative
Classification	Internal Use Confidentiality
software Version	Matlab r2018b Software (including version) used
Algorithms used	
Researcher	Soomro, Iftikhar [SATM] x
Year of Publication	29/12/2020
Project	Hardware Security & Prognostics - ICs.
Main Type	Experimental

Version: 2.0
Created at 20/01/2021 12:12 by ☐ Soomro, Iftikhar [SATM]
Last modified at 20/01/2021 12:15 by ☐ Soomro, Iftikhar [SATM]

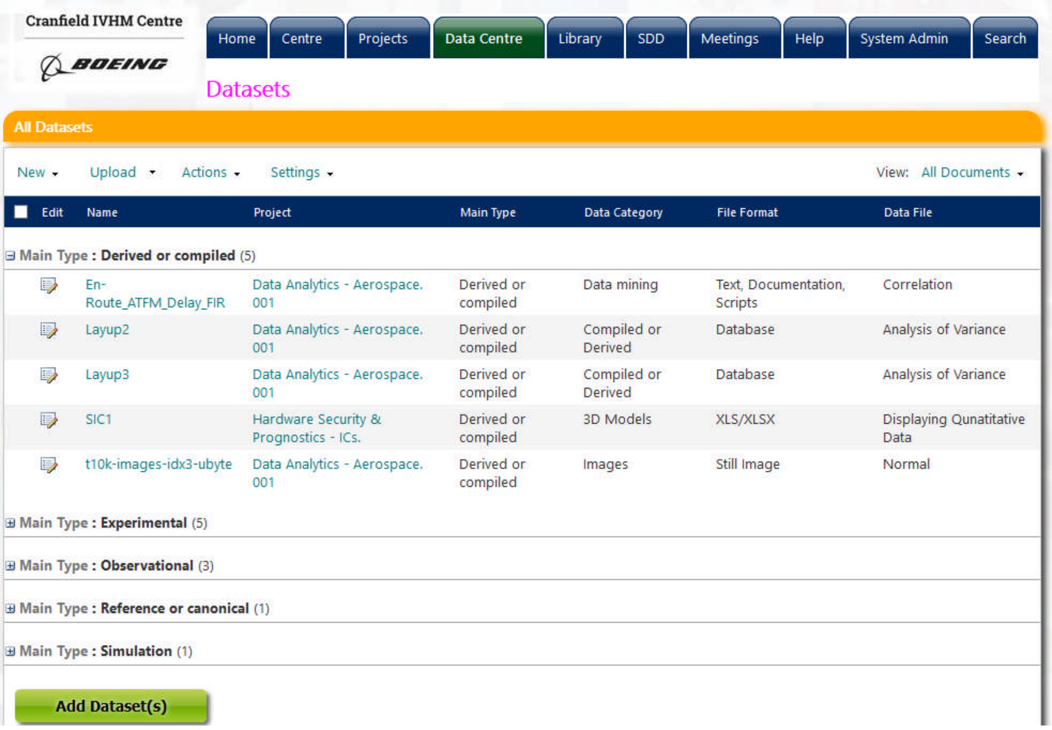
Datasets - SIC1.zip

Name *	SIC1 .zip
Title	Test Time Optimization of SIC 1
Data Category	3D Models Experimental, Observational, Simulation, Reference or Derived
File Format	XLS/XLSX File Formats should be chosen to ensure sharing, long-term access an preservation of your data.
Data File	Displaying Qunatitative Data Qualitative and Quantitative
Classification	Internal Use Confidentiality
software Version	MATLAB r2018b Software (including version) used
Algorithms used	Branch and bound algorithms
Researcher	Soomro, Iftikhar [SATM] x
Year of Publication	15/08/2019
Project	Hardware Security & Prognostics - ICs.
Main Type	Derived or compiled

Version: 2.0
Created at 20/01/2021 12:17 by ☐ Soomro, Iftikhar [SATM]
Last modified at 20/01/2021 12:21 by ☐ Soomro, Iftikhar [SATM]

Figure 7-11: System Use for Simulation and 3D Models.

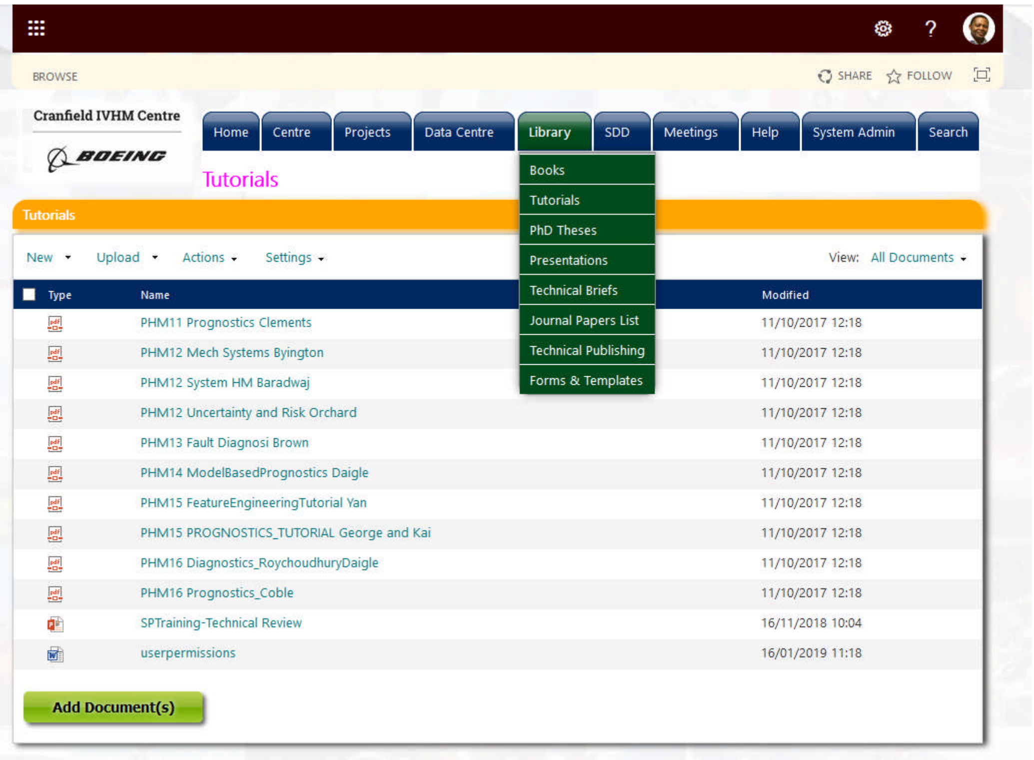
The above Figure 7.10 -11 below shows the shows the dataset creation and metadata capture form. Figure 7.11 shows the presentation of the created data sets.



Edit	Name	Project	Main Type	Data Category	File Format	Data File
Main Type : Derived or compiled (5)						
	En-Route_ATFM_Delay_FIR	Data Analytics - Aerospace. 001	Derived or compiled	Data mining	Text, Documentation, Scripts	Correlation
	Layup2	Data Analytics - Aerospace. 001	Derived or compiled	Compiled or Derived	Database	Analysis of Variance
	Layup3	Data Analytics - Aerospace. 001	Derived or compiled	Compiled or Derived	Database	Analysis of Variance
	SIC1	Hardware Security & Prognostics - ICs.	Derived or compiled	3D Models	XLS/XLSX	Displaying Qunatitative Data
	t10k-images-idx3-ubyte	Data Analytics - Aerospace. 001	Derived or compiled	Images	Still Image	Normal
Main Type : Experimental (5)						
Main Type : Observational (3)						
Main Type : Reference or canonical (1)						
Main Type : Simulation (1)						

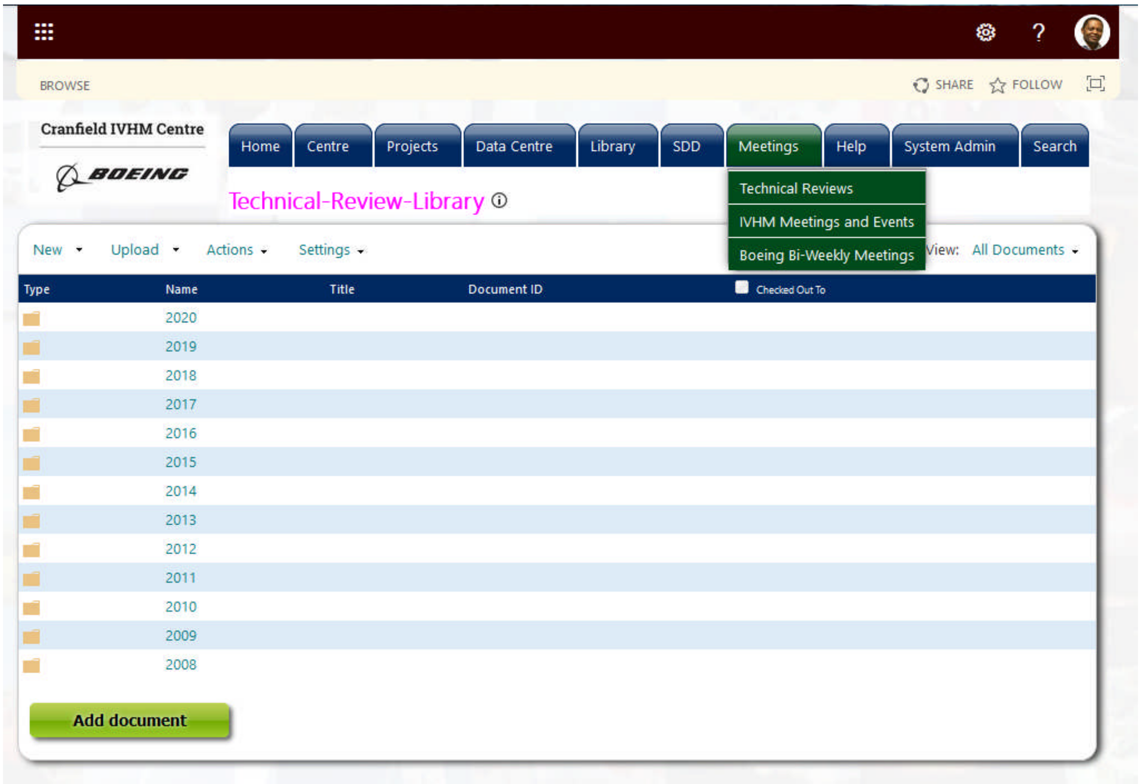
Figure 7-12: System Use for User Interface View.

