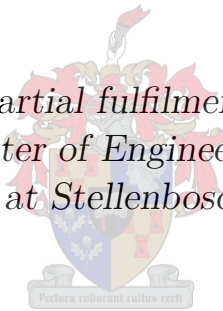


Developing a Framework for Identifying and Assessing Data Quality Issues in Asset Management Decision-Making

by

Jan Rossouw Minnaar

Thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering in the Faculty of Engineering at Stellenbosch University



Department of Industrial Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Prof. P.J. Vlok

December 2015

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature:
J.R. Minnaar

Date: 2014/11/01

Copyright © 2015 Stellenbosch University
All rights reserved.

Abstract

Developing a Framework for Identifying and Assessing Data Quality Issues in Asset Management Decision-Making

J.R. Minnaar

*Department of Industrial Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.*

Thesis: MEng (Industrial)

March 2015

Assets allow organizations to achieve their strategic objectives. Asset management translate these objectives into asset related decision and actions. A key enabler of asset management decision-making is data. Data quality, however, is a common challenge faced by many organizations, especially in the asset-intensive industry. This problem is compounded by the multitude of business intelligence systems and data standards competing for market share in some subset of an organization's data pipeline. With rapidly increasing data volumes and global competition demanding optimal management of assets, data quality is a problem that all asset managers will have to face.

With the release of the ISO 55000 series of international standards for asset management in February 2014, many asset managers will seek to implement a compliant asset management system. The ISO 55000, however, intends to be applicable to the broadest range of assets, organizations and cultures and therefore does not provide specific technical requirements. Previous studies have attempted to provide such technical requirements for data quality in asset management and subsequently contributed one more standard or system to an already crowded market.

Asset managers' continued struggle with data quality is thus not due to a lack of standards or systems. In fact, the many competing and often incompatible systems and standards are one of the many reasons for poor data quality. The severity of data quality issues and their impact on asset management decision-making were observed in a Southern African diamond mine.

A preliminary literature review confirmed that these observations were not isolated. Data quality, however, was also found to be a complex and context-specific problem, especially in asset management. Thus, instead of developing yet another standard or system to either replace or provide compatibility between existing systems, this study adopts a pragmatic approach to develop a framework to help asset managers identify *their* most critical data quality issues.

To answer the question of what such a framework would look like, a pragmatic research approach was adopted. The framework and its components were developed through an iterative cycle of development and evaluation. Applicable knowledge from a comprehensive literature review assured innovation and the business needs from the diamond mine case study ensured that the solution is relevant. The study found that the framework requires three components to be of value. The three components are: (1) a data pipeline reference model, (2) a methodology to guide asset managers in collecting the relevant data and (3) a tool to help asset managers populate their data pipeline model and identify data quality issues.

The usefulness (which is the measure of value in the pragmatic world view) of the framework was demonstrated by applying the framework in practice and fixing the critical data quality issues that it identified.

The modular nature of the framework allows future studies to be carried out to integrate the framework with various other disciplines to not only identify data quality issues, but also systematically address them. The hope is that this framework will eventually become part of a larger, pragmatic approach to allow asset managers to implement an ISO 55001 compliant asset management system.

Uittreksel

Ontwikkeling van 'n Raamwerk vir die Identifisering en Assessering van Datakwaliteitsprobleme in Batebestuurbesluitneming

*(“Developing a Framework for Identifying and Assessing Data Quality Issues in
Asset Management Decision-Making”)*

J.R. Minnaar

*Departement Bedryfsingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.*

Tesis: MIng (Bedryfs)

Maart 2015

Bates laat organisasies toe om hul strategiese doelwitte te bereik. Batebestuur omskep hierdie doelwitte na bate-verwante besluite en aksies. Data speel 'n sleutelrol in batebestuur-besluitneming. Die kwaliteit van data is egter 'n al groter-wordende uitdaging vir baie organisasies, veral in die bate-intensiewe bedryf. Hierdie probleem word vererger deur die menigte inligtingstelsels en standaarde wat meeding vir 'n aandeel in 'n organisasie se data-pyplyn. Met die vinnig toenemende data volumes en globale mededingendheid wat optimale bestuur van bates vereis, is data-kwaliteit 'n probleem wat alle batebestuurders sal moet trotseer.

Die publikasie van die ISO 55000-reeks van internasionale standaarde vir batebestuur in Februarie 2014, beteken dat al meer batebestuurders 'n batebesuutstelsel wat hieraan voldoen sal will implementeer. Die ISO 55000 standaard strewe daarna om van toepassing op die grootste verskeidenheid bates, organisasies en kulture te wees. ISO 55000 voorsien dus nie spesifieke tegniese vereistes nie. Vorige studies het al gepoog om hierdie tegniese vereistes vir die kwaliteit van data in batebestuur te ontwikkel. Hulle het egter slegs daarin geslaag om nog 'n stelsel of standaard by te dra tot die reeds oorvol mark.

Batebestuurders se voortgesette stryd met die kwaliteit van data is nie as gevolg van 'n gebrek aan stelsels of standaarde nie. Trouens, die baie mededingende en dikwels onversoembare stelsels en standaarde is een van die baie redes

vir die lae kwaliteit van data. Die erns van data kwaliteitkwessies en hul impak op batebestuur-besluitneming is in 'n Suider-Afrikaanse diamantmyn waargeneem. 'n Aanvakklike literatuuroorsig het bevestig dat hierdie waarnemings nie 'n geïsoleerde geval was nie. Boonop is bevind dat kwaliteit van data 'n komplekse en konteks spesifieke probleem is, veral in batebestuur. So, in stede van die ontwikkeling van nog 'n standaard of stelsel om óf huidige stelsels te vervang óf te versoen, neem hierdie studie 'n pragmatiese benadering. Hierdie studie poog dus om 'n raamwerk te ontwikkel wat batebestuurders help om hul mees kritieke datakwaliteitsprobleme te identifiseer.

'n Pragmatiese navorsingsmetodologie word ingespan om die vraag oor hoe so 'n raamwerk sal lyk te beantwoord. Die raamwerk en sy komponente is ontwikkel deur 'n iteratiewe siklus van ontwikkeling en evaluering. Toepaslike kennis uit 'n omvattende literatuurstudie het die nodige fokus verseker terwyl die sake-behoeftes van die diamantmyn relevansie verskaf het. Die studie het bevind dat die raamwerk drie komponente vereis om van waarde te wees. Die drie komponente is: (1) 'n data-pyplyn verwysingsmodel, (2) 'n metode wat batebestuurders lei in die versameling van die data en (3) 'n program om batebestuurders te help om hul data-pyplyn model te voltooi en sodoende datakwaliteitsprobleme te identifiseer.

Die nut (wat 'n maatstaf van waarde in die pragmatiese wêreldbeskouing is) van die raamwerk is gedemonstreer deur die toepassing van die raamwerk in praktyk. Kritieke datakwaliteitsprobleme is beide geïdentifiseer en aangespreek. Die studie het ook bevind dat die pragmatiese benadering nie net gehelp het met 'n unieke verstaan van die probleem nie, maar ook die studie in staat gestel het om bydraes te maak tot die navorsingskennispoel.

Die modulêre aard van die raamwerk moedige toekomstige studies aan om die raamwerk met ander dissiplines te integreer om nie net die datakwaliteitskwessies te identifiseer nie, maar ook stelselmatig aan te spreek. Die hoop is dat hierdie raamwerk uiteindelik deel van 'n groter, pragmatiese benadering sal word wat die implementering van 'n ISO 55001 voldoende batebestuurstelsel vergemaklik.

Acknowledgements

“I can do all things, through Him who gives me strength”
– Philippians 4:13

I would like to express my sincere gratitude to the following people and organizations that made this thesis possible:

- My supervisor, Professor PJ Vlok, for his support and motivation.
- Anglo American for partially funding this study.
- The Industrial Engineering department of Stellenbosch University for providing a friendly and productive work environment.
- Mr Grahame Fogel, for providing me with an opportunity to gain practical experience and ground this thesis in practice.
- The management and staff at Jwaneng Diamond Mine for helping me understand the problem and validating my solution.
- My parents, for their endless love and support.
- Elsabé, for showing interest in my thesis; being willing to listen to my ideas; and most off all, for her motivation and enduring support.

Contents

Declaration	i
Abstract	ii
Uittreksel	iv
Acknowledgements	vi
Contents	vii
List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Background	1
1.2 Rationale for the study	7
1.3 Research objective, questions and strategy	10
1.4 Delineation	11
1.5 Structure of the document	12
2 Literature Review	13
2.1 Asset Management	13
2.1.1 Overview	14
2.1.2 The Rise of Asset Management	15
2.1.3 Asset Management Today: ISO 55000	17
2.2 Decision-Making	21
2.2.1 Overview	21
2.2.2 Information Flow Models	23
2.2.3 The Role of Data	26
2.3 Data Quality	27
2.3.1 Data	27
2.3.2 Data Life Cycle	29
2.3.3 Data Quality	35
2.4 Information Systems	38

<i>CONTENTS</i>	viii
2.4.1 Information Systems in Asset Management	39
2.4.2 Development Considerations	40
2.4.3 Technology	44
2.5 Frameworks and Reference Models	45
2.6 Summary	46
3 Methodology	48
3.1 Philosophical World View	49
3.2 Research Methods	50
3.3 Research Design	52
3.4 Ethical and Legal Considerations	53
4 Proposed Solution	55
4.1 Introduction	55
4.2 Solution Objectives	59
4.3 The Framework	60
4.3.1 The Model	60
4.3.2 The Tool	66
4.3.3 The Process	70
4.4 Summery	71
5 Case Study	73
5.1 Understanding the Organization	74
5.2 The Data Model	82
5.3 Data Quality Issues	88
5.4 Evaluation	89
6 Closure	91
6.1 Reflection	91
6.2 Contribution	93
6.3 Conclusion	93
Appendices	95
A Reference Model	96
B Examples of Serialized Data	98
References	100

List of Figures

1.1	The problem of competing standards	5
1.2	Example of website using non-standard functionality	6
2.1	Comparison of Google search volumes for PAS 55 and ISO 55000	14
2.2	Asset Management as defined by ISO 55000	17
2.3	The IAM Conceptual Model of Asset Management	18
2.4	The Data Learning Model	23
2.5	The Imprinted Model	24
2.6	The Deeper Learning Model	25
2.7	The DIKW hierarchy	28
2.8	Overview of Data Stores	31
2.9	The System Development Life Cycle Methodology	41
2.10	The two aspects of divisibility: incrementalism and modularity	43
3.1	Design Science Research Method	52
4.1	The identification and exploration of the problem.	56
4.2	Asset management as a teleological “black box” model	56
4.3	Asset management as an ontological “clear box” model	57
4.4	Teleological “black-box” data pipeline model	58
4.5	The problem of too many systems and standards	61
4.6	Data pipeline model overview	62
4.7	Decision tree showing some of the technologies that were considered	68
4.8	User friendly interface for building hierarchies	70
5.1	Debswana ownership and subsidiaries	74
5.2	Location of Debswana’s mining operations	75
5.3	Aerial view of Jwaneng’s Main Treatment Plant	77
5.4	Jwaneng MTP process flows	78
5.5	Example view of graphical interface for SCADA	79
5.6	The as-found information flows	81
5.7	Distribution of indicator types	83
5.8	Distribution of indicator categories	84
5.9	How delays are recorded	87
5.10	The reporting tool	89

LIST OF FIGURES

x

5.11 The improved process flow	90
A.1 The complete reference model	97

List of Tables

1.1	Business intelligence market share and revenue	6
2.1	DIKW hierarchy definitions	28
2.2	Asset management key performance indicators	34
2.3	Intrinsic data quality metrics	37
2.4	System data quality metrics	37
2.5	Operational data quality metrics	38
4.1	Tag data scales and their description with examples in italics . . .	64
4.2	Aggregation Periods	66
5.1	Report Table	83

Chapter 1

Introduction

This study is about data quality in asset management decision-making. The author was first exposed to the current state of collecting, storing, analysing and reporting data in organizations at a large diamond mine in Southern Africa. From the literature and subsequent personal experience, it became clear that data quality is a common challenge faced by many organizations, especially in the asset-intensive industry. With the ISO 55000 series of international standards for asset management being published in February 2014, the time was ripe for a fresh look at this old problem. A pragmatic approach was adopted to develop a framework for addressing data quality issues in asset management decision-making.

In this first chapter, Section 1.1 provides an overview of the asset management landscape and paves the way for the rationale of the study in Section 1.2. Section 1.2 concludes with the articulation of the research problem. A summary of the research objectives, questions and strategy is given in Section 1.3 and delineated in Section 1.4. This chapter concludes with an outline of the remainder of the structure for this thesis in Section 1.5.

1.1 Background

The purpose of this section is to provide the background for the rationale of the study presented in Section 1.2. Core concepts related to asset management are briefly introduced followed by a discussion on the broader environment in which asset management is practised. Particular reference is made to data quality, standards and its role in industries and the business intelligence market to highlight some of the complexities faced by asset managers.