The following figure shows the use of the KMS for sharing of study materials.



Type	Name	Modified
pdf	PHM11 Prognostics Clements	11/10/2017 12:18
pdf	PHM12 Mech Systems Byington	11/10/2017 12:18
pdf	PHM12 System HM Baradwaj	11/10/2017 12:18
pdf	PHM12 Uncertainty and Risk Orchard	11/10/2017 12:18
pdf	PHM13 Fault Diagnosi Brown	11/10/2017 12:18
pdf	PHM14 ModelBasedPrognostics Daigle	11/10/2017 12:18
pdf	PHM15 FeatureEngineeringTutorial Yan	11/10/2017 12:18
pdf	PHM15 PROGNOSTICS_TUTORIAL George and Kai	11/10/2017 12:18
pdf	PHM16 Diagnostics_RoychoudhuryDaigle	11/10/2017 12:18
pdf	PHM16 Prognostics_Coble	11/10/2017 12:18
P	SPTraining-Technical Review	16/11/2018 10:04
W	userpermissions	16/01/2019 11:18

Figure 7-13: Learning Materials (user interface view)



Type	Name	Title	Document ID	☐ Checked Out To
📁	2020			
📁	2019			
📁	2018			
📁	2017			
📁	2016			
📁	2015			
📁	2014			
📁	2013			
📁	2012			
📁	2011			
📁	2010			
📁	2009			
📁	2008			

Figure 7-14: Events and Meetings

The IVHM KMS also provides a section for capturing details of events and meetings. It holds details of presentations, proposals, agendas, minutes and other content arising from the events.

7.7 Conclusion

The purpose of a data life cycle model is to inform the management of data throughout its life from creation to disposal. It provides the requirements specification for various stages of the data lifecycle through the provision of processes and practices required to ensure quality, reliability, security and provenance of data.

This implementation supported this view and understanding of the data life cycle model in improving the requirements specification for data management infrastructure, data science and data curation. It provides the foundation for review and adoption of data and knowledge management tools, including best practice guidelines, policies and procedure.

The use of the model in this implementation demonstrated the value of the data life cycle in requirements gathering and specification for systems development, Research and engineering work. This approach also shows huge potential in costs reduction and projects implementation.

7.8 References

Choudhary, R., Perinpanayagam, S. and Butans, E. (2016) ‘Design and Analysis of Communication Model for Implementation of CBM Systems Based on OSA-CBM Framework’, in 2016 IEEE Aerospace Conference, pp. 1–7. doi: 10.1109/AERO.2016.7500821.

Chua, A. (2004) ‘Knowledge Management System Architecture: A Bridge Between KM Consultants and Technologists’, International Journal of information management. Elsevier, 24(1), pp. 87–98.

Kuan, Y. W. (2005) ‘Critical Success Factors for Implementing Knowledge Management in Small And Medium Enterprises’, Industrial Management & Data Systems. Emerald Group Publishing Limited, 105(3), pp. 261–279. doi: 10.1108/02635570510590101.

Maindze, A. (2018) ‘Developing a Knowledge Management System for Integrated Vehicle Health Management Using a Data Life Cycle Model’, in Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM.

Maindze, A., Skaf, Z. and Jennions, I. (2019) ‘Towards an Enhanced Data- and Knowledge Management Capability : A Data Life Cycle Model Proposition for Integrated Vehicle Health Management’, in Proceedings of the Annual Conference of the PHM Society, pp. 1–14. doi: 10.36001/phmconf.2019.v11i1.842.

Skaf, Z. and Jennions, I. (2017) Knowledge Management System for the IVHM centre.

CHAPTER 8: DISCUSSION AND CONTRIBUTION

8.1 Introduction

This research aimed primarily at developing a Knowledge Management System for Integrated Vehicle Health Management (IVHM), using an appropriate data life cycle model. The primary phase of the IVHM research, or engineering, is data collection using methods such as Simulations, Observations, Derivations, Experiments and Referencing (SODER). The effective and efficient management of this data is imperative for its medium and long-term activities. A data life cycle model provides the steps and principles for consideration in data management activities from planning, acquisition, pre-processing, processing analysis, preservation, and dissemination and archiving. Fundamental to these activities and for ensuring long-term data availability and provenance are quality, integrity and security, supported by robust metadata capture. Therefore, the systems or tools that sustain these activities need to be initiated through the integration of the data life cycle model as a first step. Thus, an IVHM Knowledge Management System is developed on the basis of the data life cycle model, presenting a new approach to enhancing data and knowledge management.

This thesis focused on various elements of the data life cycle model as a foundation for IVHM data and knowledge management and gives answers to the following queries:

- Which data life cycle model is suitable for IVHM data and knowledge management?
- How can IVHM data management requirements for medium to long-term data availability be addressed?
- What can be done to guarantee, the security, reliability and quality of IVHM data?

Where permissible, and indication of future work or application to be carried out for more insights will be given.

8.2 Recapitulation of Research Aim and Objectives

Objective 1. To review the literature on data/ knowledge management and data life cycle models.

This objective was attained through a combination of a systematic review as main method and traditional literature review as scoping review. The use of systematic reviews is the reproducibility of both the approach and results. In Chapter 2, we provide a definition of what it is that we called knowledge management and introduced and the concept of ‘intellectual paradox’. The definition operationalised Knowledge Management in the context of this research whilst at the same time addressing the philosophical gap in the literature. Inconsistencies in the history of Knowledge Management (Chapter 3) are presented and discussed. Chapters 2 and 3 are outcomes of the scoping study; the first step in the systematic reviews which aims at developing the protocol for implementing the systematic review.

The interest of IVHM researchers and engineers is in a model that encapsulates a cross-section of relevant standards and reflects their activities. In Chapter 4, we established the deficiency of existing models in addressing IVHM data and knowledge management challenges. By reviewing relevant standards, connection to existing data life cycle models and their implication for IVHM data and knowledge management sets the foundation for addressing its data and knowledge management challenges. Understanding the presence and or absence of these linkages will be invaluable in developing the IVHM data life cycle model.

Summarising, succeeded in determining “Which data life cycle model is suitable for IVHM data and knowledge management?” and “How can IVHM data management requirements for medium to long-term data availability be addressed?” Contextualised knowledge management internally and externally whilst addressing its inconsistent historical narrative. Succeeded in creating the protocol, implemented the protocol and determined the criteria for a suitable data life cycle model for IVHM.

Objective 2. To develop and validate an appropriate data life cycle model for IVHM.

The realisation of this objective was paramount to the success of this research. The objective represents the foundation of this research effort. This objective was achieved by integrating key elements of essential standards and models to create a hybrid data life cycle model (Chapter 5). The data sources, creation methods and categories of IVHM research and engineering were imperative to developing this model. Furthermore, the peer review and publication of the model (Chapter 5 and 6) asserted the capability of the model in addressing IVHM data management challenges as well as specific prognostic health management issues. Further verification and validation of the model is carried out through the design and implementation of surveys (Chapter 6) that showed 84% approval.

This approach to developing and validating the data life cycle model represents proves to be a reliable technique for non-mathematical, non-statistical conceptual models.

Objective 3. To use the data life cycle model as a foundation to build a Knowledge Management System for IVHM.

This objective is achieved by implementing the data life cycle model as a framework for requirements analysis and for designing the Knowledge Management System architecture. This was then followed by the actual development of the system (Chapter 7), based on an existing platform. The data lifecycle model provided a high-level summary of the specific actions and processes that must be undertaken at different phases of data management. The model informed the integration of the basic tenets and practices into the requirements gathering process. Noteworthy is the fact that the IVHM data life cycle model comprises a physical and a theoretical implementation. The most relevant for this work was the theoretical implementation. This means it guided the overall requirements gathering and analysis process. The model assists in creating the requirements that ensure the quality, integrity, security both data and well-curated data resources.

8.3 Contribution

8.3.1 Data and Engineering Management

The research proposes a scalable, reconstituted and integrated data life cycle model applicable to IVHM research and IVHM engineering practices. These include multidisciplinary activities from structural sciences to social sciences, covering various types of research as well as varied project sizes. It provides a reference for enhanced and integrated short to long-term requirements specification, data management and systems development. The verification and validation of the model through peer review, presentation publication and expert surveys demonstrates its potential in addressing the various data and knowledge management challenges.

8.3.2 Methodology

The research has produced an approach for validating non-statistical, non-mathematical conceptual models. This approach combines peer review, face validity, structured walkthrough and surveys. This requirements gathering process that starts with the data life cycle model as a foundation is introduced by this research and has proved to facilitate the process.

8.3.3 Knowledge Management (KM): Improved Conceptual Definition

Prior knowledge management research shows that there is no universally agreed definition on knowledge management, and there is little or no distinction between knowledge management [Phrasal verb] and knowledge Management [phrasal noun] discipline. This lack of consensus is also reflected on the way the history of knowledge management has been captured in the literature to date. The history of KM in the literature is fragmented (Jasimuddin 2006; Spender 2015)(Maindze, Jennions and Skaf, 2017) and inconsistent on when it was or who created it. The research work addresses these fundamental challenges in knowledge management: lack of definitional consensus in knowledge management, the confusion that persists between “knowledge management [phrasal verb]” and “Knowledge Management [Phrasal noun]” as well as the lack of clarity on the constituent elements of knowledge management [phrasal noun].

8.3.4 Education and Training

The research provides academics, students and knowledge management professionals with a basis for designing study materials, organising their perceptions of knowledge management and the development of knowledge management strategies.

8.3.5 Impact

The primary theme of this project is multidisciplinary in nature and the knowledge generated by this research contributes to, benefits and influences data and knowledge management in both academia and industry. It contributes to the development of new perceptions of knowledge management and its origins and contributes to high quality applied research. It will allow data managers, curators, researchers and systems developers to consider the framework for commencing and maintaining their respective activities to the desired levels of efficiency and effectiveness.

The research provides an improved understanding of knowledge management and its history. This allows readers to learn more about the history of knowledge management and may help to improve the development of knowledge management courses and programme, as well as contribute to improved research data and knowledge management.

The proposed model furthermore the potential to facilitate reduced costs, reduced data loss, increased data quality, security and long term data availability through improved planning, acquisition, processing, storage and dissemination. In this way, the model may aid to improve both the systems development and research processes.

8.4 Limitations and Future Research

This study has its own limitations. It focuses on DLCM model for IVHM and validated only in one non engineering setting- the IVHM research centre thereby imposing some limitations on the generalisability of the model. Because the proposed model is intended to be the foundational pillar for IVHM data management standards, tools, and techniques, it must be extensively reviewed. It has already been reviewed in academic research setting. Additional formal reviews are necessary in purely engineering settings or other research facilities of variable sizes that accurately describes how asset data in IVHM is managed from creation to deprecation. Furthermore, to develop strategies for advanced operational analytics based on big data for prognostics, an investigation should consider the Batch, Near-Real-Time, and Real Time Ingestion components for IVHM data types to compare the costs of the respective components with resultant system performance.

In addition IVHM activities are generating large amounts of data (Big Data) throughout the life span of an asset. Big Data encompasses a variety of processes, tools and data requirements, which means high speed (near-real-time ingestion) and handling very large data volumes (more than several terabytes). Understanding IVHM data provisioning for big data and data warehouse is critical for data-driven prognostic health management, as they compute different components of Big Data. Further research is needed into the impact of this computational variations on prognostics health management.

Another direction is to study IVHM Data Provenance. In order to verify the history and authenticity of an item, the provenance of the item is needed. However, source and transformation provenance are not completely isolated from each other, so studying under what circumstances it is possible to change one into the other and how much redundancy is introduced by storing both would be interesting.

Although the model and application has been guided by the V-model, consideration should be given to Agile and other approaches to validate the scalability and flexibility of the DLCM.

Further research into the verification and validation of non-statistical, non-mathematical models to strengthen the approach use in this research. Application of the IVHM model in non-research based, non-engineering environments.

8.5 References

Maindze, A., Jennions, I. and Skaf, Z. (2017) ‘Knowledge Management Yesterday and Tomorrow: Exploring an “intellectual paradox”’, in James Gao, Mohammed El Souri, S. K. (ed.) Advances in Transdisciplinary Engineering. Amsterdam: IOS Press, pp. 522–527. doi: 10.3233/978-1-61499-792-4-522.

Spender, J. C. (2015) ‘Knowledge Management: Origins, History, and Development’, in Bolisani, E. and Handzic, M. (eds) Advances in Knowledge Management: Celebrating Twenty Years of Research and Practice. Cham: Springer International Publishing, pp. 3–23. doi: 10.1007/978-3-319-09501-1_1.

APPENDIX

Appendix A Ethical approval

12 July 2019

Dear Mr Nyuyfoghan Maindze ,

Reference: CURES/8806/2019
Title: Algorithmic Data- and Knowledge Management Approach for integrated Vehicle Health Management

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/8806/2019 **has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

Page 1 of 1

Appendix B Survey

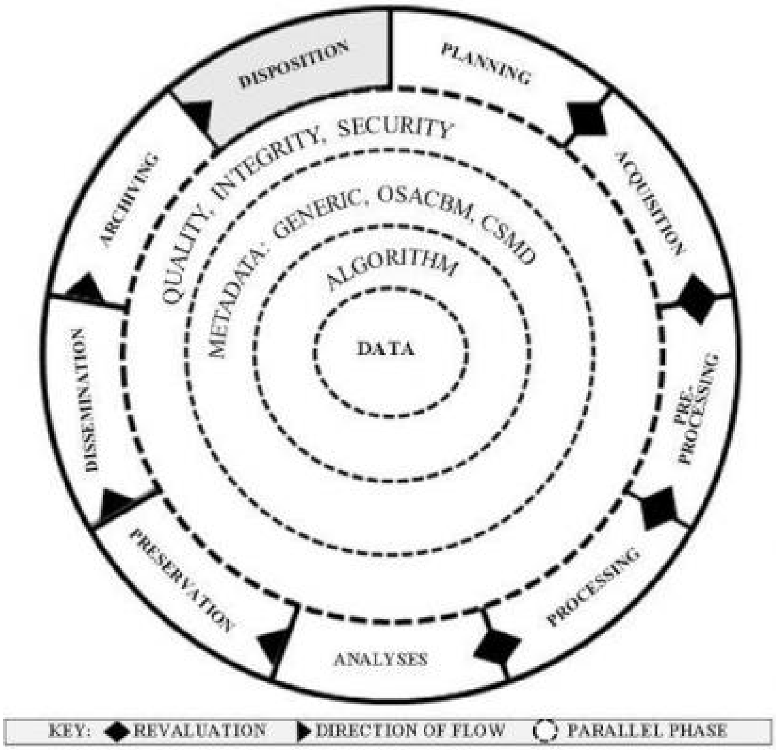
A Data Life Cycle Model Proposition for Integrated Vehicle Health Management.

Start of Block: Default Question Block

THE INTEGRATED VEHICLE HEALTH MANAGEMENT DATA LIFECYCLE MODEL

The proposed IVHM Data lifecycle model (IVHM-DLCM) is a hybrid that integrates relevant standards, frameworks and models that fit the profile of IHVM research and engineering activities. The IVHM-IVHM Data lifecycle model is scalable and intended for use in diverse IVHM activities of all depths and breadths. It takes into consideration interoperability, integrity, quality, security, provenance and preservation of data throughout the lifecycle. The IVHM-DLCM has nine discrete phases with three other phases running parallel. There is also the revaluation process that runs parallel to phases One, Two, Three and Four.

A data life cycle model is “*…a formal representation of all the possible states and all the valid state transitions of a data item, when handled by a particular system or by an user application, e.g. created, duplicated, deleted, backed-up.*” (Simonet et al., 2015:26)



Phase 1: Planning

The planning phase is the conception and beginning of simulations, observations, derivation, experiments and referencing (SODER) activities. Other activities at this phase include explorations, conceptualisation, design, manufacturing, testing, operation, service and maintenance. In this phase, the research or engineering task is defined and planned – resources and planned deliverables for each phase are explored and explained. The requirements for success, quality, integrity and security are defined, including systems requirements. The metadata parameters both generic and standard are considered in this phase. The data management plan, retention schedules and requirements specification are some of the recommended deliverables of this phase. In this phase, we also do the selection of the sensor modules and other accessories like software tools required for the SODER activity. File formats, data storage and sharing plans are developed at this stage.

Phase 2: Acquisition

The data acquisition phase represents data creation from scratch or the *extraction of existing raw or derived data for reuse. In this stage, observational, experimental, simulation data creation takes place as well as the retrieval of existing datasets, derived or reference data for reuse. Data governance standards and best practices ensuring integrity, security, quality and metadata are considered. The quality of the process is also vital as it has a direct impact on the quality of the data created of retrieved for reuse.*

Phase 3: Pre-Processing

The pre-processing phase aims to flag out of range data values, missing values to mitigate the risk of making decisions based on misleading results. It is fundamental to IVHM machine learning activities. It is used to clean the original signal by eliminating noise and to improve object component condition. In other words, this phase represents the low-level computation of sensor data and constitute a key element of the OSA-CBM architecture. Sensor data is transformed into an understandable format. This is the stage where data cleansing –detecting and correcting mistakes, incomplete, inaccurate, irrelevant and incorrect records from datasets. This step employs best practices to ensure that data is free of inconsistencies, correct, usable and reliable.

Phase 4 : Processing

In this phase meaningful and relevant is extracted in suitable formats from raw datasets created after pre-processing for future use. This step involves activities such as validation, aggregation, summarization, sorting, classification and validation. It includes the conversation of data in usable and desired forms and formats. It can take the form mechanical, manual or digital processing.

Phase 5: Analyses

This represents activities like organization, interpretation and presentation of data. It involves statistical data analytics, simulations, modelling and other computation activities that reveal trends, facts, faults and tests theories and assumptions. In this step raw data is transformed into information and communicated to the stakeholders.

Revaluation: Phase 1 – Phase 5

Revaluation is one of the most important tools in The Systems Engineering Process. Re-evaluation observation of outputs and using the information to modify the system, the inputs, the product or the process (INCOSE 2017). It takes place in the first five phases. Feedback is collected at each phase on a situational basis to continually improve the steps and eliminate problems. The loops are used specifically when issues are identified as data moves from one phase to the next in the lifecyle.

Phase 6: Preservation

Preservation includes steps and processes for active data storage for the duration that it might be needed. It also involves access control and backup for security. These are the security actions taken to reduce the chances of data corruption and data loss. They also include submission to reliable data repositories.

Phase 7: Dissemination

This step includes the preparation and dissemination of datasets, derived data as well as findings or outcomes to relevant stakeholder communities. It improves accessibility as well as being a recommended best practice.

Phase 8: Archive

This the last step before disposal. It represents the retraction of data and related outputs from active circulation. The retention schedule is applied to the data and only accessible on demand.