Assets allow organizations to achieve strategic objectives and meet stakeholder needs. Managing assets optimally is crucial for organizations to remain competitive. Optimized asset management requires consistent decision-making on activities that impact asset-related risks, performance, and cost profiles. By

implication, management must have access to quality data as well as decision-making tools that result in objective, predictable, and consistent decisions.

Data quality has, however, become a complex issue. Information overload is a real problem. Global markets pressure organizations into making more decisions, more frequently. “Business intelligence” has become a thriving industry, but their products are rarely compatible with each other and may require extensive training to use. Problems such as inconsistently defined key performance indicators and ambiguous data are commonplace. These are just some of the challenges that add to the complexity of data quality. Data quality is no longer just a function of its existence, but a function of many interrelated factors (Lin et al. 2006).

In recent times of international cooperation (and competition), ever increasing access to data and automated processes, it has become essential for organizations (and applications) to be able to communicate (internally and externally) in terms that are well-defined and universally understood, even by computers.

Perhaps the simplest illustration of the importance of well-defined, universally agreed means of communicating data is ISO 1, the first international standard published by the International Standards Organization in 1951. ISO 1 simply specifies the standard reference temperature for geometrical product specification. Before this standard, due to thermal contraction and expansion of products, it was difficult to communicate geometric measures, especially since there was no international standard for communicating temperature at the time either. For the first time, organizations were thus able to communicate geometric measurements internationally, without any ambiguity (International Organization for Standardization 1997).

ISO is only one organization that is involved with standardizing. Many industries have their own trade associations or oversight bodies that publish their own standards and reference models:

- The Supply Chain industry have their universally recognized Supply-Chain Operations Reference-model (SCOR) that identifies more than 150 key indicators to measure supply chain performance (Supply Chain Council 2014).
- The World Wide Web Consortium (W3C) develops standards for the web (W3C 2014). They are an international community and work with member organizations to develop web standards (W3C 2014). These standards include the Hypertext Markup Language (HTML) for defining the content of web pages and Cascading Style Sheets (CSS) for specifying the formatting of HTML documents.
- The Association for Retail Technology Standards (ARTS) is responsible for the widely adopted retail standards. The four standards are: The Standard Relational Data Model, UnifiedPOS, ARTS XML and

the Standard Request for Proposal standard. These standards and their global adoption are the reason retail organizations enjoy almost universal hardware and data compatibility (NRF 2014).

- The Institute of Electrical and Electronics Engineers Standards Association is one of the leading standards-making organizations in the world. They have published more than 900 standards, with the most notable ones relating to the wired and wireless communication between electronic devices that enables the internet and worldwide web (IEEE 2014).

There is also no shortage of global, independent, non-profit organizations dedicated to the standardization of creating, managing and sharing data:

- The Organization for the Advancement of Structured Information Standards (OASIS) has participants from more than 600 organizations across 65 countries. Their standards include the OpenDocument Format for Office Applications, the Universal Business Language and several other standards for health, crisis, legal and government data (OASIS 2014).
- DAMA (Data Management) International defines ten “knowledge areas” and is currently developing the second edition of their “Data Management Body of Knowledge” framework to be published in 2015 (DAMA 2014).
- The Open Group (known for the TOGAF Enterprise Architecture) is currently revising their “Universal Data Element Framework (UDEF)”. UDEF aims to make it easy to describe data and enable interoperability (The Open Group 2014).
- The Semantic Web, sponsored by the W3C, provides a common framework for sharing and reusing data. Despite being far from a reality, many organizations have built their products based in whole or part on the basic building blocks described by the Semantic Web (2014). Some of the most widely adopted building blocks include:
 - Resource Description Framework (RDF): a general method for describing information
 - Web Ontology Language (OWL): a family of knowledge representation languages
 - Extensible Markup Language (XML): a markup language that defines a set of rules for encoding human and machine readable documents
- The International Organization for Standardization (ISO) has published various standards pertaining to data and information management. These

standards range from the minimal ISO 8601, which dictates the formatting of a date (yyyy-mm-dd), to the 28-part ISO 2382 family of standards for information technology and data processing.

The asset management industry is no exception: there are also several global, non-profit, vendor-neutral organizations dedicated to standardizing asset management, asset management systems and asset management data.

- MIMOSA (2014) develops open standards for operations and maintenance data in manufacturing, fleet and facility environments.
- The Exploration, Mining, Metals and Minerals Vertical Forum (hosted by The Open Group) develops architecture frameworks and reference models for industries involved with exploration, mining, metals and minerals (The Open Group 2014).
- POSC Caesar Association (PCA) promotes open standards for data and software integration and initiated ISO 15926, a standard for the “Integration of life-cycle data for process plants including oil and gas production facilities”. They are currently working on an OWL implementation of ISO 15926 (PCA 2014).
- The Institute for Asset Management (IAM) was the main driving force behind publicly available specification for the optimal management of physical assets (PAS 55) which preceded the ISO 55000 series of international standards for asset management (discussed in more detail in section 2.1).

Despite this abundance of standards, asset managers are still not getting the data they need to make decisions. This can be due partly to the relative newness of asset management standards, but if time was all that was required for standard adoption, the United States of America should be using the metric system by now: ISO already published the International System of Units in 1960!

When looking at examples such as the retail industry where the ARTS standards are almost universally adopted, the problem seems to be that standards only benefit an industry if a majority adhere to the same standard. The many (competing) standards available for asset management and asset management data (none of which has widespread adoption) makes the appeal of an international standard for asset management clear, especially since evidence point to standardization promoting competitive innovation (Panel for the review of the European standardization system 2010). At the same time, there are industries where organizations feel that standardization opposes innovation and competition. An example of this can be found in the browser industry: developers of browsers continually extend the W3C’s standards with features

that only their browsers can display. The technical committee responsible for ISO 55000 seemed to take the risk of suppressing competition and innovation into consideration when they developed the ISO 55000 series by only specifying what an asset management system should look like and leaving implementation details (the *how*) to individual organizations and consultants.

Despite the ISO 55000 series finally providing asset managers with a universal standard for asset management systems without prohibiting innovation, it will not be easy to get widespread adoption. The comic in Figure 1.1 does an excellent job of illustrating the complexities of standards. A new standard on its own, even one published by the International Standards organization, will not get asset management decision-makers the data they need.

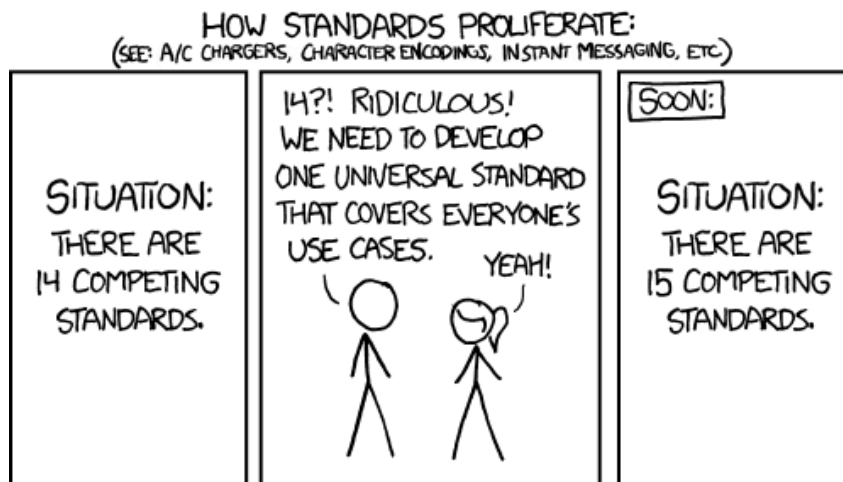


Figure 1.1: The problem of competing standards (Munroe 2011)

A further complication to the asset management environment is its reliance on the highly competitive business intelligence industry. Enterprise asset management (EAM) systems, decision support systems (DSS), computerized facilities management system (CFMS), computerized maintenance management system (CMMS), business intelligence (BI) and expert systems (ES) are all software systems commonly found in asset-intensive organizations. Table 1.1 shows the market share and revenue for 2013 of the top five business intelligence and analytic software vendors worldwide. IBM, Oracle and SAP are three of the top competitors in the business intelligence market and have shown consistent growth over the past five years.

A recent report from Gartner (2014a), however, shows that it is the smaller companies that showed the biggest growth resulting from the increasing demand for data discovery and analytics. These “other” companies account for almost 30% of the market. Compare that with the browser industry: after subtracting the big four (Internet Explorer, Firefox, Chrome and Safari) the “other” browsers make up around 2% of the market. Yet, despite HTML

Table 1.1: Business intelligence market share and revenue in millions of US dollars (Gartner 2014a)

Company	Revenue 2013	Market share (%)
SAP	3,057	21.3
Oracle	1,994	13.9
IBM	1,820	12.7
SAS Institute	1,696	11.8
Microsoft	1,379	9.6
Other	4,422	30.8

and CSS being standards controlled by a vendor-neutral consortium and there only being four dominant browsers, websites displaying messages like the one in Figure 1.2 is common.

**Figure 1.2:** "Please consider switching browsers" – Example of website using non-standard browser specific functionality.

Fortunately, the top four browsers are all free and easy to install and users can easily import and export their browsing history and bookmarks across browsers. For asset managers, it is not so easy as "vendor lock-in" is a common problem due to the high capital invest and incompatibility of business intelligence systems. Yet, the abundance of asset management data standards allows vendors to truthfully advertise their products as "standards compliant". For instance, Oracle and IBM (two of the top business intelligence vendors) are platinum members of The Open Group (2014), while Ivara (a CMMS vendor) and Invensys (an industrial automation software and analytics specialist) are both corporate members of MIMOSA (2014). As if this was not bad enough,

Gartner (2014b) recently found that, despite the growing market, “governed data discovery — the ability to meet the dual demands of enterprise IT and business users — remains a challenge unmet by any one vendor”. This means that many asset-intensive organizations have systems from more than one software vendor that does not interoperate for storing, processing and reporting asset related data. Section 2.4.1 describes these systems in more detail.

The recently published ISO 55000 family of standards, thus, provides a much needed “universal” standard for asset management and signifies the gaining recognition and maturity of asset management. However, with the technical details (such as an asset management data model) left for organizations to figure out, implementation of an ISO 55000 compliant asset management system still poses a significant challenge.

The next section discusses the rationale for a study to help asset managers address some of these challenges.

1.2 Rationale for the study

This section presents the rationale for the study from both a practical and scientific point of view, culminating in an articulation of the research problem that the study addresses.

As discussed in the previous section, asset managers finally have an international standard to guide them in the establishment, implementation, maintenance and improvement of an asset management system. However, as the ISO 55000 series was designed to be used in conjunction with other standards, “sector-specific, asset-specific, or activity specific technical requirements” were deliberately omitted (ISO 55000 2014, Section 0.2).

One of these sector-specific standards is ISO 14224, a standard for the “collection and exchange of reliability and maintenance data for equipment in the petroleum, petrochemical and natural gas industry”. ISO 14224 states the following in its introduction:

“Data collection is an investment. Data standardization, when combined with enhanced data-management systems that allow electronic collection and transfer of data, can result in improved quality of data for reliability and maintenance.”

Unfortunately, with the abundance of these technical standards and the many incompatible business intelligence platforms, this means there is still a significant challenge to implementing an asset management system. Section 1.1 made it clear that developing yet another technical standard, or designing another business intelligence platform for the highly competitive market is not the solution. Yet, data quality, which is what these standards and platforms are supposed to address, is a key issue.

Gartner, an information technology research and advisory group, frequently report on the state and impact of data quality in their reports:

- “More than 25 percent of critical data in the world’s top companies is flawed.” – Gartner (2007)
- “In [another] study, 36% of participants estimated they are losing more than \$1 million annually because of data quality issues” – Friedman (2010)
- “Poor data quality is a primary reason for 40% of all business initiatives failing to achieve their targeted benefits” – Friedman and Michael (2011)

Lee et al. (2002) observed that, despite years of research, the field of information management still lacked “comprehensive methodologies” for assessing and improving information quality. To address this issue they developed a methodology for “information quality assessment”.

There have also been a few studies directly aimed at data quality in asset management. These studies generally follow a common theme: assets are important, but lack of quality data impedes the optimal management of these assets. Consequence resulting from mismanaged assets range from operational annoyances to organizations shutting down. These studies all had slightly different focus areas and proposed different solutions. Saša Baškarada et al. (2006) propose a maturity model for assessing the quality of asset information and stated the problem as:

“The process of managing engineering assets is profoundly affected by the information used to make relevant decisions. Thus, poor quality information adversely affects EAM, potentially resulting in great financial losses”

Lin et al. (2006) used interviews to identify common problem areas and developed a data quality model based on this knowledge. Again, data quality was a core issue:

“There is strong evidence that most organizations have far more data than they possibly use; yet at the same time they do not have the data they really need.”

Likewise, Rokstad (2012) noticed the reliance of a “wide array of tools aiding asset management strategies” on “data about the assets”, which simply was not available. He suggests a “diary database” for capturing all life cycle data and a cost benefit analysis for selecting tools.

Common themes identified from these studies and personal experiences include:

- too much (unprocessed) data;

- too many competing systems with too little compatibility between them;
- low asset management maturity resulting in too much fire-fighting;
- data quality and asset management seen as low priority in boardrooms;
- departmental silos inhibiting communication;
- lack of skills/knowledge;
- all-in-one systems require big capital investments; and
- too much faith in software (the business intelligence vendors provides software, not solutions).