Phase 9: Dispose

The data has reached the end of its useful and potential useful life. It is purged at this stage. It is securely disposed of to ensure that there can be no unwarranted access.

PARALLEL PHASES

Parallel phase 1: *Metadata: Generic; Core Scientific* Metadata *model; Open System Architecture for Condition-Based Maintenance (OSA-CBM)*
The strategy for metadata generation and documentation is referenced throughout the data lifecycle. As data transitions from one phase to the next, it might also change platforms and devices making this phase particular important for addressing provenance issues. The metadata phase draws on the best generic metadata created by systems, the core scientific metadata model and the OSA-CBM built in metadata to create a scalable metadata model within the lifecycle. This allows for easy contraction and expansion of research and engineering activities. It accommodates small to large scale data generating activities.

Parallel phase 2: Data Integrity, Data Quality and Data Security
Data integrity involves creating, processing and maintaining the assurance, accuracy, consistency and completeness of data throughout its lifecycle. The content and meaning of data is maintained throughout its lifecycle. This also includes compliance with statutory requirements. Quality represents the use of best practice protocols and methods of collecting and organizing data that ensures its accessibility, completeness, validity, accuracy, consistency, availability and timeliness. Security involves the protection of data from unauthorized access to modify, use, delete and disclose. It includes protection against theft, breach of agreements, data protection laws and unintended or hateful modification. The computer system security, physical security and file security are all part of this step.

Parallel phase 3: Algorithms
Refers to the algorithms and agents that automate research and engineering tasks. They are maintained as datafiles and at the same time they represent agents within the data- and Knowledge Management System.

Q1 Participant Background - Work Designation

○ Engineer (1)
○ Manager (2)
○ Researcher (3)
○ Professor (4)
○ Lecturer or Senior Lecturer (5)
○ Designer (6)
○ Architect (7)
○ Analyst (8)
○ Curator (9)
○ Other (10)

Q2 **To what extent do you agree with the following statements?**

	Strongly Agree (1)	Agree (2)	Neither agree nor disagree (3)	Disagree (4)	Strongly disagree (5)
1 - This model accurately represent the data life cycle (1)	○	○	○	○	○
2 - The model is clear and easy to understand (2)	○	○	○	○	○
3 - The model layout is easy to read and pleasant to the eye (3)	○	○	○	○	○
4 - The model design sequence is accurate (4)	○	○	○	○	○
5 - The model enhances requirements specification (5)	○	○	○	○	○
6 - The terminology is clear and correct (6)	○	○	○	○	○

Q3 Identify facilitators to the model's implementation

____
____
____
____
____

Q4 How likely are you to recommend this data life cycle model to friend, colleague or organisation?

○ Extremely likely (1)
○ Somewhat likely (2)
○ Neither likely nor unlikely (3)
○ Somewhat unlikely (4)
○ Extremely unlikely (5)

Q5 Identify barriers to the model's implementation

____
____
____
____
____

Q6 **Additional observations**

____
____
____
____
____

End of Block: Default Question Block

Appendix C Question 3

A Data Life Cycle Model Proposition for Integrated Vehicle Health Management.

March 1st 2020, 5:10 am MDT

Q3 - Identify facilitators to the model's implementation

Identify facilitators to the model's implementation

This model is based on SharePoint at this stage, but I believe it can be used in the data centre of many major groups and organizations.

I am not quite clear what you are wanting here. Do you mean standards bodies, interested organisations such as INCOSE, MOD, major manufacturers and operators?

Yes

OEMs

This question is difficult to understand what you are asking.

University research project funded by external organisations.

The model is clearly articulated. The underlying data models are very relevant to IVHM philosophy.

Computational power and memory needs to be accounted for. The cycle of refresh rates and archiving for the live data could be included.

The facilitators will be part of a variety of stakeholders starting from operators, maintenance engineers, life cycle and cost engineers, strategy and innovation managers etc. On the academic side, it will be researchers working in the area of product life-cycle.

IVHM staff and participants in the project.

Industries, third parties, academics

industries, Academics and Government organizations

Qualified trained engineers, Software, Hardware, Time Management

Refer to the System Engineering V diagram if the goal is to represent requirements specification processes. Require more depth and descriptions.

Its similar to the spiral method already utilised in product development

Appendix D Question 5

A Data Life Cycle Model Proposition for Integrated Vehicle Health Management.

March 1st 2020, 5:09 am MDT

Q5 - Identify barriers to the model's implementation

Identify barriers to the model's implementation
Maybe will meet the challenge from some existing data management system. For a more professional perspective, it is unknown from my current level.
Probably a key barrier is getting the model adopted by a key player or organisation. I think you need both a simplified version - of the sort that would be on presentation slides or a poster - because what you have just now is quite a lot to take in - and a development where you map what the implications of each phase are for organisations.
Validation, data security and confidentiality
It's very high level; there isn't sufficient detail to actually implement.
A use case implementation may be helpful.
Loss of information between the stages
This is tricky as it is dependent on the organisation, and issues associated with data storing and sharing practices and the gate keepers who handle this.
In the course of Phase 7, parallel phase 2 Data Integrity, Data Quality and Data Security mayn't be complied with by all members on the project.
Missing information
lack of completeness
Poorly written Software, Inefficient Hardware.
Apply this to machine learning approaches, how does training fit in to the model? How about simulation and generation of data for machine learning?
Model induces vertigo when the eye tries to follow it. The word ALGORITHM is off centre Unfamiliar terms in model (OSACBM, CSMD) require scrolling back and forth to text and back to model.
lack the importance of an agile approach when applying model in enterprise environment, during continuous iteration of the lifecycle sometimes steps are skipped, need for a flexible approach throughout successive steps in cycle
There similarities in its approach may deter uptake as it could be seen as an unnecessary replacement

Appendix E Question 6

A Data Life Cycle Model Proposition for Integrated Vehicle Health Management.

March 1st 2020, 5:08 am MDT

Q6 - Additional observations

Additional observations
Null
You are effectively asking us to do "peer review" of a model. I don't see the context, the empirical grounding, nor how this is related to existing theories and models. I therefore cannot meaningfully evaluate it
Every project has unique peculiarities. So it is extremely hard to correlate all different fields in a single model. However, it is a very good approach for a generic architecture.
A flow chart description may help the reader easily understand the overall model.
Contingencies for backing up or recovering lost data required. There are some corrections required in the text. In the first paragraph, I believe it is IVHM cycle, not IVHM-IVHM cycle. In phase 4 paragraph, validation is mentioned twice in the list of activities and the last line says conversation of data instead of conversion of data.
By using this model in academics, researchers will learn about the new contexts in knowledge management and a good framework for managing research data. In industries, the model is robust and suitable for long-time industrial information management.
N/A
Overall, this model is impressive. It is correct, to the point and avoids un-necessary 'jargon'. The diagram could be re-designed slightly as the dotted lines that form the circumference of the inner circles can play 'mind-tricks', causing the reader to lose focus, so a solid line or some larger dashes may suit better.
The headings of the development life cycle phase is good, but direction is lost in top-level descriptions.
Using this approach in a smaller acquisition process to prove the concept would strengthen the case for its use

Appendix F Metadata and Content Types Development Strategy

Metadata and Content Types Development Strategy

Contents

1. Introduction

This document describes how the metadata and content types for the Knowledge Management System will be created.

2. Aim

- To create a plan for managing growing knowledge and information requirements
- To tackle issues of data quality, data integrity and data reuse
- To facilitate the consistent and accurate acquisition of strategic information from operational data
- To enable ease of information Sharing with researchers, staff and business partners

- To highlight the significance of organisational data management
- To enhance efficiency through element progress management and reuse
- To reduce duration in the software development cycle

3. Benefits

The advantages of a standard approach to metadata include:

- **Improved interoperability**: different datasets can be combined into a single view for analysis and demonstration.
- **Improved accessibility to services**: allow researchers, staff and partners to interact with multiple stakeholders or organisations through a single interface.
- **Enhanced decision making:** recognize trends, provision planning and targeted service delivery
- **Decreased administrative costs:** automate data management processes, modernise reporting, reduce duplication of effort through suitable sharing and re-use of knowledge and information assets.

4. Metadata Standards and Procedures

The terms store or taxonomy will be created taking into consideration the following standards used to enable sharing, access, interpretation and re-use of research data. They include:

a) ISO 15836 Dublin Core metadata element set AS 5044-2010 AGLS metadata element set. http://dublincore.org/ (**Content** – title, subject, Description, source, Language, relation, coverage; **Intellectual Property** – Creator, Publisher, Contributor, rights; **Instantiation** – Date, Type, Format, identifier)

b) Common European Research Information Format (CERIF),

c) Core Scientific metadata model

d) Data Documentation Initiative (DDI)

e) ISO 19115 Geographic Information – Metadata

5. Ways of Collecting the Metadata

a) Direct Metadata Data:

Generated automatically by the system.

b) Indirect Metadata (Managed Metadata)

This will be created from organisational structure, system artefacts, transactions, and uses of the system, analysis and decision making. This will include IVHM terms entities and attributes, technical and subject area names, query and report definitions etc metadata will be collected in a variety of ways including text files provided specifically to IVHM, from technical publications.

6. Metadata Elements To be Collected

The metadata elements to be created will focus on IVHM business context and will include terms, information governance rules, labels, and stewards. Technical metadata providing details about source and target systems, database table and field structures, and dependencies fundamental to system as well as operational metadata describing the job runs and the database table or data files.

TYPE	GOAL	HOW
Descriptive Metadata	Used for describing and identifying information resources. Descriptive metadata labels an information asset or declares certain characteristics at the local level that support searching, discovery and retrieval. Information to discover, identify, select and obtain the resource - Title, subject, date, description, type of resource, identifier, technical knowledge. Links to related resources help users understand the meaning and context of data or information.	**Human generated:** This will be generated directly from outputs of the department and from the requirements specification. **Autogenerated** when server is installed: author, description, site name, type, modified, last modified, URL, Title…
Structural Metadata	Facilitates navigation and presentation of electronic resources. It details the navigation structure of the system and relationship between resources: Information about the structured relationship between components. Chapter numbers, indexes, pages, sections, tables of contents	**Autogenerated** through crawling whole text indexing. Harvested from documents. **Human generated:** File plan - Document Classification

		Autogenerated: page, section, chapter numbering, indexes, and table of contents
Administrative Metadata	Administrative metadata supports the management of knowledge and information assets. Facilitates both short-term and long-term management and processing of digital collections. Version numbers, dates of significant actions, metadata about rights management, access control and use requirements, acquisition information, location, information, are examples of administrative metadata. These could support automated data disposal, proactive publishing or open access information. Metadata about a workflow, such as approval date, authorising party, revisions or amendments, also build trust and accountability into digital processes. ***Preservation Metadata***: Physical characteristics of the resource. Retention schedules. Used to Preservation Metadata.	**Human generated** **Autogenerated**
Technical Metadata	Functioning of system and behaviour of metadata. Software, hardware, security data.	**Autogenerated**

7. Meta Data Quality

Metadata about the level of accuracy, coherence and interpretability of a resource, how often it is updated, and other characteristics helps users to determine whether the resource can meet their requirements. Audit trails, creator or custodianship metadata describe integrity or authenticity. They help users evaluate data quality and whether it can be re-purposed. Data quality standards enable potential users to establish the relevance and fitness-for-purpose of data and information, for publication, exchange or re-use.

- Data Accuracy: how well data values represent the actual business requirements
- Data Completeness: how well available data meets current and future business information demands
- Data Currency: how timely are data values
- Data Consistency: are data definitions and values are the same across all data stores

- Data Integrity: conformation of source data values to data values) default values, referential integrity constraints, derived those allowed by business rules (data characteristics,

8. Resources

i. http://dublincore.org/
ii. http://www.ulb.ac.be/ceese/meta/meta.html
iii. http://metadata-standards.org/
iv. Data Documentation Initiative (DDI)
v. Common European Research Information Format (CERIF)

Appendix G IVHM Knowledge Management System Requirements

Integrated Vehicle Health Management Centre

IVHM Knowledge Management System Requirements

[Algorithmic Data and Knowledge Management Approach]

REVISION DATE: 21-02-2018

Revision Number	Scope	Date	Description
Version 1.0	Requirements	12/02/2018	Initial Creation

Approval of the System Requirements indicates an understanding of the purpose and content described in this deliverable. By signing this deliverable, each individual agrees with the content contained in this deliverable.

Contents

1 Aim of Document

As described below the purpose of the System Requirements document is to specify the overall specifications that will guide the development and implementation of the IVHM Knowledge Management System. The document will also establish any needed initial security, training, capacity and system architecture requirements, as well as, system acceptance criteria agreed upon be the project sponsor and key stakeholders.

2 Background

Knowledge management over the years has become a central substructure for business improvement, sustainability and competitiveness. This on the back drop of organizations seeking to manage their soft and hard assets. The IVHM Centre has existed since 2008 and runs an active research program that generates high value knowledge through theoretical and experimental work. It produces a wide range of algorithms and processes to capture and analyse data from experiments carried on rigs, facilities and process. The Centre working on more than 40 projects has generated more than 100 technical papers, published at least 6 books and generated terabytes of experimental data. The Centre has dozens of algorithms and 10 experimental rigs. With this growth in the creation of research data, data automating algorithms, technical papers, reports and theses the Centre needs both a data management model and a Knowledge Management System to storing, organizing and sharing its research outputs that is secure, scalable with high level of interoperability. We do not yet have a system and coherent approach to IVHM knowledge and data management. The absence of a Knowledge Management System means that valuable knowledge is lost or is very difficult to find. Data visualization is fragmented and done on project by project basis which increases costs [IVHM Centre].

3 Gathering of the Requirements

The research started by requesting for any previous requirements documents used or developed including from the existing SharePoint 2010 site. The documents including the “IVHM Centre – Core Capabilities and Vision” as well as the existing knowledge environment have been analysed to develop a comprehensive requirements for the Data and knowledge management system.

4 System Overview

The Integrated Vehicle Management (IVHM) Knowledge Management System (KMS) amongst other functions will be primarily combining the management of experimental Data and general content from multiple sources. SharePoint (SP) 2010 is deficient in the essential end-user and administrator requirements for the IVHM KMS and in addition, Microsoft support for SP2010 is to be scaled back by about 80% and above by 2020. The table below lists some of the critical requirements that are provided in SP2013 and not available in SP2010. SharePoint 2010 further lacks the flexibility, scalability, reliability and capabilities required for the IVHM KMS in the medium to long term. A key element; Content Management Interoperability Services (CMIS) remains a huge admin problem in SP2010; a feature that has been Simplified and improved significantly in later versions.

END USER REQUIREMENTS	ADMINISTRATOR REQUIREMENTS
1. Metadata-driven Navigation 2. Managed navigation 3. Hybrid search Capability 4. Quick preview 5. Search results sorting 6. Filter Search 7. Related Items 8. Follow Items 9. Content Recommendations 10. Web Content Management o Catalogue o Cross-site publishing o Faceted navigation o Image Renditions	1. Remote Event Receiver 2. Analytics Platform 3. Business connectivity Services ▪ Alerts for External Lists ▪ App-Scoped External Content Types (ECTs) ▪ Open Data Protocol (OData) connector ▪ Tenant-level external data log 4. Open authorization (OAuth) 5. Forms on Spreadsheets

o Search Engine Optimizations (SEO) 11. Office Web Apps Server integration 12. Preservation hold library 13. Decoupled Pivot Tables and Pivot Charts 14. Field List and Field Support 15. Filter Enhancements 16. Extensible content processing 17. Advanced Content Processing 18. Continuous Crawl 19. Custom entity extraction 20. Project workspace 21. One-click Share	

Table 1: End user and Administrator requirements for IVHM KMS

The sections below briefly describes the components of the IVHM Knowledge Management System:

Research (data acquisition): All about research activities. Ongoing projects, new projects, future projects, proposals and collaborative projects.

Document/ Records Management (preservation dissemination): Document content types, creation and routing workflows. Retention schedule applied on inactive documents.

Rigs (data acquisition): Rig configuration manuals, sensor details, H&S, operation, link to experiments, tests data, algorithms, KPIs, and publications.

Business Management: Financial models, fleet simulations, investment opportunities, SLAs/contracts and KPIs

Collaboration: Collaboration projects, proposals and partners

Knowledge Base (Library): Conference presentations, technical reports, technical papers, publications, user guides, linked to rigs, algorithms and data

Courses: Short courses and workshops

Support: Point of contact, health & safety, DSE, interactive guides' and troubleshooting

People: Profiles, expertise and research interests, board

Assets: Rigs, Computers and accessories …etc

External Resources: Relevant and useful resources from other sources

5 General System Requirements

5.1 Major KM System Capabilities

- The KM System should be available online
- The KM System should be accessible on mobile devices
- System can be easily upgraded.
- Records of system actions cannot be deleted.
- Integration with Outlook.
- The KM System must be web-based, conform to IVHM's Core Capabilities and Vision, support OSA-CBM standards, provision distributed computing, reside on SQL platform and link with MS Outlook/Exchange, and most file formats.
- The KM System should be able integrate with different functional applications like document management, correspondence tracking, reporting tools, documenting imaging, knowledge management, BI dashboard and e- forms.
- The KM System requires a tracking capability for alerts, logging, timestamping, reporting, close-out, and routing based on business rules, workflow, and prioritizing.
- The KM System should have the facility to install and activate/deactivate functionality.
- All system actions shall be date/time stamped and tracked.
- The KM system should be based in an approved cloud platform.
- The system should have lookup lists capability to improve routine activities.
- The KM System automated alerts system by email.
- The KM System can prioritize requests, projects, documents, and electronic communications.
- The KM System can deliver alerts and documents through email clients including standard handheld devices.
- Dashboard reporting and analytics.

5.2 Major System Conditions

The system should conform to the following attributes:

- **Scalable** — must be able to support a large number of users and a robust, industrial strength database;

- **Extensible** — capable of expanding as needed by the organization;
- **Compliant** with industry standards, allowing companies to leverage existing resources;
- **Secure**;
- **Collaborative** — although many efforts start with a single department of group, the best KM programs grow to encompass input from across the organization;
- Allow for complex queries;
- **Fast and easy** to administer and deploy;
- **Flexible** - The technology should be able to handle knowledge of any form, including different subjects, structures and media. It should be able to handle forms which do not as yet have been defined;
- Heuristic - The systems should learn about both its users and the knowledge it possesses as it is used. Over time, its ability to provide users with knowledge should improve. For example, if the solution deals with many requests on a particular subject, it should learn how to assist users in more depth on that subject;
- **Suggestive** - The solution should be able to deduce what a user's knowledge needs are and suggest knowledge associations that he is not able to do himself;

5.3 System Interfaces

The system will be based on office 365 and cloud hosted. A SharePoint site collection using a three level navigation structure.