David Berger (2010) sketches an apt picture of the current situation when he observes that:

“It’s quite astonishing just how reliant management is on a CMMS (computerized maintenance management system) to provide information such as budget variances, asset availability and performance, energy consumption, payroll hours consumed, work backlog and so on. Yet despite our thirst for information, there’s sometimes little thought as to where the data is coming from and whether it reflects reality. It’s our inexplicable blind faith in technology that is our weakness - as if anything the CMMS outputs to screen or paper must be accurate because a computer processed it. As many maintenance managers discovered over the years, the quality of data input into the CMMS can be sadly lacking.”

The problems listed above could have severe impact on data quality and an organization’s ability to make informed decisions. Low quality data in an asset management system can be characterized by asset managers as:

- spending time and money measuring non-statutory metrics that are never used in decision-making;
- compiling reports that are never read;
- duplicating measurements;
- calculating the same KPI using different standards;
- storing data in incompatible systems;
- not having critical data in an accessible format and
- implementing expensive systems to automate data analysis without understanding what the results mean.

Vigon and Jensen (1995) conducted a comprehensive survey of database practitioners to determine how these common data quality issues were addressed. Vigon and Jensen found that application of “quality assessment procedures [were] by no means widespread, uniform or rigorous”. At the same time, they found that database practitioners were “aware of the need for assessing and maintaining data quality”.

The discussion above makes it clear that data quality is an important issue in asset management. Although some solutions have been suggested, none of them claims to be a complete solution. This still leaves asset managers to decide where to begin and how to prioritize their data quality improvement efforts. In “Asset Management – an anatomy” The Institute for Asset Management (2014) supports the notion that organizations must “prioritize” data collection and data cleaning activities:

“The quality of Asset Data & Knowledge should be assessed, understood and managed in order to ensure that it provides effective support to business decision making and processes. Typically, asset-intensive organizations do not have all the asset information they would ideally require, and the information they have may not be to the required quality. Therefore, organizations will need to assess and prioritize data gathering and data cleansing activities to focus on areas that will be beneficial” – Institute for Asset Management (2013)

Before data quality issues can be prioritized, they must first be identified. Thus, the research problem addressed in this study can be formulated as:

There currently exists no practical framework for helping asset management decision-makers identify and prioritize data quality issues in their organization.

1.3 Research objective, questions and strategy

The purpose of this section is to state the primary research objective and research question of the study. Additional solution objectives and secondary research questions are also identified.

Based on the rationale for the study as discussed in the previous section, the research objective of the study is:

To develop a framework for identifying and assessing data quality issues in asset management decision-making

In order to achieve this primary research objective, the following primary research question is considered:

What would a framework for identifying and assessing data quality issues in asset management decision-making look like?

To answer this question, a study within the philosophical paradigm of pragmatism is conducted. Within this pragmatic paradigm, a qualitative approach with a broad literature review and a case study for research methods is selected to drive an iterative design cycle. Chapter 3 presents the research methodology in more detail.

To ensure relevance and innovation, prior literature on the three knowledge domains addressed in the research question (asset management, decision-making and data quality) and their interfaces are studied. The following questions are considered to keep the literature review focussed:

- What is asset management?
- What role does decision-making play in asset management?
- How are decisions made?
- What role does data play in decision-making?
- What is data and when is it considered to be of “good” quality?
- How can information systems help maintain quality data?

The literature review is presented in Chapter 2. Once these questions have been answered, the solution objectives are formulated and the framework is designed and developed in Chapter 4.

To validate whether the framework can successfully identify and assess data quality issues as per the research objective, it is tested in practice. This demonstration and evaluation of the framework is presented in Chapter 5.

1.4 Delineation

The purpose of this section is to limit the scope of the study through various delineations. This is done in order to manage expectations regarding findings of the study.

“Not everything that counts can be counted, and not everything that can be counted counts.”

— Albert Einstein

This thesis is about the intersection of what counts, and what can be counted. This means that the study acknowledges that data is just one aspect of asset management and strategic, cultural and human resource factors are not considered.

Likewise, decision-making is a vast and complex field. This thesis will not attempt to cover decision-making in its entirety; rather, it will focus on the data aspect of decision-making. Decision-making is only considered in its context as the eventual goal of data collection and processing.

Although analytical tools and methods for analysing data play an important role in processing data and are mentioned in this study, the intention

is not to make a comprehensive study of analytical tools for asset managers. Mentions are made and examples of tools are used only to portray the different requirements of data, and how data quality affects the outcome of analyses using these tools.

Finally, the time-frame of the study is another limiting factor. Even though a case study was performed, the time-frame did not allow for a thorough analysis of the effectiveness of the framework. For instance, the long-term effect on the organization's financial record could not be studied.

1.5 Structure of the document

The methodology used in this study (documented in Chapter 3) consists of interlinked activities that were executed in parallel over several iterations of designing, developing and evaluating a solution to the problem described in this chapter.

To document this asynchronous and iterative development process, the following document structure was chosen:

Chapter 1: Introduction

Chapter 2: Literature Review

Chapter 3: Research Methodology

Chapter 4: Proposed Solution

Chapter 5: Case study

Chapter 6: Closure

The first three chapters provide the foundation for this study; they document the real world problem, the theoretical context of the problem and the research methodology used to derive a solution. Chapter 4 documents the solution objectives and the resulting framework for identifying and assessing data quality issues in asset management decision making. The application of this framework (to solve a part of the original real world problem) is documented in Chapter 5. The concluding chapter, Chapter 6, reflects on the study and makes future recommendations.

In keeping with the outlined structure of this document, the next chapter documents the literature review of this study.

Chapter 2

Literature Review

The problem described in Chapter 1 is not unique to asset management, yet the solution for it might be unique to asset management. Industries share so many characteristics that it is possible to learn from other industries' solutions, but merely copying and applying it to asset management is likely to fail. For example, in other industries such as the retail industry, there is much more standardization that would allow a solution to rely only on a few compatible standards. In asset management, however, several competing and even incompatible standards would need to be considered. This chapter studies the context of asset management and its related disciplines to build on prior knowledge and ensure innovation.

Section 2.1 reviews the history and context of asset management and gives a summary of ISO 55000. Decision-making and how it relies on data is discussed in Section 2.2 while data quality is investigated in Section 2.3. An overview of information systems and their development considerations is also given in Section 2.4.

2.1 Asset Management

The primary context of this study is asset management as defined in the recently published ISO 55000 series on international standards:

Asset management translates the organization's objectives into asset-related decisions, plans and activities using a risk-based, information driven planning and decision-making process.

– ISO 55000 2014: 2.1

This section provides an overview of asset management, its definitions, its long journey to how it is perceived today and its benefits and challenges.

2.1.1 Overview

An asset is an item, thing or entity that has potential or actual value to an organization.

– ISO 55000: 2.3

An asset can be tangible or intangible, financial or non-financial. It includes human, information, financial and physical assets (PAS 55-1 2008). *Assets* are what enable organizations to achieve their objectives. In today's global, competitive market, optimal management of these assets has become vital to organizations sustained competitiveness and operation.

Managing assets is not a new concept, but the purpose and scope of asset management have significantly evolved over the years. The release of the first series of international standards aimed at asset management in 2014 (ISO 55000, ISO 55001, ISO 55002) signals a new era for asset management. This series is specifically aimed at asset management and systems for managing asset management (asset management systems).

At its core, the purpose of the ISO 55000 series is to enable organizations to achieve their objectives through effective and efficient management of assets. The practices outlined in these standards can be applied to the “broadest range of assets, to the broadest range of organizations, across the broadest range of cultures” (ISO 55000 2014).

Even though the ISO 55000 series has only recently been released it seems to be quickly gaining in popularity. Figure 2.1 shows the Google search volume for PAS 55 (blue) and ISO 55000 (red) over the last seven years.



Figure 2.1: Comparison of Google search volume for PAS 55 (blue) and ISO 55000 (red)

The next section, Section 2.1.2, gives an overview of the origin of asset management. Section 2.1.3 summarizes the definitions, requirements and benefits of ISO 55000.

2.1.2 The Rise of Asset Management

After World War II, the world experienced the Golden Age of Capitalism (Marglin and Schor 1991). During this period, most of the world experienced high economic growth. Industries were pressured by national and global demands to expand rapidly. New materials, chemicals and technologies emerged at an accelerated pace. Processes were automated to cut labour costs. Machines became more complex, operating speeds increased and equipment was subjected to heavier loads and longer operating hours. Industries were producing at record rates, but regulations, standards and maintenance strategies were still stuck in the pre-war era (Brown and Sondalini 2012).

The lagging of regulations described above is a common example of the “pacing problem” in innovation (Marchant et al. 2013). Technological innovation, pushed by economic and social incentives, is inherently fast. Governmental regulators (and other oversight bodies) are inherently slow (reactive) and typically require external “motivators”. Unfortunately, as will be discussed in the next few paragraphs, these motivators can be quite severe in the asset-intensive industries.

In 1968, 14 000 Americans were killed and 2.5 million injured in work related accidents (Percy 1980). On 29 December 1970, the Occupational Safety and Health Act was signed into law. It was also during this period that the need for better maintenance led to planned preventative maintenance (fixed interval maintenance) derived from the belief that equipment follow the bath-tub-curve when failing (high frequency of failure during initial use and with old age) (Brown and Sondalini 2012).

With safety standards in place, the next motivator for regulations became lack of quality. After more than 20 years of sustained growth, the world was hit by the 1973 oil crisis. This event had profound impacts around the world and drastically changed manufacturing, especially in the automotive industry. Smaller, more efficient cars imported from Japan became a threat to western car manufacturers, particularly since Japanese manufactures were achieving very high levels of quality. To compete, Ford Motor Company launched a quality initiative in the 1980s led by Edwards Deming, who many consider as the father of modern quality management (Gabor 1990).

Competition, international trade and subsequent outsourcing led to the eventual creation of the ISO 9000 family of standards, one of the most widely used management tools in use today (Bird 2010). The demand for quality required equipment that was more reliable than what maintenance strategies commonly used at the time could achieve. To improve their reliability, the manufacturing industry looked towards the aviation industry.

The aviation industry has been questioning the bath-tub-curve belief as early as the 1960s. United Airlines (Smith 1993) studied the different failure characteristics of equipment and observed six different patterns of failure. They also found that 89% of non-structural equipment on aircraft fails at

random intervals. As such, achieving the required level of reliability using preventative maintenance was not financially feasible. Out of necessity, thus, the aviation industry developed a new technique for maintaining their equipment and defined it in MSG-1 (Maintenance Steering Group 1). It is from this study that condition based maintenance and reliability centred maintenance (RCM) was born. Variations of RCM and condition based monitoring proved to be just as successful in other industries.

With safety and quality regulated, the third major motivator became the environment. In the years following the energy crises of the 1970s, several major environmental disasters such as the Shell Niger Delta oil spills since 1976 and the Exxon Valdez oil tanker running aground in 1989, increasingly drew attention to the impact of asset-intensive organizations on the environment. Growing awareness of global warming and the need for environmental sustainability lead to a third relevant standard: the ISO 14000 series of standards published in 1996.

Until recently, these three series of standards, OSHA 18000, ISO 9000 and ISO 14000 provided asset-intensive organizations with an “auditable golden triangle” within which to operate (Fogel 2013). But today’s equipment is becoming more and more complex, global competitiveness is getting fiercer and shareholders are demanding reduced costs. All of this is increasing risk towards the business. These circumstances require organizations to manage risks. At the core of safety, quality and environmental risks lies physical assets. Yet, none of the standards discussed above directly addresses the management of physical assets.

The Institute for Asset Management identified this gap in the available standards and in 2008, the British Standards Institute published the publicly available specification for asset management (PAS 55). Following on the success of PAS 55, the International Standards Organization released the ISO 55000 series in 2014.

The ISO 55000 series is meant to fill the gap left by the golden triangle: OSHA 18000 manages risk to employees, ISO 9000 manages risk to customers, ISO 14000 manages risk to the environment and ISO 55000 manages risk to the organization.

R. Moore (2011) provides evidence that there is a correlation between reliability, safety, cost effectiveness and environmental soundness. This is no surprise, as PAS 55 and ISO 55000 was designed to integrate closely with the safety, quality and environmental standards. With today’s renewed focus on environmental issues, safety issues, quality issues and global economic uncertainty, an integrated asset management solution is critical for any organization’s sustained competitiveness.

2.1.3 Asset Management Today: ISO 55000

“Asset management translates the organization’s objectives into asset-related decisions, plans and actions, using a risk based, information driven, planning and decision-making process.” ISO 55000 2014, Section 2.1 and Section 2.4.2(b)(1).



Figure 2.2: Asset Management as defined by ISO 55000

Today, asset management can be concisely defined and “the broadest range of organizations across the broadest range of cultures” can implement and benefit from an ISO 55001 compliant asset management system for the “broadest range of assets” (ISO 55001 2014).

This section details the benefits of an asset management system, puts the data quality issue in context of the rest of asset management field and looks in more detail at the fundamentals, elements and requirements of an asset management system as set out in the ISO 55000 series. When correctly implemented and integrated, asset management benefits the whole organization. It improves financial performance, risk management, safety, efficiency, sustainability as well as numerous other aspects of an organization.

Before discussing the benefits in more detail below, it is important to understand what an asset management system is and where data and decision-making fit in. Figure 2.3 gives an overview of the conceptual model of asset management that can be found in “An Anatomy of Asset Management” published by the IAM (2012).

The conceptual model consists of six subject groups, namely: “asset management strategy and planning”, “asset management decision-making”, “life cycle delivery activities”, “asset knowledge enablers”, “organization and people enablers” and “risk and review”. Of importance to this study is mainly the asset knowledge enablers and asset management decision-making subject groups, each relating to a number of subjects. For example, under the asset knowledge enablers, the subjects are asset information strategy, asset knowledge standards, asset information systems and asset data & knowledge. Under all six subject groups, there are 39 subjects in total. This model shows that

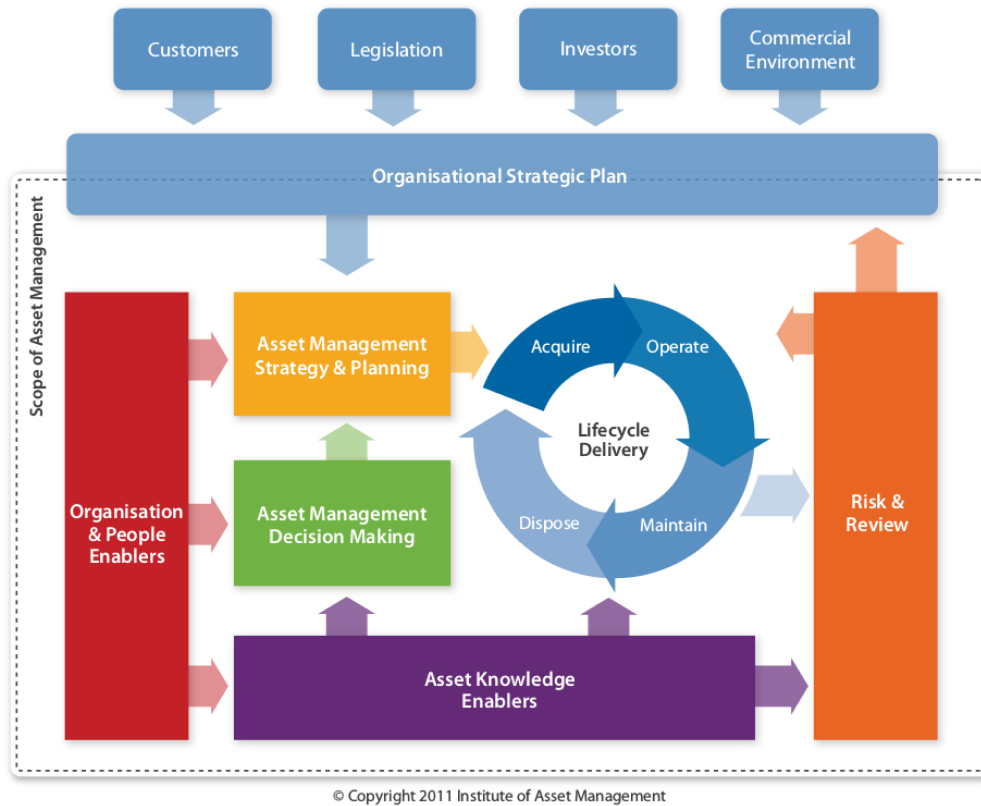


Figure 2.3: The IAM Conceptual Model of Asset Management

asset knowledge is a key enabler of asset management, and specifically, asset management decision-making. Yet, it is only one subject group out of six.

ISO 55000 similarly makes it clear that asset management relies on accurate data while acknowledging that it is only part of the bigger asset management system:

Asset management requires accurate asset information, but an asset management system is more than a management information system

– ISO 55000 2014, p. 2.5.1

ISO 55000 also identifies the fundamentals on which asset management is based in Section 2.4.2. There are four fundamentals, namely value, alignment, leadership and assurance. In identifying these fundamentals, this thesis will only focus on the parts that are relevant to the process of decision-making.

Value: asset management exists to provide value to the organization, especially through the establishment of decision-making processes that reflect stakeholder needs and define value.

Alignment: Asset management entails the making of technical, financial and operational decisions. These decisions entail the implementation of

risk-based, information-driven, planning and decision-making processes that transform organizational objectives into asset management plans.

Leadership: this principle includes clearly defining responsibilities and making sure that it relates to the right people.

Assurance: There exists a need in organizations for effective government. In order to achieve this it needs to implement processes for monitoring and continual improvement.

Building on these four fundamentals, ISO 55000 presents four core benefits of an asset management system. These benefits are:

- **Creating an asset management system provides benefits for itself.** Because an asset management system is data intensive, it is often necessary to use new tools to collect and manage data. By creating such new tools, the organizational knowledge and decision-making can be improved.
- **Top management benefits from new insights and cross-functional integration.** An asset management system can help the top management to understand assets, their performance and the risks associated with the management of assets better. They will also understand the asset value as an input to decision-making and organizational strategic planning better. Asset management further supports energy and environmental management and sustainability.
- **Financial functions benefit from improved data and linkages.** Because a strategic asset management plan provides a more effective taxonomy, it enables an integrated financial and technical view of asset systems.
- **Many parts of the organization benefit from an asset management system.** In order to achieve the above-mentioned benefits, asset management must be integrated. It is often the case that data in certain systems are isolated from that in other systems. An asset management system allows the integration of the data, which can lead to new information and improved decision-making.

The fundamentals and benefits listed above give an indication of just how important asset management is to the successful functioning of an organization. Next, the elements of an asset management system, as set out in Section 2.5.3. of ISO 55000 will be briefly discussed. The seven elements of an asset management systems are:

1. Context of the organization
2. Leadership

3. Planning
4. Support
5. Operation
6. Performance evaluation
7. Improvement

Under the support element, it is clear that the asset management system provides information to support the development of asset plans and to evaluate their effectiveness. Because asset management systems can be very large and complex, there are frequently issues involved in the collection and consolidation of asset data. A crucial function of the asset management system is thus to create, control and document the information from the assets. It is furthermore clear that organizations have to evaluate the performance of their assets and asset management. However, this is often an indirect and complex process. Effective data management and the transformation of data to information are key to measuring asset performance. The evaluation of performance should furthermore be a continuous process. In the evaluation process, the asset management system should be evaluated against set objectives, and cases where the objective was either not met or exceeded have to be examined.

After the preliminary listing of these elements in ISO 55000, the requirements for each of the elements are discussed in ISO 55001. The elements and their requirements, where relevant to this study, are presented below.

Under the first element, namely context of the organization, the only relevant requirements relate to understanding the needs and expectations of stakeholders. In this context, it is important that the organization determines criteria against which asset management decision-making can be judged, and also determines what the stakeholder requirements with regard to recording information relevant to asset management are.

For the planning element, the relevant requirements are those related to the asset management objectives (they must be measurable and monitored) and the planning to achieve these objectives (which requires identifying the method and criteria for decision-making and then prioritizing the activities and resources).

Regarding the operation element, the most important requirement is that the organization keeps documented information to the extent necessary to indicate that processes were carried out according to plan. This can only happen if data from processes are accurately captured and stored.

The performance evaluation element has requirements relating to monitoring, measurement, analysis and evaluation that are relevant, and these include questions such as what needs to be monitored, the methods used to monitor and when it shall be monitored. These questions have significant impact on the availability and quality of data.

The information requirements under the support element for an asset management system are especially relevant to this study. In the process of determining the information requirements, the organization must look at the significance of identified risks, roles and responsibilities for asset management, and the *impact of quality, availability and management of information*. It is also imperative, in this regard, that the organization determines the attribute requirements of identified information, the quality requirements of the information and how and when information is to be collected, analysed and evaluated. Thereafter, the organization has to implement processes for managing data and must ensure that there is *consistency and traceability between data*. Lastly, the organization must document their information flows and processes. These requirements go to the core of this study.

From the discussion above, it is clear that data (“asset knowledge”) plays an important “supporting” or “enabling” role. Figure 2.3 also shows that there is an important interface between data and decision-making. ISO 50000 also repeatedly state the “information driven” aspect of asset management decision-making. The next section is, therefore, dedicated to the anatomy of a decision and how data influences decision-makers and the decision-making process.

2.2 Decision-Making

Decision-making is a very broad field studied by many disciplines. This section gives a high level overview of decision-making and the different views on this subject, discusses process models in detail and discusses the role of data in decision-making.

2.2.1 Overview

Decision-making has many definitions, but it essentially is the choosing of an option. A decision signals the end of a process: it irreversibly eliminates the other options. It is for this reason that Perdicoúlis (2012) considers a decision “the most valued, appreciated or important part” of many professional activities. Towler and Keast summarize decision-making as the period preceding implementation, preceded by a process of “inference from the evidence”.

In complex processes, there might be several decisions before the *final* decision, but each decision will reduce the available options or influence the final decision. Perdicoúlis observes that decision-makers occupy the top ranks of their professions. They are typically the ones that determine the actions of the organization. “With great power, comes great responsibility” rings especially true for decision-makers. It is, thus, not a surprise that amongst the many qualities looked for in decision-makers, the one considered most crucial is: the decision-maker must be informed (Perdicoúlis 2012).

This simple notion forms part of the foundation of this study: a decision-maker must be informed. To understand what an informed decision entails, it is necessary to understand how decisions are made (Perdicoúlis 2012). The importance of understanding how decisions are made is also emphasized in asset management standards (ISO 55001 2014, Section 4.2.4)

The field of decision-making is predominantly concerned with two types of decision-making models: prescriptive and descriptive (Bazerman and D. A. Moore 2012). Prescriptive models, common in the field of engineering, are built on the assumption of rationality and typically prescribe mathematical models for optimal decision-making. A typical example of a prescriptive decision-making model might follow these general steps (Hammond et al. 2002):

1. Define the problem
2. Identify the criteria
3. Weight the criteria
4. Generate alternatives
5. Rate each alternative on each criterion
6. Compute the optimal decision

These types of models are invaluable for consistent decision-making and are well suited for mature organizations in asset-intensive industries. This study will later discuss useful numeric tools that asset managers can utilize to help with rational decision-making. In reality, however, few organizations have the infrastructure and support in place to take the time required for the steps outlined above.

To benefit these organizations, it is thus important to also understand how decisions are *actually* made. Descriptive decision-making models attempts to do just this, study how decisions are actually made. Due to its human nature, these models are usually found in the social and business sciences. Researchers studying descriptive decision-making models typically distinguish between two primary “functions” or methods of decision-making. These two methods loosely correlate with the amount of “thought” that is put into making the decision.

Researchers from different fields have developed different theories to describe these two methods. Stanovich, a professor in applied psychology and human development, called it System 1 and System 2 thinking (Stanovich 1999). Argyris, a business theorist, called it Single Loop and Double Loop Learning (Argyris 1977). Perdicoúlis, a professor in engineering, relates this to decision-makers’ role as creator of decision options and criteria and their

role as selector of the best option as per the criteria. There are subtle differences between these theories, but a full analysis thereof falls outside the scope of this study.

Information flow models are useful for visualizing how information flows during the decision-making process and will be used to better illustrate the various methods of decision-making.

2.2.2 Information Flow Models

In the simplest form of decision-making, the decision-maker is a “selector” drawing information from two sources: the available options and the criteria for selection. The first information flow model (Figure 2.4) illustrates the case where these options and criteria (“rules”) are fixed.

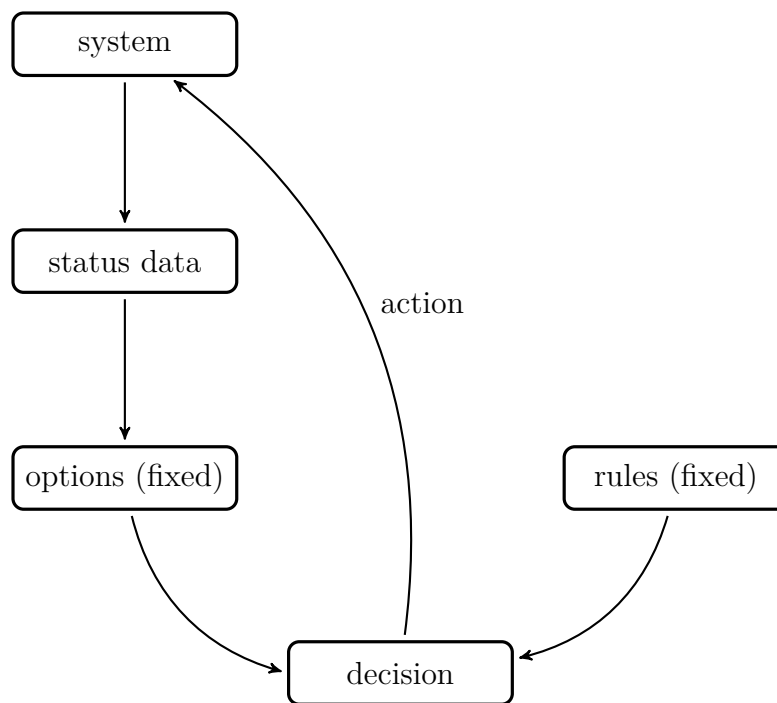


Figure 2.4: The Data Learning Model

Perdicoúlis describes the type of decision-making illustrated in Figure 2.4) as the “Data Learning Model”. This type of decision-making is commonly found at the operational level of asset-intensive industries. For example, an operator of an automated manufacturing process will continuously monitor her control panel indicating the status of the equipment. If any indicator turns purple, the whole process must be stopped. If an indicator turns red, only the associated equipment needs to be stopped and if it is green, no action is required. In this example, there are only three options, stop the whole process,

stop only specific equipment or no action. The decision to determine which action to take is determined by the rules outlined above.