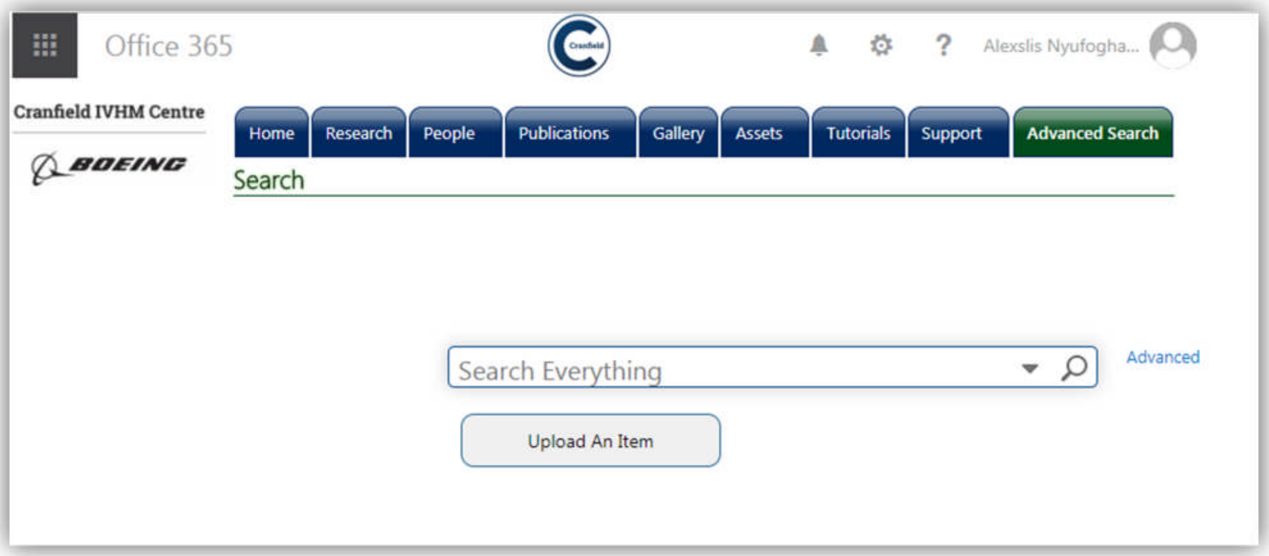


Figure 1: View of the pilot user-interface

5.4 System User Characteristics & requirements

Identify each type of user of the system by function, location, and type of device. Specify the number of users in each group and the nature of their use of the system.

- System Owner (Create, Design, Branding, Workflows, Taxonomy, Delete, Policies etc)
- Administrator(Contribute with Delete)
- Staff (Contribute without delete)
- Board members (Targeted, Contribute with or without Delete)
- Sponsors (Read only, Contribute without delete. Apply targeting)
- Partners (Contribute without delete)
- Researchers/ Research Fellows (Contribute without delete, Apply targeting)
- Public (read only)

✓ Role-based login capabilities that provides encryption and lost password recovery.
✓ Authorized users can create, edit, and delete email or in-system alerts that are automatically triggered by an event or date.
✓ Only authorized users can view system actions.
✓ The system shall have a new user’s tutorial.
✓ System shall provide standard navigation aids with the ability for users to skip through repetitive or unnecessary navigation.
✓ The system will have interactive on-line help that tells the user how to use the program and knowledge-management related information with access to constantly updated online user manuals.
✓ On-demand ability to fax requested documents to an authorized user’s workstation.
✓ The system will allow for collaboration with external users.

5.5 Search Requirements

✓ Searches can be performed by research project, publications, people, file type, data type, assets and rigs.
✓ Searches can be saved, modified, deleted, and shared.
✓ Ability to search within a search.
✓ Searches can be by full-text, keywords, and metatags.
✓ The ability to search different file types (.doc, .xls, .mdb, etc.), scanned files (PDFs), and text within graphics.
✓ Searches can be by index, Boolean (simple and multiple), wild card, near spell, proximity, synonyms, exact phrase, and exclusion of terms.

- ✓ Search terms will be highlighted in the document or web page.
- ✓ Search results can provide a list of hits along document object histories, progress through workflow routes, summaries of documents, or profile data of documents.
- ✓ List of most often requested results for a specified search term.
- ✓ Searches can be exported into common document formats (.doc, .xls, .pdf, etc.).

5.6 Data Management Requirements

- ✓ Consistent data entry formats (including pop-up calendar for dates), data validation, ability to cut and paste into data fields.
- ✓ System shall allow definition of custom fields including calculated fields with customizable error messages.
- ✓ Validating messages for deletion.
- ✓ Fields can be locked for read-only views.
- ✓ All electronic data shall be kept under conditions to prevent degradation, periodically spot-checked for data integrity, comply with IVHM's Record Management policies where available, and have a back-up [provided by IT].
- ✓ Ability to exempt specific documents from regular scheduled destruction date.
- ✓ All electronic documents shall be immediately available and metatags shall be updated at least yearly by document owners.
- ✓ Data security shall comply with all Data protection regulations and research ethics guidance.
- ✓ Data access will be by group or user-based roles.
- ✓ Security access can be customized by metadata, document, database column, and/or database row to allow read, edit, or delete functions.
- ✓ Data can be accessed securely from off-site and will apply to any information exchanged between systems and based on workflow.
- ✓ The KMS shall provide an interface for data sharing with our sponsors and partners.

5.7 Document Management Requirements

- ✓ Document management is available for documents in different file formats and as well as scanned documents.
- ✓ The system shall have a library of standard documents and templates, version control, document storage, and ability to create customized documents.
- ✓ The system shall use best practice naming conventions, can link documents, support meta tagging/profiling of documents, annotation of documents, package documents of different file-types together, and previewing of files in HTML format.

- ✓ System shall allow protection of documents as “permanent” and allow for auto-population of document/profile fields from correspondence system.
- ✓ Documents will have their own audit trail system with alerts and notifications when actions are taken with the documents.
- ✓ System shall be able to capture documents in various sizes and thicknesses and convert these captured documents to PDFs.
- ✓ System shall be able to modify and redact sections of scanned documents.
- ✓ The ability to freeze destruction automation for emails and documents relevant to an impending e-Discovery request.
- ✓ The ability to track retention periods for documents
- ✓ Ability to route documents through workflow for approvals/electronic signatures
- ✓ System must provision application of retention schedules, automate reporting and purging actions and facility for exporting records and metadata.

  Metadata shall be managed metadata and automated metadata. The term store metadata shall be created using IVHM Core capabilities and vison document, as well as content generated since 2008. Content types and lookup lists shall be used to ease consistency and ease of system management.

5.8 Knowledge Management Requirements

- ✓ System shall store explicit knowledge and allow for self-service searches of a knowledge repository.
- ✓ System shall have a searchable audio and video library devoted to capturing tacit knowledge.
- ✓ System will record and display who posted what content and on what date/time.
- ✓ Ability to track versions of content submissions.
- ✓ Ability to roll-back to previous content submissions.

6 Policy and Regulation Requirements

The system shall be developed taking the following policies and regulatory into consideration where applicable. They include but not limited to:

- ✓ Research data Management
- ✓ Research ethics
- ✓ Data Protection Act
- ✓ Strategic Export/Import Control
- ✓ University policy, Design & branding guidelines

7 Security Requirements

The System owner (Super Administrator) must be able to limit access to records, files and metadata to specific users and groups. Furthermore they should be able to apply targeted access. Restrict user access to content after it has expired and limit to strictly on demand. The system must be able to capture audit trail all activities performed on content including date and time.

8 Training Requirements

Self-help training materials for shall be developed alongside FAQs as required.

9 Initial System Architecture

Microsoft SharePoint Architecture

10 System Acceptance Criteria

The new system must meet stakeholder data management and Knowledge Management requirements.

11 Roles and responsibilities

Director of IVHM Center

IT Service

Data security and knowledge Management officer

Researchers and Research Fellows

12 Resources

Document No.	Document Title	Author
01	Knowledge Management System-MSc Project	IVHM Centre
02	IVHM Centre Core Capabilities and Vision	Prof Ian K Jennions
03	https://www.cranfield.ac.uk/research/phd/knowledge-management-system-for-the-ivhm-centre	IVHM Centre
04	IEEE Recommended Practice for Software Requirements Specifications, IEEE Std. 830-1998.	IEEE
05	IEEE Guide for Developing System Requirements Specifications, IEEE Std. 1233-1998.	IEEE

13 Document Revision History

Version	Date	Description
01	12/02/2018	Initial Creation

Revisions [initial creation, change request, new mandated change]

Appendix H IVHM Data Management Requirements 2.0

Integrated Vehicle Health Management Centre

IVHM Knowledge Management System Requirements 2.0

[Algorithmic Data and Knowledge Management Approach]

REVISION DATE: 21-02-2018

Revision Number	Scope	Date	Description
Version 1.0	Requirements	12/02/2018	Initial Creation

Approval of the System Requirements indicates an understanding of the purpose and content described in this deliverable. By signing this deliverable, each individual agrees with the content contained in this deliverable.

Contents

1 Aim of Document

The aim of this document is to specify the overall data management requirements that will govern the development and implementation of the of the IVHM data management within the Knowledge Management System.

2 Background

Data management has been described as the organizational process of acquiring, validating, storing, protecting, and processing in order to guarantee the accessibility, reliability, and timeliness for user needs and satisfaction(Gonzalez, 2014; Mbabu *et al.*, 2014) . Knowledge Management over the years has become a central substructure for business improvement, sustainability and competitiveness. This on the back drop of organizations seeking to manage their soft and hard assets. However, much of the emphasis has been on Knowledge management, with not much focus on the foundation layer which is Data. The IVHM Centre has existed since 2008 and runs an active research program that generates high value knowledge through theoretical and experimental work. It produces a wide range of algorithms and processes to capture and analyse data from experiments carried on rigs, facilities and process. The Centre working on more than 40 projects has generated more than 100 technical papers, published at least 6 books and generated terabytes of experimental data. The Centre has dozens of algorithms and 10 experimental rigs. With this growth, the Centre is beginning to face the challenges of data storage, sharing, preservation, reproducibility, quality and security which further creates the problems of replication, provenance (Yang, Matthews and Wilson, 2013), data underutilization and data loss (Hsu *et al.*, 2015). These eminent challenges therefore necessitate both a data management model and a knowledge management system for storing, organizing and sharing its research outputs that is secure, scalable with high level of interoperability to cater for the growing volumes of research data, data automating algorithms, technical papers and reports. We do not yet have a system and coherent approach to IVHM knowledge and data management. Data visualization, data sharing are fragmented, incoherent and done on project by project basis which increases costs [IVHM Centre].