The Data Learning Model can be considered a variation of the Recognition-primed Decision Model described by Klein (1998). This second model essentially describes an “if [status] then [option]” reaction that even inexperienced decision-makers can follow. Another variation of the Recognition-primed Decision Model is where an expert “recognizes” a situation and, primarily from experience, determines the action to be taken (instead of following explicit rules). This loosely correlates with System 1 thinking where decision-making is “fast, automatic, effortless [and] implicit” (Stanovich 1999). In the example of the operator monitoring an automated process, this type of decision-making might occur when a specific indicator stops emitting light: from experience, the operator knows that this requires restarting the associated equipment.

The second variation of the Recognition-primed Decision Model described above introduces another element of decision-making: the mental model. Mental models represent a decision-maker’s personal understanding of the situation (or more general, their view of the world) (Towler and Keast 2009, p. 13). These “cognitive maps” describe not only the decision-makers understanding and knowledge of the situation, but also their assumptions and generalizations.

Mental models allow decision-makers to make fast and instinctive decisions. Perdicoulis dubbed this decision model “The Imprinted Model” (depicted in Figure 2.5).

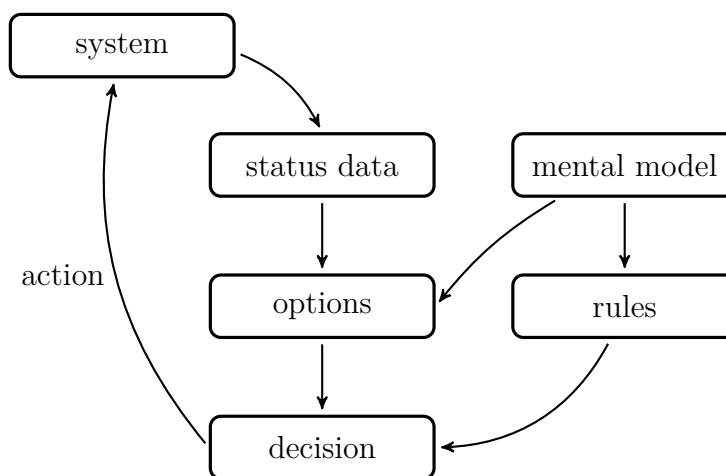


Figure 2.5: The Imprinted Model

Both the models referenced above exhibit characteristics of System 1 thinking. Fortunately, the assumption of rationality in prescriptive decision-making models is not just theoretical. Researchers have also observed decision-making that is more logical, calculating, slow and conscious, characteristics of System 2 thinking (Kahneman 2013).

System 2 thinking is observed when decision-makers “update” their mental model by studying the system. This means that decision options (alternatives) and decision criteria (rules) are influenced not just by the status data, but also by the *reason* for the status data (the causal data). An example of this decision-making is when (referring back to the example of the operator of the automated production process) the operator has to decide what to do after switching off a machine whose indicator turned red. Her options are: call the engineer, call the electrician or record the incident for later attention. From experience she knows that this specific machine usually indicates red when the v-belt becomes misaligned, but instead of making the instinctive decision to call engineering, she decides first to inspect why the indicator turned red. Upon inspection, she observes that the cause is a blown fuse which prompted her rather to call the electrician.

Perdicoúlis calls the decision-making process exhibited in the example above the “Deeper Learning Model” and represents System 2 thinking by the flow of information (causal data) from the system to the decision-maker’s mental model (see Figure 2.6).

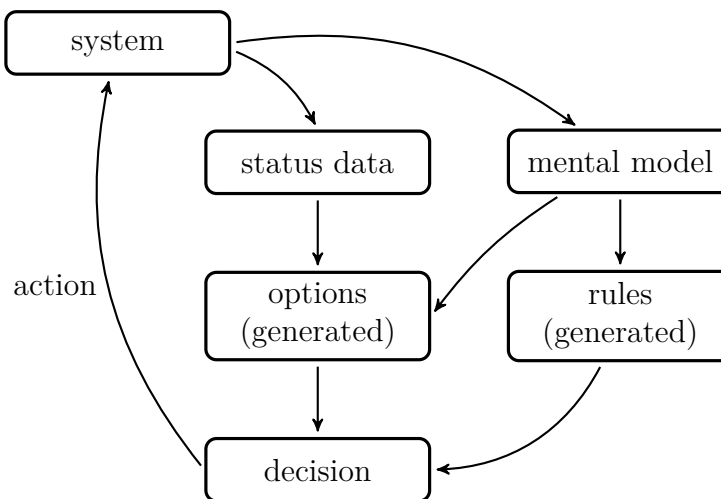


Figure 2.6: The Deeper Learning Model

All the models presented up to now embody Single Loop Learning. Single Loop Learning, on the one hand, is the repeated attempt to solve a problem without changing the goal (Argyris and Schon 1978). The decision-maker stays in “selector” mode and does not re-evaluate the goals (which determine the decision criteria and feasible options). Double loop learning, on the other hand, is when a decision-maker, after evaluating the options or observing the consequences, challenges the objectives (typically by asking “why?”). The decision-maker switches from being the selector of options to the creator of options. Perdicoúlis describes this last model in his “Systems Thinking Model”.

To demonstrate the Double Loop Learning aspect of the Systems Thinking Model, consider the last example where the operator decided to call the electrician. The electrician might warn the operator that the blown fuse is likely due to a worn wire shorting at random intervals. The operator, who up to that point made her decisions with the criteria and options derived from the objective of getting the system up and running as soon as possible, might reconsider the objective. By updating her mental model with the new causal data from the system (worn wires) she might conclude that the objective should no longer be production orientated, but safety orientated. Based on this new objective (with new options and criteria) she decides to kill the power to that machine and submit a job request for full inspection of the condition of the machine.

2.2.3 The Role of Data

From the decision-making models presented above, it is clear that data plays an important role in decision-making. Particularly, it was illustrated how data plays two critical roles in the decision-making process:

1. the acquisition of “status data”, which initiates the decision-making process, and
2. causal data is used to “update” decision-makers’ mental models, which informs their decision-making.

In the first instance, the need for timely, accurate data is obvious: if asset managers are not being made aware of new developments in their plant, they will not be able to initiate a decision-making process that might lead to some action. The second role of data is summarized by a term well known in the fields of computer science and information technology: “*Garbage in, garbage out*”. The quality of a decision directly reflects the quality of information it was based on.

It is, thus, no surprise that Rud (2009), in his *Business intelligence success factors: tools for aligning your business in the global economy*, warns that “the penalties of not knowing are harsher than ever” as personal accountability “at the highest levels is not only prudent, it is now legally mandated”. Rud also makes the argument that information is no longer a by-product of a business, but “the lifeblood of business”.

The next section studies data and data quality in more detail to better understand the two roles of data, its life cycle and how the quality of data affects these roles.

2.3 Data Quality

Previous sections in this chapter dealt with the larger contexts of asset management and decision-making, how they relate and why they are relevant to the problem identified in Chapter 1. As discussed in Section 2.2, the quality of data (including its existence) is arguably the most important factors of good decision-making, especially in the context of asset management. This section expands on the concepts of data, quality and data quality in asset management decision-making.

2.3.1 Data

“Data is raw. It simply exists and has no significance beyond its existence (in and of itself). It can exist in any form, usable or not. It does not have meaning of itself.”

– Ackoff (1989)

For lack of a better word, the term “data” has, up until now, been used in the broadest sense to refer to all forms and stages of information pertaining to asset management decision-making. However, to understand data quality and all its intricacies properly, it is important to differentiate between the various incarnations of data that leads up to a decision.

This notion that a data “leads to a decision” forms one of the basic assumptions of this study and has been discussed in Section 1.4. By only acknowledging the value of data in terms of its ability to inspire action (a decision), the discussion of the concept of data can be simplified to four, closely related concepts: raw data (simply referred to as “data” for the remainder of this section) which can be viewed as the lowest level of abstraction, information which is derived from data, knowledge which is derived (in part) from information and wisdom which takes knowledge into account and results in a decision being made (which leads to action).

The idea that data is part of a hierarchy is an old one formalized in the data-information-knowledge-wisdom (DIKW) hierarchy (see Figure 2.7) frequently attributed to Ackoff (1989). In a comprehensive review of textbooks in information management, information systems and knowledge management, Rowley (2007) found that the DIKW hierarchy is often quoted, or used implicitly in definitions of data, information and knowledge. Rowley also notes that long-standing debates on the nature of information and knowledge have led to many variations (or even rejection) of the DIKW hierarchy and its definitions.

The use of the DIKW hierarchy in this study is thus intended as a tool to illustrate the role, context and importance of data in asset management decision-making, not as a discussion on the nature of knowledge. The DIKW hierarchy (Figure 2.7) and summary of definitions in Table 2.1 are adapted from Rowley’s review of various textbooks.

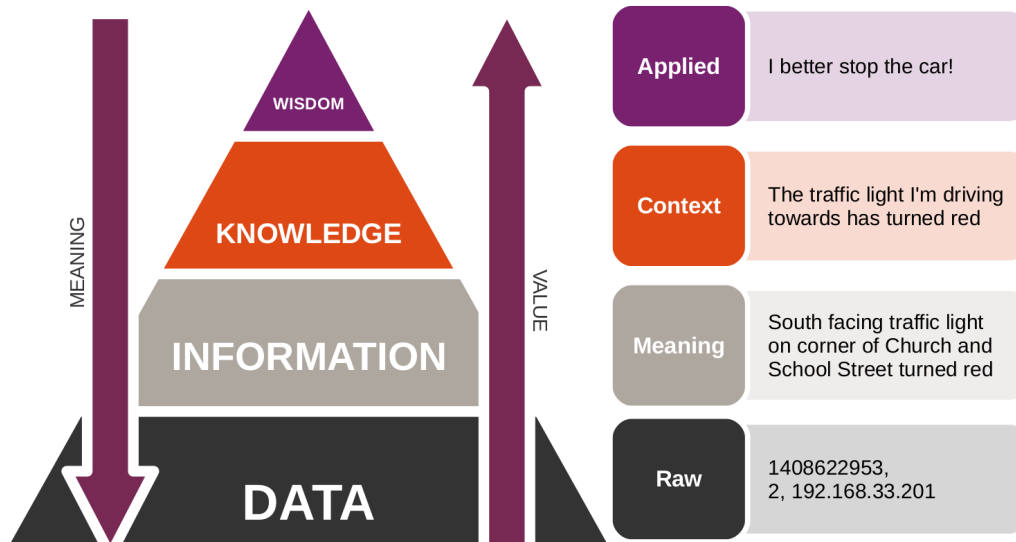


Figure 2.7: The DIKW hierarchy

Table 2.1: DIKW hierarchy definitions adapted from Rowley (2007)

Element	Definition
Data	The lowest level of abstraction, lacks meaning or value and is unorganized and unprocessed.
Information	Data that has been processed or interpreted for a specific purpose so that it becomes meaningful.
Knowledge	Knowledge is the synthesis of information, context and experience.
Wisdom	The capacity to put into action the most appropriate behaviour, taking into account what is known (knowledge) and what does the most good (ethical and social considerations).

Data quality is an important but often overlooked problem. Almost 20 years ago Wand and Wang (1996) identified the same issue. In their article they mention that an estimated 60% of organizations in a survey of 500 medium sized corporations with annual sales of more than \$20 million had problems with data quality. Wand and Wang attribute this to the rapid increase in complexity that comes with technological advancements.

Almost 20 years later, data quality issues are still abundant, if not more so than a few years ago. ISO 55000 (2014) highlights this issue:

“Asset information systems can be extremely large and complex in some organizations, and there are many issues involved in collecting, verifying and consolidating asset data in order to transform it into asset information. Creating, controlling, and documenting this information is a critical function of the asset management system”

– ISO 55000 2014, Section 2.5.3.5

The data life cycle gives valuable insight into the “extremely large and complex” reference above and is discussed below.

2.3.2 Data Life Cycle

Since “data” is such a broad concept, data life cycles are often highly context-specific and depend on an organization’s business processes. In a document published by the Committee on Earth Observation Satellites, their Working Group on Information Systems and Services presents several data life cycle models from various industries (CEOS WGISS 2011). Most of these life cycles were variations (with context-specific additions) of acquire, use, and maintain.

The data life cycle stages presented below is an adoption of the U.S. Geological Survey’s science data life cycle (USGS 2014). The life cycle stages are: acquire, process, analyse, preserve and share. These stages were chosen primarily for facilitating the discussion below on the complexities of data.