3 Gathering of the Requirements

The research started by requesting for any previous requirements documents used or developed including from the existing SharePoint 2010 site. The documents including the "IVHM Centre – Core Capabilities and Vision" as well as the existing knowledge environment have been analysed to develop a comprehensive requirements for the Data and knowledge management system.

3.1 Data Management Requirements

3.1.1 High level requirements

1. Create a data life cycle model to eliminate waste, improve efficiency, and ensure use of data management best practices.
2. Create specification for data to be passed between collection, experimental rigs and knowledge managements system.

3. Integrate these requirements with Knowledge Management system.
4. Design a library (database) for storing details of experiments, algorithms and assets.
5. Create a GUI interface that allows researchers, research fellows and Staff to view experiments data, share experiments data as well as derived data.

3.1.2 System integrated specification

- ✓ Consistent data entry formats (including pop-up calendar for dates), data validation, ability to cut and paste into data fields.
- ✓ System shall allow definition of custom fields including calculated fields with customizable error messages.
- ✓ Validating messages for deletion.
- ✓ Fields can be locked for read-only views.
- ✓ All electronic data shall be kept under conditions to prevent degradation, periodically spot-checked for data integrity, comply with IVHM’s Record Management policies where available, and have a back-up [provided by IT].
- ✓ Ability to exempt specific documents from regular scheduled destruction date.
- ✓ All electronic documents shall be immediately available and metatags shall be updated at least yearly by document owners.
- ✓ Data security shall comply with all data protection regulations and research ethics guidance.
- ✓ Data access will be by group or user-based roles.
- ✓ Security access can be customized by metadata, document, database column, and/or database row to allow read, edit, or delete functions.
- ✓ Data can be accessed securely from off-site and will apply to any information exchanged between systems and based on workflow.
- ✓ The KMS shall provide an interface for data sharing with our sponsors and partners.

3.2 Search Requirements

- ✓ Searches can be performed by research project, publications, people, file type, data type, assets and rigs.
- ✓ Searches can be saved, modified, deleted, and shared.
- ✓ Ability to search within a search.
- ✓ Searches can be by full-text, keywords, and metatags.
- ✓ The ability to search different file types (.doc, .xls, .mdb, etc.), scanned files (PDFs), and text within graphics.
- ✓ Searches can be by index, Boolean (simple and multiple), wild card, near spell, proximity, synonyms, exact phrase, and exclusion of terms.

- ✓ Search terms will be highlighted in the document or web page.
- ✓ Search results can provide a list of hits along document object histories, progress through workflow routes, summaries of documents, or profile data of documents.
- ✓ List of most often requested results for a specified search term.
- ✓ Searches can be exported into common document formats (.doc, .xls, .pdf, etc.).
- ✓ Actions and facility for exporting records and metadata.

4 Security Requirements

The System owner (Super Administrator) must be able to limit access to records, files and metadata to specific users and groups. Furthermore they should be able to apply targeted access. Restrict user access to content after it has expired and limit to strictly on demand. The system must be able to capture audit trail all activities performed on content including date and time.

Document No.	Document Title	Author
01	Knowledge Management System-MSc Project	IVHM Centre
02	IVHM Centre Core Capabilities and Vision	Prof Ian K Jennions
03	https://www.cranfield.ac.uk/research/phd/knowledge-management-system-for-the-ivhm-centre	IVHM Centre
04	IEEE Recommended Practice for Software Requirements Specifications, IEEE Std. 830-1998.	IEEE
05	IEEE Guide for Developing System Requirements Specifications, IEEE Std. 1233-1998.	IEEE

5 Document Revision History

Version	Date	Description
01	12/02/2018	Initial Creation

Revisions [initial creation, change request, new mandated change]

Integrated Vehicle Health Management Centre

IVHM Knowledge Management System Requirements 3.0

[Data- and Knowledge Management Approach]

REVISION DATE: 21-02-2018

Revision Number	Scope	Date	Description
Version 1.0	Requirements	12/02/2018	Initial Creation

Approval of the System Requirements indicates an understanding of the purpose and content described in this deliverable. By signing this deliverable, each individual agrees with the content contained in this deliverable.

Contents

1 Aim of Document

The aim of the System Requirements document is to specify the overall system requirements that will govern the development and implementation of the IVHM Knowledge Management System. The document will also establish any needed initial security, training, capacity and system architecture requirements, as well as, system acceptance criteria agreed upon be the project sponsor and key stakeholders.

2 Background

Knowledge management over the years has become a central substructure for business improvement, sustainability and competitiveness. This on the back drop of organizations seeking to manage their soft and hard assets. The IVHM Centre has existed since 2008 and runs an active research program that generates high value knowledge through theoretical and experimental work. It produces a wide range of algorithms and processes to capture and analyse data from experiments carried on rigs, facilities and process. The Centre working on more than 40 projects has generated more than 100 technical papers, published at least 6 books and generated terabytes of experimental data. The Centre has dozens of algorithms and 10 experimental rigs. With this growth in the creation of research data, data automating algorithms, technical papers, reports and theses the Centre needs both a data management model and a knowledge management system to storing, organizing and sharing its research outputs that is secure, scalable with high level of interoperability. We do not yet have a system and coherent approach to IVHM knowledge and data management. The absence of a knowledge management system means that valuable knowledge is lost or is very difficult to find. Data visualization is fragmented and done on project by project basis which increases costs [IVHM Centre].

3 Gathering of the Requirements

The research started by requesting for any previous requirements documents used or developed including from the existing SharePoint 2010 site. The documents including the “IVHM Centre – Core Capabilities and Vision” as well as the existing knowledge environment have been analysed to develop a comprehensive requirements for the Data and knowledge management system.

4 System Overview

The Integrated Vehicle Management (IVHM) knowledge management system (KMS) amongst other functions will be primarily combining the management of experimental Data and general content from multiple sources. SharePoint (SP) 2010 is deficient in the essential end-user and administrator requirements for the IVHM KMS and in addition, Microsoft support for SP2010 is to be scaled back by about 80% and above by 2020. The table below lists some of the critical requirements that are provided in SP2013 and not available in SP2010. SharePoint 2010 further lacks the flexibility, scalability, reliability and capabilities required for the IVHM KMS in the medium to long term. A key element; Content Management Interoperability Services (CMIS) remains a huge admin problem in SP2010; a feature that has been Simplified and improved significantly in later versions.

END USER REQUIREMENTS	ADMINISTRATOR REQUIREMENTS
22. Metadata-driven Navigation 23. Managed navigation 24. Hybrid search Capability 25. Quick preview 26. Search results sorting 27. Filter Search 28. Related Items 29. Follow Items 30. Content Recommendations 31. Web Content Management o Catalogue o Cross-site publishing o Faceted navigation o Image Renditions o Search Engine Optimizations (SEO) 32. Office Web Apps Server integration 33. Preservation hold library 34. Decoupled Pivot Tables and Pivot Charts 35. Field List and Field Support 36. Filter Enhancements 37. Extensible content processing 38. Advanced Content Processing 39. Continuous Crawl 40. Custom entity extraction	6. Remote Event Receiver 7. Analytics Platform 8. Business connectivity Services ▪ Alerts for External Lists ▪ App-Scoped External Content Types (ECTs) ▪ Open Data Protocol (OData) connector ▪ Tenant-level external data log 9. Open authorization (OAuth) 10. Forms on Spreadsheets

41. Project workspace 42. One-click Share	

Table 1: End user and Administrator requirements for IVHM KMS

The sections below briefly describes the components of the IVHM knowledge management system:

Research (data acquisition): All about research activities. Ongoing projects, new projects, future projects, proposals and collaborative projects.

Document/ Records Management (preservation dissemination): Document content types, creation and routing workflows. Retention schedule applied on inactive documents.

Rigs (data acquisition): Rig configuration manuals, sensor details, H&S, operation, link to experiments, tests data, algorithms, KPIs, and publications.

Business Management: Financial models, fleet simulations, investment opportunities, SLAs/contracts and KPIs

Collaboration: Collaboration projects, proposals and partners

Knowledge Base (Library): Conference presentations, technical reports, technical papers, publications, user guides, linked to rigs, algorithms and data

Courses: Short courses and workshops

Support: Point of contact, health & safety, DSE, interactive guides' and troubleshooting

People: Profiles, expertise and research interests, board

Assets: Rigs, Computers and accessories …etc

External Resources: Relevant and useful resources from other sources

5 General System Requirements

1.1. Major System Capabilities

- System must be available on the Internet
- System must be available 24 hours per day

- System must be accessible by mobile devices
- System can be easily upgraded.
- Records of system actions cannot be deleted.
- Integration with Outlook.
- System must be web-based, conform to IVHM's Core Capabilities and Vision, support OSA-CBM standards, support distributed processing, reside on an Intel platform, operate in a MS Windows environment, support VPN, server-based fax pool, data encryption/PKI, Assistive Technology, interface with MS Outlook/Exchange, and common file formats.
- System must integrate with different functional applications such as document management, correspondence tracking, reporting tools, documenting imaging, knowledge management, IVR systems, executive dashboard and electronic forms.
- System shall have a correspondence tracking system including logging, tracking, timestamping, notification, reporting, close-out, and routing based on business rules, workflow, and prioritizing.
- System is modular with ability to deploy and activate/deactivate functionality.
- All system actions shall be date/time stamped and tracked.
- All or most of the KMS will be based in an approved cloud environment.
- System actions are to be stored in a secured database, searchable, and can trigger alerts.
- Whenever appropriate, system will use look-up lists to streamline routine operations.
- Automated reminder system by email.
- System can prioritize requests, projects, documents, and electronic communications.
- System can deliver alerts and documents through email clients including standard handheld devices.
- Dashboard reporting and analytics.