Planning: Data does not come into existence by itself. Building a data pipeline to drive the data life cycle and preserve data quality requires extensive planning. Designing and implementing a data pipeline across the life cycle of data requires a skilled data scientist. Unfortunately, it is also very easy for an organization to build such a data pipeline on an ad hoc basis, the results of which has been described in Chapter 1.

Acquiring Data: The data life cycle starts with its acquisition. The term “acquire” is used here to collectively refer to data collection, data capture and data creation processes. In a survey of database practitioners, Vigon and Jensen (1995) used the following broad categories to indicate the sources of data (ordered by popularity):

- Internal company data
- Technical trade association data
- Books and statistical compilations

- Industry Average Data
- Peer reviewed literature

In asset-intensive organizations, the majority (by volume) of data is automatically generated from digital or analogue signals captured and processed by industrial control systems. Another common source of internal company data is maintenance logs, safety logs and production logs. These are typically manually populated records of events or observations and might include unstructured comments and descriptions.

This combination of high volume, real-time, automated data streams coming from equipment and manual, unstructured logs of events and observations is unique to the asset-intensive industry. Both of these data sources come with their own complexities. When using data originating from industrial control systems, it is important to know the limitations of the sensors. For instance, the update frequency, interpolation techniques and noise filtering techniques can all influence the quality of the data. Knowing when a sensor was last calibrated and by whom might give an asset manager additional insight into the quality of the data.

For manual observation and event logs, metadata such as the “who” and “when” become important quality indicators for decision-makers. For example, a consistent deviation in one operator’s recording of equipment vibration might indicate that they require additional training to correctly use the measuring tool. Manual records also frequently require a supervisor to sign off on the record.

The discussion so far has focused on data from automated industrial control systems and data from manual records. In the first instance, the acquisition process is best described as *data collection*. This is because the data is being generated and stored by the industrial control system irrespective of whether it is being used. The second instance, describing manual data acquisition, is an example of *data capturing*. Events occur independent of whether they are recorded. For example, a mechanic can fix a broken gearbox, but if she does not record the event, then there is no data. The term “capturing” thus refers to the elusiveness of this type of data source.

The third method of acquiring data is *creating data* and corresponds with a much under-appreciated source of data: simulation. Changes to an asset-intensive organization typically require massive capital investments which makes simulating different scenarios a fast, inexpensive solution for creating “what-if” data that can be used for anything from production scheduling to asset investment decisions (Von Petersdorff and Vlok 2013).

Processing Data: Any digitizing, moving, copying, translating, cleaning or checking of data constitutes processing. Data processing also include managing and storing data.

In an asset management system that relies on many disconnected or incompatible subsystems (like the automated data streams and manual logs discussed above), a lot of data processing is required. Hand-written records must be digitized, spell checked, categorized and verified. Automated data streams may need to be normalized to account for calibration errors. Even the format of the date- and timestamps might require translation from one format to another. For example, old American control systems might report the date and time using the month-first convention (mm/dd/yyyy), while Unix based systems might record it as a unix timestamp (which is a running total of seconds since 1 January 1970).

Lack of training can also exponentially increase the amount of manual processing that occurs. For example, asset managers familiar with Excel might try to copy and paste data from one application to another, while asset managers without the required training might manually retype all the values.

Human fatigue, lack of skills and malicious intent, amongst other, makes manual data processing error-prone and slow. Organizations thus typically attempt to minimize manual data processing by acquiring business intelligence tools to automate data processing. Figure 2.8 shows a popular “data warehouse” architecture where data from various systems is processed and stored in a central “data vault”. Smaller “data marts” make subsets of the data available in a predefined format suitable for the particular analysis that is required.

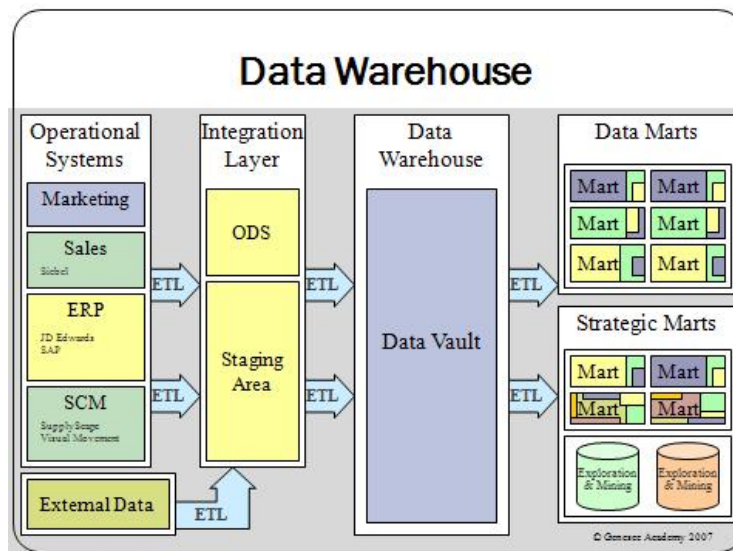


Figure 2.8: Overview of Data Stores

As discussed in Section 1.1, however, over-reliance on automated systems is just as problematic as manual data processing.

Analysing Data: Data analysis include any activity related to interpreting, deriving or forecasting data. The ever increasing volume of data being generated by organizations necessitates the use of tools for both processing and analysing data.

Understanding data and how it can be analysed, however, is not something that can be automated. Data analysts must be familiar with tools for analysing data and they must know how the data was acquired. In maturing asset management systems, this is often not the case. Several previous studies have attempted to identify simple tools that asset managers can use. Minnaar et al. (2013) presents a high level overview of potential numerical tools specifically relevant for asset managers implementing an ISO 55001 compliant asset management system. Burnett and Vlok (2014) developed a simplified numerical decision-making methodology for physical asset management decisions.

With knowledge of the various tools available for analysing data, care should be taken that the right tool is used for the right job. A popular typology proposed by Stevens (1946) that is frequently used in statistical analysis is “scale type” or “level of measure”. There are four levels of measure: nominal, ordinal, interval and ratio measurements. These “levels of measure” are frequently employed when making judgements on what type of analysis can be performed and what type of graphs can be used to represent this data.

Velleman and Wilkinson (1993), in a critique of Stevens’ topology, however, warn that:

“Scale type, as defined by Stevens, is not an attribute of the data, but rather depends upon the questions we intend to ask of the data and upon any additional information we may have. It may change due to transformation of the data, it may change with the addition of new information that helps us to interpret the data differently, or it may change simply because of the questions we choose to ask.”

Velleman and Wilkinson (1993) further argue that Stevens’ topology can be misleading if blindly enforced:

“Responsible data analysis must be open to anomaly if it is to support scientific advancement. Attempts to narrow the range of relationships that may be considered, restrict the transformations that may be applied, or proscribe the statistics that may be computed limit our ability to detect anomalies. Textbooks and computer programs that enforce such an approach to data mislead their readers and users.”

These two opposing forces (enforcing certain patterns versus allowing “creative” analysis) is common theme in business intelligence affecting both the

data analysis and the information system components. A later section, Section 2.4.2, elaborates on the information system component and how restricting user “improvisations” affect the success of the system.

Preserving Data: Traditionally, preserving data primarily dealt with filing and extending the life of physical documents. Today, preserving data is concerned much more with selecting appropriate file formats, data serialization formats, encoding standards and compression algorithms. Digital files do not “degrade” over time, but they are susceptible to corruption and data loss. An important aspect of data preservation is thus making backups of data. Despite the prevalence of online (“cloud”) backup services with capacities far exceeding most organizations’ needs, many still use magnetic tapes or similar physical on-site backup mediums.

Another (often neglected) function of preserving data is to future-proof it (Pinola 2012). Future-proofing is not just about minimizing the risk of data loss over years; it also ensures that data is compatible between different systems and protects organizations from vendor lock-in. With fierce competition and rapidly changing business models, an organization should be able to seamlessly upgrade or change their hardware and software, move to the “cloud” or bring their data back in-house (Satzger et al. 2013).

When dealing with large sets of structured data, a core enabler of future proofing and compatibility is data serialization. Unfortunately, software publishers often deliberately obfuscate or encrypt their data serialization formats to prevent competitors from creating compatible products. This means organizations must rely on their own serialization strategies for exporting and preserving data.

When serializing data, an entity, object or record is converted to human and machine readable plain text format in such a way as to preserve its attributes, data types and relationships. Using a well-defined serialization format, data can be stored, shared and eventually reconstructed to an identical representation of the original object.

Popular serialization formats include Extensible Markup Language (XML), JavaScript Object Notation (JSON), Structured Query Language (SQL), YAML Ain’t Markup Language (YAML) and Character-separated Values (CSV). Examples of each of these formats can be found in Appendix B.

A back-up strategy and archival requirements must also be in place together with metadata and documentation describing what data is stored where and in what format.

Sharing Data: The final stage of the data life cycle is to share it. Sharing data, however, is a balancing act of making it as easy as possible for authorized people to access it, while securing it against unauthorized access. Before sharing data, questions like these must be considered:

- Who may access the data?
- How will this restriction be enforced?
- What is the intended use of the data?
- In what format will it be shared?
- How will the data be promoted?

Sharing data is also about communicating data, and as already discussed in Section 1.1, this means a well-defined ontology. One of the most common forms of sharing data is in the form of key performance indicators (KPIs). KPIs is usually a well-defined derived value that can be compared to previous values or other departments or organizations values.

Table 2.2 shows examples of asset management KPIs categorized by domain.

Table 2.2: Asset management key performance indicators

Category	KPI
Costs	Budget compliance
	Unit cost
Production	Overall equipment effectiveness (OEE)
	Availability
	Utilization
	Quality
Work Management	Schedule compliance
	Work orders opened, closed
	Backlog
	Process Compliance
Reliability	Mean time between failure (MTBF)
	Mean time to repair (MTTR)
	Planned vs Unplanned %
	Condition-based maintenance%
Resources	Overtime %
	Callouts
	Stores service levels
	Labour utilization

This section described the various stages of the data life cycle. It also identified some issues that could arise at each stage. The next section discusses data quality in more detail.

2.3.3 Data Quality

The previous sections in this chapter have established that data plays a critical role in asset management decision-making. The quality of the data used, thus, has a direct impact on an organization's ability to achieve its objectives. Having discussed data and its life cycle, this section is dedicated to data quality.

Ever since computers first allowed organizations to collect large data sets, data quality has become a frequent topic of study. Amongst the many definitions given to data quality, two popular themes have emerged: (1) data quality is multidimensional and (2) the quality of data depends on what it is being used for (Gao and Koronios 2010). The second theme, data quality being a function of its use, makes it difficult for researchers to agree on the various dimensions of data quality and their definitions. The result is a multitude of data quality dimensions with little agreement on their definitions.

In a systematic review of the literature, Wand and Wang (1996) encountered 25 dimensions of data quality. Accuracy, reliability, timeliness and relevance were the most popular, despite researchers differing on exactly how it is defined. Wand and Wang identified this lack of a well-defined set of dimensions to be detrimental to designing information systems that deliver high quality data. To address this issue, they defined four data quality dimensions that they deem to be intrinsic to data and, thus, independent of its use. The derivations of these dimensions were based on their data quality model and its four fundamental principles. The principles are formulated as the following four assumptions:

1. **The Representation Assumption:** An information system is a representation of a real world system.
2. **The Interpretation Assumption:** The representation of the real-world system in the design-phase and interpretation of the information system in the use-phase is done by the same user.
3. **The Inference Assumption:** The information system can create a perceptible representation (such as a visual dashboard) from which the user can infer a view of the real-world system.
4. **The Internal View Assumption:** The internal view concerns the details of the construction necessary to satisfy a set of requirements determined by the external view.

The *Interpretation Assumption* excludes issues arising from differences in how people perceive the world (see the discussion on mental models in Section

2.2.2). Interface related issues (implied by the *Inference Assumption*) and issues relating to the external view, such as why the data are needed and how it is used, are also outside the scope of Wand and Wang's data quality model.

Based on their model, Wand and Wang identified three "design deficiencies" and one "operational deficiency". The design deficiencies are summarized below:

- **Incomplete representation:** Lawful states in the real-world system cannot be represented in the information system.
- **Ambiguous representation:** More than one real-world state is mapped to the same information system state.
- **Meaningless states:** There exist lawful states in the information system that cannot be mapped back to the real-life system.

For example, if Scotland did gain independence, all border control information systems would have been incomplete representations for the time it took to add Scotland as a "country of origin" option. Similarly, an ambiguous representation is when an ethnographic study requires participants to fill in their country of birth but does not allow citizen of the United Kingdom of Great Britain and Northern Ireland to specify whether they are from England, Scotland, Northern Ireland or Wales. In this case, four possible real-world states are mapped to a single state in the information system.

Precision problems are another typical example of ambiguous representation: if there are not enough digits to represent a real-world state, it might not be possible to infer which real-world state is represented in the information system. A meaningless state does not prevent an accurate representation of a real-world system, but it does create the potential for meaningless data. For example, listing all countries for teams to pick from (instead of just the common wealth nations) when entering the Glasgow 2014 Common Wealth Games, the information system creates the possibility of a common wealth team from Russia. This renders the entry meaningless.