1.2. Major System Conditions

The system should conform to the following attributes:

- **Scalable** — must be able to support a large number of users and a robust, industrial strength database;
- **Extensible** — capable of expanding as needed by the organization;

- **Compliant** with industry standards, allowing companies to leverage existing resources;
- **Secure**;
- **Collaborative** — although many efforts start with a single department of group, the best KM programs grow to encompass input from across the organization;
- Allow for complex queries;
- **Fast and easy** to administer and deploy;
- **Flexible** - The technology should be able to handle knowledge of any form, including different subjects, structures and media. It should be able to handle forms which do not as yet have been defined;
- Heuristic - The systems should learn about both its users and the knowledge it possesses as it is used. Over time, its ability to provide users with knowledge should improve. For example, if the solution deals with many requests on a particular subject, it should learn how to assist users in more depth on that subject;
- **Suggestive** - The solution should be able to deduce what a user's knowledge needs are and suggest knowledge associations that he is not able to do himself;

1.3. System Interfaces

The system will be based on office 365 and cloud hosted. A SharePoint site collection using a three level navigation structure.

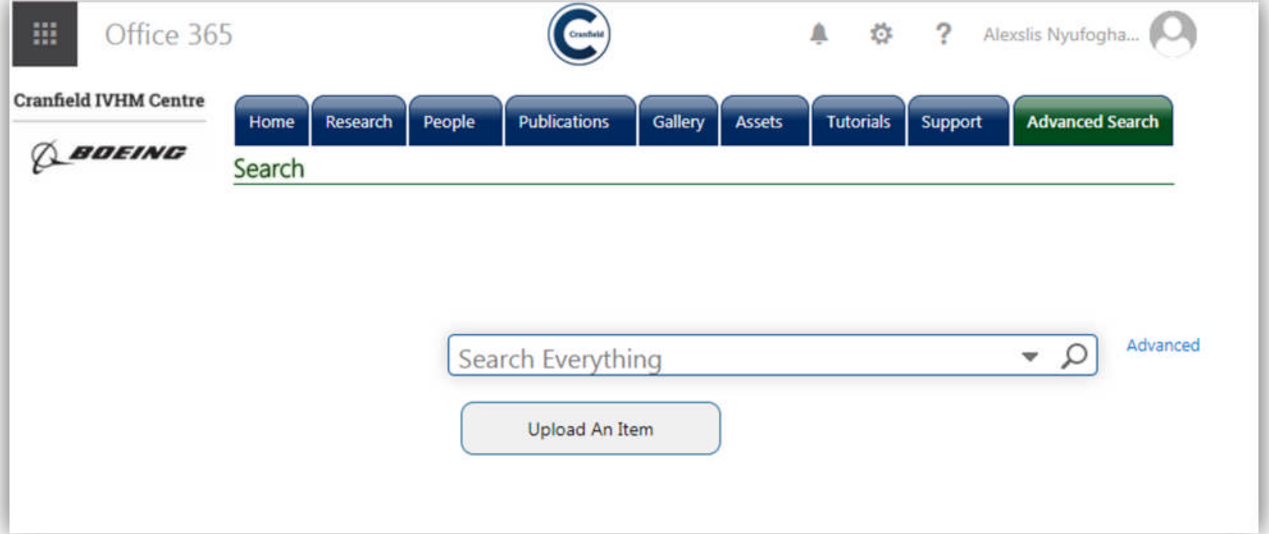


Figure 1: View of the pilot user-interface

1.4. System User Characteristics & requirements

Identify each type of user of the system by function, location, and type of device. Specify the number of users in each group and the nature of their use of the system.

- System Owner (Create, Design, Branding, Workflows, Taxonomy, Delete, Policies etc)
- Administrator(Contribute with Delete)
- Staff (Contribute without delete)
- Board members (Targeted, Contribute with or without Delete)
- Sponsors (Read only, Contribute without delete. Apply targeting)
- Partners (Contribute without delete)
- Researchers/ Research Fellows (Contribute without delete, Apply targeting)
- Public (read only)

- ✓ Role-based login capabilities that provides encryption and lost password recovery.
- ✓ Authorized users can create, edit, and delete email or in-system alerts that are automatically triggered by an event or date.
- ✓ Only authorized users can view system actions.
- ✓ The system shall have a new user’s tutorial.
- ✓ System shall provide standard navigation aids with the ability for users to skip through repetitive or unnecessary navigation.
- ✓ The system will have interactive on-line help that tells the user how to use the program and knowledge-management related information with access to constantly updated online user manuals.
- ✓ On-demand ability to fax requested documents to an authorized user’s workstation.
- ✓ The system will allow for collaboration with external users.

1.5. Search Requirements

- ✓ Searches can be performed by research project, publications, people, file type, data type, assets and rigs.
- ✓ Searches can be saved, modified, deleted, and shared.
- ✓ Ability to search within a search.
- ✓ Searches can be by full-text, keywords, and metatags.
- ✓ The ability to search different file types (.doc, .xls, .mdb, etc.), scanned files (PDFs), and text within graphics.
- ✓ Searches can be by index, Boolean (simple and multiple), wild card, near spell, proximity, synonyms, exact phrase, and exclusion of terms.
- ✓ Search terms will be highlighted in the document or web page.

- ✓ Search results can provide a list of hits along document object histories, progress through workflow routes, summaries of documents, or profile data of documents.
- ✓ List of most often requested results for a specified search term.
- ✓ Searches can be exported into common document formats (.doc, .xls, .pdf, etc.).

1.6. Data Management Requirements

- ✓ Consistent data entry formats (including pop-up calendar for dates), data validation, ability to cut and paste into data fields.
- ✓ System shall allow definition of custom fields including calculated fields with customizable error messages.
- ✓ Validating messages for deletion.
- ✓ Fields can be locked for read-only views.
- ✓ All electronic data shall be kept under conditions to prevent degradation, periodically spot-checked for data integrity, comply with IVHM's Record Management policies where available, and have a back-up [provided by IT].
- ✓ Ability to exempt specific documents from regular scheduled destruction date.
- ✓ All electronic documents shall be immediately available and metatags shall be updated at least yearly by document owners.
- ✓ Data security shall comply with all Data protection regulations and research ethics guidance.
- ✓ Data access will be by group or user-based roles.
- ✓ Security access can be customized by metadata, document, database column, and/or database row to allow read, edit, or delete functions.
- ✓ Data can be accessed securely from off-site and will apply to any information exchanged between systems and based on workflow.
- ✓ The KMS will provide an interface for data sharing with our sponsors and partners.

1.7. Document Management Requirements

- ✓ Document management is available for documents in different file formats and as well as scanned documents.
- ✓ The system shall have a library of standard documents and templates, version control, document storage, and ability to create customized documents.
- ✓ The system shall use best practice naming conventions, can link documents, support meta tagging/profiling of documents, annotation of documents, package documents of different file-types together, and previewing of files in HTML format.

- ✓ System shall allow protection of documents as “permanent” and allow for auto-population of document/profile fields from correspondence system.
- ✓ Documents will have their own audit trail system with alerts and notifications when actions are taken with the documents.
- ✓ System shall be able to capture documents in various sizes and thicknesses and convert these captured documents to PDFs.
- ✓ System shall be able to modify and redact sections of scanned documents.
- ✓ The ability to freeze destruction automation for emails and documents relevant to an impending e-Discovery request.
- ✓ The ability to track retention periods for documents
- ✓ Ability to route documents through workflow for approvals/electronic signatures
- ✓ System must provision application of retention schedules, automate reporting and purging actions and facility for exporting records and metadata.

1.8. Metadata management requirements

Metadata shall be managed metadata and automated metadata. The term store metadata shall be created using IVHM Core capabilities and vison document, as well as content generated since 2008. Content types and lookup lists shall be used to ease consistency and ease of system management.

1.9. Knowledge Management Requirements

- ✓ System shall store explicit knowledge and allow for self-service searches of a knowledge repository.
- ✓ System shall have a searchable audio and video library devoted to capturing tacit knowledge.
- ✓ System will record and display who posted what content and on what date/time.
- ✓ Ability to track versions of content submissions.
- ✓ Ability to roll-back to previous content submissions.
- ✓

6 Policy and Regulation Requirements

The system shall be developed taking the following policies and regulatory into consideration where applicable. They include but not limited to:

- ✓ Research data Management
- ✓ Research ethics
- ✓ Data Protection Act
- ✓ Strategic Export/Import Control
- ✓ University policy, Design & branding guidelines

7 Security Requirements

The System owner (Super Administrator) must be able to limit access to records, files and metadata to specific users and groups. Furthermore they should be able to apply targeted access. Restrict user access to content after it has expired and limit to strictly on demand. The system must be able to capture audit trail all activities performed on content including date and time.

8 Training Requirements

Self-help training materials for shall be developed alongside FAQs as required.

9 Initial System Architecture

Microsoft SharePoint Architecture

10 System Acceptance Criteria

The new system must meet stakeholder data management and Knowledge Management requirements.

11 Roles and responsibilities

Director of IVHM Centre

IT Service

Data security and knowledge Management officer

Researchers and Research Fellows

12 References

Document No.	Document Title	Author
01	Knowledge Management System-MSc Project	IVHM Centre
02	IVHM Centre Core Capabilities and Vision	Prof Ian K Jennions
03	https://www.cranfield.ac.uk/research/phd/knowledge-management-system-for-the-ivhm-centre	IVHM Centre
04	IEEE Recommended Practice for Software Requirements Specifications, IEEE Std. 830-1998.	IEEE
05	IEEE Guide for Developing System Requirements Specifications, IEEE Std. 1233-1998.	IEEE

13 Document Revision History

Version	Date	Description
01	12/02/2018	Initial Creation

Revisions [initial creation, change request, new mandated change]