The operational deficiency is:

- **Garbling:** A real-world state is mapped to either a meaningless or an incorrect state.

By design, Wand and Wang's model does not consider deficiencies resulting from erroneous perceptions of the real-world system or malicious intent. Garbling is thus strictly limited to examples like erroneous data entry or failure to record changes in the real-world system.

Based on the four deficiencies discussed above, Wand and Wang (1996) defined four *intrinsic* data quality dimensions summarized in Table 2.3. These

Table 2.3: Intrinsic data quality metrics

Quality Metric	Description
Ambiguity	Can information system states be mapped backed to only one real-world state?
Completeness	Can all real-world states be represented?
Meaningfulness	Can all information system states be mapped back to meaningful real-world states?
Correctness	Does all information system states map back to the correct real-world state?

four intrinsic data quality metrics provides a useful basis for discussing data quality issues in well-defined, ontologically grounded terms.

Despite this conciseness, it is also useful to look at other, less well-defined, but frequently used data quality metrics. Some researchers, like David Berger (2010), rely on their experience to define data quality issues that they frequently encounter in practice. Others, such as Lee et al. (2002), performed comprehensive literature review. Besides the intrinsic data quality metrics, two other categories of data quality metrics were identified: system and operational quality metrics. System metrics related to the larger context of asset management data and is highly dependent on what the data is being used for. These metrics are summarized in Table 2.4.

Table 2.4: System data quality metrics

Quality Metric	Description
Accessibility	How easy/convenient is it for authorized users to add or retrieve data?
Flexibility	How feasible is it to migrate the data to another system?
Redundancy	Is the data unique?
Timeliness	Is the data current / shared in reasonable time?
Traceability	Can the origin of the data be traced?
Relevance	Is the data useful / relevant to the organization?
Reliability	How reliable is the storage systems?

In contrast, operational data metrics, are directly related to data records

and is independent of the system or processed that is used. It is also less dependent on what the data is used for. Examples of operational data quality metrics from the literature is summarized in Table 2.5.

Table 2.5: Operational data quality metrics

Quality Metric	Description
Accuracy	Does the data accurately reflect reality?
Completeness	Are all the required data filled in?
Consistency	Is the same data recorded each time for the same event or observation?
Credibility	How reasonable/believable, is the data?
Integrity	How big is the risk for corrupting data (either intentionally or accidentally)?
Objectivity	How big is the risk for the data to be biased?
Precision	How precise is the data?
Validity	Does the data conform to business rules, format and standards?
Currency	When was the data last updated, confirmed?

2.4 Information Systems

“It is sometimes said that we now have more and more information that we know less and less about.” – (Madnick and Zhu 2006)

Today, it is impossible to separate data from information systems. The volume of data and rate of generation has become too large to allow manual processing. This is especially true for asset intensive industries where equipment generates very large volumes of data that is only useful when analysed over longer periods. For example, hundreds of data points need to be aggregated every hour just to get a meaningful indication of the production rate of an ore processing plant. The role of information systems are explained in Section 2.4.1.

Despite greatly enhancing an organization’s ability to manage data, information systems, unfortunately, also have their own problems that impact data quality as described in the previous section. Information systems are also notoriously difficult to sustain. Since Chapter 4 includes a software tool as part of

the framework, it is important to be aware of the development methodologies and the common pitfalls preventing the tool from being useful. These topics are discussed in Section 2.4.2.

When developing an information system or attempting to understand a process that relies on information systems, it is also important to look at the underlying technologies that enable them. These technologies are considered in Section 2.4.3.

2.4.1 Information Systems in Asset Management

As discussed in Section 2.2, decision-making is the end-goal of data. A decision signifies the end of a process and the start of an action. Today's competitive, knowledge-based economy requires decision-makers to make more decisions more frequently. With the added abundance of alternative scenarios, decision-makers often suffer from information overload. Information overload is "a gap between the volume of information and the tools we need to assimilate it." (Rochat 2002). To mitigate this issue, asset managers rely on information systems and decision support systems, sometimes collectively referred to as "business intelligence" systems.

Dr Dan Power, founder of the Association for Information Systems Special Interest Group on Decision Support, Knowledge and Data Management Systems (SIG DSS), defines a decision support system (DSS) as "an interactive computer-based system or subsystem intended to help decision-makers use communications technologies, data, documents, knowledge and/or models to identify and solve problems, complete decision process tasks, and make decisions". Power 2007

Another, more recent field that aids decision-making is Business intelligence (BI). BI is a set of theories, methodologies, architectures, and technologies that transform raw data into meaningful and useful information for business purposes (Rud 2009).

The key similarity in these two definitions is "making business decisions". In particular, both concepts are focused on helping to make these decisions in a better and easier way. The other important similarity is they both involve decision-making "based on data" (Kopáčková and Škrobáčková 2006).

In Chapter 1, it was mentioned that despite the abundance of business intelligence systems, no product currently fulfils all the requirements of a modern asset management system. In any asset management system, it is thus common to find many specialized systems that are required by different aspects of an asset management system. These systems may include:

- Supervisory Control and Data Acquisition (SCADA)
- Computerized Maintenance Management System (CMMS)
- Geographic Information Systems (GIS)

- Enterprise Asset Management (EAM)

The broader field of business intelligence also includes (amongst others):

- data warehousing;
- online analytical processing (OLAP);
- relational databases;
- report writing and
- data mining.

Data warehousing is used to aggregate data from various operational systems and make it accessible for multidimensional queries through OLAP. Relational databases store structured data from where reports can pull dynamic values. Data mining is a group of concepts related to “knowledge discovery”. These systems provide the building blocks for an integrated, real-time business intelligence data pipeline. As discussed in Chapter 1, the business intelligence market is well established and highly competitive. But with multiple business intelligence vendors each providing a bigger or lesser (and often overlapping) part of an asset management system, organizations rarely achieve this utopia of interconnected, real-time reporting systems. Instead, expensive, incompatible and under-utilized systems act as source for “copy-and-paste” spreadsheet reports.

In casual conversations with asset managers, the most frequently cited reason for this fall-back to spreadsheets was that the business intelligence systems were “too restrictive”. Since the solution, as will be presented in Chapter 4, also includes an information system component, the next section looks at why so many information systems fail and how to prevent them from being “too restrictive”.

2.4.2 Development Considerations

Software development methodologies play a vital role in the success of business intelligence systems. K. E. Kendall and J. E. Kendall (2014) identify three main system development methodologies that occur in literature: System Development Life Cycle (SDLC), the Agile approach and Object-Oriented Analysis and Design. A brief discussion of each is given below.

The oldest methodology, SDLC, originated in the 1960’s in response to the increased data processing and analysis needs of organizations. The SDLC consists of seven phases as shown in Figure 2.9.

SDLC strongly emphasises careful planning and thorough documentation of the system. Considered a more formal approach, it is generally praised for

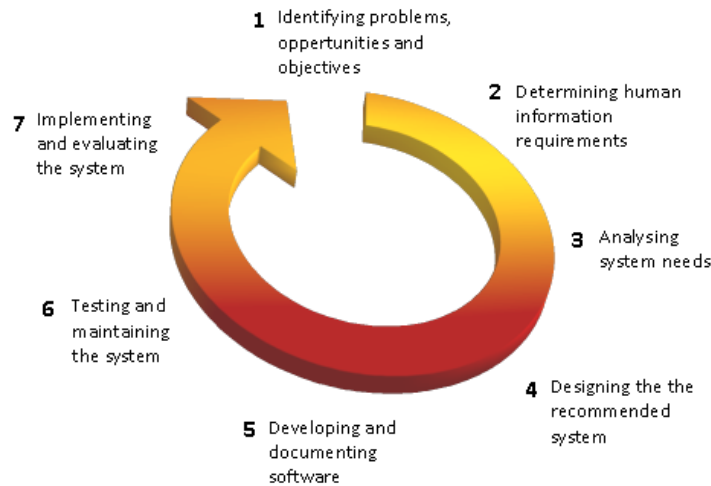


Figure 2.9: The System Development Life Cycle Methodology (K. E. Kendall and J. E. Kendall 2014)

its systematic approach and traditional, high quality products and documentation. However, with the additional time and resource required for planning and designing a system, alternative methodologies more suited for today’s volatile business environments have emerged.

The Agile Alliance, a group of 17 developers, published their Agile Manifesto in 2001. Today, the Agile approach is a broad group of commonly used methodologies that still incorporate the specific values, principles and core practices set out in the Agile manifesto. The twelve Agile principles emphasize: customer priority, frequent (working) software releases, cross-functional collaboration and face-to-face communication, simplicity, self-organizing teams, trust and embracing of change.

Agile methodologies are thus preferred if the system needs to be developed quickly in response to a dynamic environment. However, strict adherence to the Agile principles often require an Agile champion to be present. Agile methodologies, due to their encouragement of change and frequent releases, have also been criticized for lack of proper documentation (K. E. Kendall and J. E. Kendall 2014).

The third common approach is Object-Oriented Analysis and Design (OOAD). Like the Agile methodologies, it broke away from the structured approach of SDLC and emerged from the need for developing continuously evolving systems. OOAD, however, approaches the problem of change very differently from the Agile methodologies. Instead of reacting to change when it occurs, the ability to adapt to change is incorporated in the initial design phase. To do this, OOAD introduces a new way of modelling a system: it reduces a complex system into objects (which can be events or things) and groups them into classes that define a set of shared attributes and behaviours. These “building blocks” are described by using the industry standard unified modelling lan-

guage (UML) and allow systems designed with this methodology to be quickly adapted. This is because (at a minimum) only one “block” needs to be updated, instead of the whole system (K. E. Kendall and J. E. Kendall 2014).

The three formal methodologies presented above provide a good foundation from which to start, but Heeks (2002) argues that these methodologies were mostly developed by Western, industrialized countries and might not be appropriate for all contexts. Common pitfalls and design recommendations for preventing information system failure are discussed below.

There are countless examples of failed information systems. Most recently, the State of Oregon has taken Oracle (one of the big five business intelligence vendors mentioned earlier) to court for the failed Obamacare website, healthcare.gov. Wall Street Journal (2014), reports that over a period of three years, Oracle billed the state \$240,280,008 in “false claims” and failed to produce a working site in September 2013. Allegedly, the “Oracle Solution’ was not flexible, was not integrated, and most importantly, did not work out-of-the-box” (Wilhelm 2014).

Heeks (2002) provides evidence that many information systems can be categorized as having either failed partially or completely. He explains that an information system is likely to fail if the “design-actuality gap” (the difference between the current system and the future system as represented by the design) becomes too large.

Several issues contribute to the existence and extent of this gap, first of which is simply that people’s view of the future is subjective. The second contributor to the gap between design and actuality is the lack of objectivity in the design process. The designer’s cultural values and ideals are often incorporated into the design. Heeks also notes that assumptions about the user’s “activities, skills, culture and objectives” as well as the organization’s structure and infrastructure constrain the design to a specific context, one which cannot hold indefinitely.

Heeks presents a third contributor: the rationalist mind-set with which most developers (educated in Western design methodologies) approach system design. When this mind-set is applied in organizations where employees time is mostly spent “fighting fires”, the gap between the “hard” rational design (with emphasis on formal, standardized procedures and quantitative information) and the prevalent “soft” political environment (with emphasis on flexible and informal procedures and qualitative information) becomes another possible cause for failure. Unfortunately, organizations far too commonly attempt to “fix” this “soft” actuality with information systems that enforce more formal procedures and data standards.

A report by Melissa Jonson-Reid and her colleagues (2001), is an example of how the design-actuality gap, if not addressed, can lead to failure. They documented and analysed the design and implementation of an information system for social workers in a school environment. They highlighted the issue of computer illiterate users who struggled with the interface and hardware

and software incompatibilities, but both these problems were overcome by training users and buying new hardware. Their biggest obstacle was identified after a pilot semester: they realized that they had hard-coded a lot of legacy procedures and data requirements. These requirements were simply ignored when filling out forms by hand, but with the new information system, the social workers had to spend additional time recording unnecessary data. This gap between design and actuality caused some social workers to revert to the old system of hand-written records.

Fortunately, Heeks (2002) also proposes a solution for the design-actuality gap: local improvisations. Local improvisations happen when the users, after implementation, creatively adapt a system to either keep it relevant to the changing environment or extend its functionality beyond what it was originally designed for. The extent of these improvisations is largely limited by the design itself. On the one end, there is design-imposing applications and on the other end, actuality-supporting applications. Design imposing applications restrict improvisations by hard coding, and thus requiring, specific processes and usage patterns. At the other end of the spectrum are actuality supporting applications. These applications have relatively few design assumptions and support local improvisations. An example of this type of design is a spreadsheet application, which can be used for multiple purposes. For example, some people use spreadsheets only as a calculator, while others use it to build elaborate database-like systems. Neither of these use cases comply with the original purpose of spreadsheets.

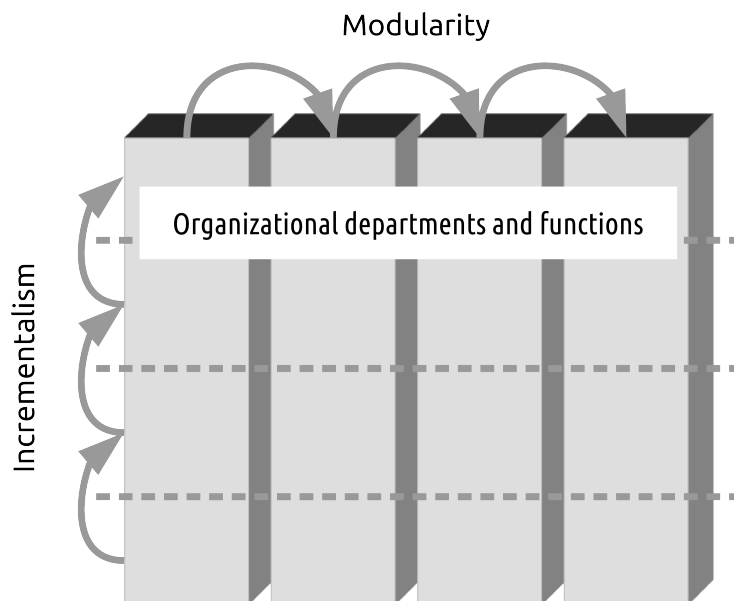


Figure 2.10: The two aspects of divisibility: incrementalism and modularity (Heeks, 2002)

To facilitate local improvisations and improve the general sustainability of an information system, Heeks (2002) encourages “divisibility”. The two aspects of divisibility are modularity and incrementalism (see Figure 2.10). Modularity, on the other hand, allows users to learn the system gradually and adapt it for their context. Google’s services are a good example of modularity. Users, for example, can choose to only use Google’s email service as each service is functional on its own. As users become more comfortable with Google’s services, they can start using more services from Google, all of which form part of the same suite. Incrementalism, on the other hand, allows a system to be functional even if only a part of the organization uses it or if the data is incomplete. An example of this is Wikipedia or OpenStreetMaps, which respectively allows users to incrementally build an entire encyclopaedia or global street map.

The discussion above supports the idea that information system, especially in organizations with immature asset management systems must be *actuality supporting* as opposed to *design imposing*.

The next section discusses the technology aspect of information systems.

2.4.3 Technology

Traditionally, an information system was as much a physical hardware problem as it was a design problem. Large organizations often had to invest in additional infrastructure to house and support servers with business critical applications and data. Servers had to have a reliable power source, humidity and temperature had to be regulated and backups were stored on physical containers that had to be locked in safe. Typically, only one or two people in the organization had the privilege of connecting to these systems.

Today, computers of all shapes and sizes can be found everywhere. Desktops, laptops, tablets and smart-phones are everywhere and everything is connected. Virtual servers allow organizations to deploy services in minutes and large business intelligence vendors even allow instant deployment and “pay-per-minute” services from the “cloud”.

Yet, in large asset-intensive organizations, these new devices and services often coexist with remnants of the old systems described above. Any project undertaken in such an environment (that relies on data in any form), must thus evaluate the consequences of the technology they choose to use.

When choosing a technology, consideration must be given to its:

- cost
- compatibility
- license
- support

- community
- track record of company

Cost is an obvious consideration. Upfront investments versus monthly fees must be carefully considered.

With so many devices, old and new, the compatibility of any technology is a serious consideration. An iPad, for example, might be an attractive and affordable technology to gather data in the field, but it is not compatible with old Microsoft Exchange still used by many organizations for e-mail, contact and calendar synchronizing.

With 80% of servers and 60% of mobile phones running on open source software, licensing has become a hotly debated topic. Open source software claims better security, faster bug fixing and better sustainability amongst a long list of benefits. With the recent focus on privacy and security, open source technology vendors emphasize that their products “puts you in control of your data”. As previously mentioned, vendor lock-in is a serious problem. With so much hardware being virtualized and provided as a service, licensing has become an important consideration.

Support is an often overlooked, but important aspect of choosing technologies. Both paid support from the vendor and community support are important. A technology with an active user base means more people are familiar with it and can use or support the technology.

Finally, in the age of start-ups, it is important to look at the track record of a company. Even large companies like Google is famous for sudden discontinuation of products that some organizations relied on for their business.

2.5 Frameworks and Reference Models

The data quality issues discussed in Chapter 1 and Chapter 2 have their roots in many disciplines. It can be as much a social or corporate issue as it can be an engineering problem. Even from an engineering perspective many approaches exist.

For this study, the research objective in Section 1.2 states that a “framework” for identifying and assessing data quality issues in asset management must be developed. This objective was directly derived from the research problem. How the research problem was defined was largely influenced by the context where the problem was first identified (see Section 5.1), the literature review presented in this chapter and the philosophical world view and research approach discussed in the next chapter. This section defines what a framework and its related concepts are.

Modelling is a key part of understanding and solving problems and is at the core of any scientific or engineering activity. A *model* is simply an abstraction of a system to represent reality.

A *reference model* is a generalized abstraction for understanding entities, their attributes and the relations among those entities in some problem domain. A reference model is not directly tied to any standards, technologies or other concrete implementation details. Its purpose is to provide a common ontology (“semantics”) that can be used unambiguously across different implementations (OASIS, 2014). There are four important concepts that relate to a reference model. Reference models are abstract, describe entities types and their relationships, are restricted to a specific environment and are technology agnostic.

In modelling, a framework is a container artefact for collecting procedures, methods tools and reference models. Where a reference model helps to understand a problem, a framework accompanies it with procedures, methods and tools to solve the problem. For instance, a reference model might describe how windows, doors and walls (entities) relate to each other in a reference model of a house. A framework will incorporate tools and methods such as provided by drafting software to create a specific model of a house and possibly a procedure for getting the house approved and building it.

2.6 Summary

The literature review presented in this chapter provides a holistic view of the problem identified in Chapter 1. Section 2.1 began with the historical context of asset management and concluded with what ISO 55000 defines as asset management today: a risk based information driven planning and *decision-making* process. Asset management converts organizational objectives into asset related decisions. Section 2.1 not only highlighted the importance of data in asset management, but also the *management* of data.

From the problems observed in practice, and from asset management’s emphasis on data, it was clear that data played an important role in decision-making. Section 2.2 investigates this role by discussing several decision-making models. Besides defining what a decision is and comparing various decision-making models, it was also discovered that all the models had one thing in common: the decision-making process starts with a “status update” from the system. More complex models introduced the concept of “mental models” and the process of using “causal data” to update the decision-makers mental model. This means that data plays to very different, but equally important roles in decision-making.

Knowing how data influences decision-making and where it fits into asset management, Section 2.2 presented an anatomy of data quality. It was shown that data does not have any meaning in itself and, thus, has no value. Only upon progressing through the data life cycle does it gain meaning and value until ultimately informing action. Here, it was discovered that each stage of the data life cycle adds a myriad of possible data quality issues.

Another discovery from Section 2.3 was that information systems play a crucial part of nearly every stage of the data life cycle. Section 2.4 thus presented an overview of the prevalence and role of information systems in asset management (first noted in Section 1.1). Section 2.4 also argues that the high rate of failure in information systems is likely due to the “design-actuality” gap. “Local improvisations” and application “divisibility” were recommended to limit the extent of the design-actuality gap. This knowledge, as well as the discussion on technological considerations that followed, is also used in Chapter 4 when designing a tool.

Finally, Section 2.5 defined what reference models are and how they fit into a framework. As mentioned in the previous section, the environment in which the problem was identified (Chapter 1) and the literature review in this chapter are only two of the three main influences that lead to the research objective formulated in Section 1.3. The third primary influence is the methodology and its underlying philosophical world view as defined in the next chapter.

Chapter 3

Research Methodology

“Empirical research begins in the field of practice and requires a certain amount of pragmatism. However, this pragmatism must be philosophically informed.” – Mesel (2013)

The research methodology is the strategy that determines how a research study is undertaken. Creswell (2013) explains the interconnection of three components of a research methodology: the research approach, the philosophical world view and the specific research methods.

Traditionally, there were two approaches to scientific enquiries: quantitative research (adopted by the positivist world view) and qualitative research (adopted by the interpretivist world view). Bryman (2006), however, argues that this dichotomy has largely been replaced by mixed methods research stemming from a pragmatic viewpoint. This pragmatic viewpoint means that the use of any approach that allows the research question to be answered, is acceptable, and that no single method is preferred over another. For the pragmatists, the purpose of science is not to find truth or reality. Truth and reality are concepts whose existence is perpetually in dispute. The purpose of a pragmatic approach is thus rather to focus on the facilitation of human problem-solving (Powell 2001).

Johnson and Onwuegbuzie (2004) argue that the “paradigm wars” (the long running dispute between advocates of quantitative and qualitative research approaches), have lost relevance in today’s era of mixed methods research. However, Mesel (2013) maintains that, especially now, “transparency on the philosophical level is important for validity and consistency as well as for attempts to integrate or establish an interface to other research”.

This chapter describes the philosophical assumptions (world view) underpinning this study (Section 3.1) and the particular methods employed (Section 3.2) to develop a research methodology (Section 3.3). The ethical and legal considerations of the proposed methodology are also discussed (Section 3.4).

3.1 Philosophical World View

The philosophical world view is influenced by the field of study and the researcher's experience and personal beliefs. The world view holds important implications for which research methods will be used, how the problem is formulated, how value is measured and how the study is validated.

Previous studies addressing data quality issues in asset management decision-making were mostly grounded in an interpretivist world view and thus followed a qualitative approach with no practical component. For example, in "A Data Quality Model for Asset Management in Engineering Organizations", Lin et al. (2006) identified the same problem as the one identified in this study, namely, the lack of quality data available for asset management decision-making. However, they addressed this very similar problem through a qualitative approach rooted in an interpretivist world view with an interview-based case study as their primary research method.

Contrary to previous studies done on a similar topic, this study will follow a pragmatic world view. Pragmatists are in agreement that research never occurs in isolation, but that it is always situated within a social, historical, political and legal context. As was noted in Chapter 2.4.2, it is crucial that solutions are flexible and that a system should not be hard-coded. It must be open to improvisations that fit in with the context in which it is applied. The implementation of a solution thus clearly, and to a large extent, relies on the world view underlying it. In this way, pragmatic studies are able to provide localized solutions that reflect the social, economic and political context of an organization.

Philosophical assumptions affect three domains that are of concern to research approaches, namely ontology, epistemology and axiology.

Ontology: Ontology, a branch of metaphysics, mainly deals with questions concerning which entities exist or can be said to exist. It thus looks at how entities are divided into categories based on their similarities and differences. Pragmatists see many realities that are ever changing and believe observations can only be made through affecting change in the system and measuring the outcome.

Epistemology: Epistemology is the theory of truth or knowledge. It poses the question: what is true, and how do we come to know that truth? Epistemology also specifies what constitutes appropriate knowledge in the field, where it is and how it can be represented and transferred.

Axiology: Axiology is the theory of value or worth. It thus defines a value system in the field. It dictates how a study is validated and how its value is measured.

For this study, the implication of following a pragmatic approach is that the question of which entities exist and how they relate is determined by bringing about change in a real world system. A pragmatic approach also implies that the value of the solution proposed in this study will primarily be assessed against its usefulness.

This section has shown the application of the pragmatic world view on the three main areas of philosophy. This means that determining how entities should be grouped in a solution, whether the solution is true and whether it has value, will be determined through a pragmatic lens. Solutions will be judged on their usefulness in solving a specific problem and not whether it has complied with a specific philosophical theory. This is a consideration that will be specifically important in Chapter 6 when the success of the study to achieve the stated aims will be reflected on.

Having established the philosophical world view and related fields, it is now necessary to look at the research method that is followed in this study.

3.2 Research Methods

The research objective of this study is to develop a framework. To ensure that the development process produces a framework that is both relevant and innovative, two research methods were adopted: a comprehensive literature review of all the relevant contexts for this study, and a case study where the framework can be iteratively designed, developed and evaluated. The field of information systems research has long dealt with the design of "knowledge artefacts" such as software programs, database models or technical frameworks. The remainder of this section will thus describe a research method that originated from information system researchers' need for more rigid and reproducible studies: design science research. This discussion intends to highlight some of terminology and aspects of design science research that are incorporated into the research design for this study.

Design science research involves the design of novel or innovative "artefacts". Artefacts include, among others, models, frameworks and tools and are knowledge containing entities. Design science research analyses the performance of such artefacts to achieve an understanding of a problem. The creation and evaluation of artefacts form a crucial part of the design science research process (A. R. Hevner et al. 2004). Design science research is thus "learning through the act of building".

Other examples of artefacts that the design science research process applies to include, methods, constructs, instantiations, explanatory theories and social innovations. It further applies to both new and previously unknown properties of resources which can be of a technical or informational nature.

A. R. Hevner et al. (2004) provide a conceptual framework for design science research, consisting of seven guidelines which help one understand, exe-

cut and evaluate the research.

1. **Design as an Artefact** - Design science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation.
2. **Problem Relevance** - The objective of design science research is to develop technology-based solutions to important and relevant business problems.
3. **Design Evaluation** - The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.
4. **Research contributions** - Effective design science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies.
5. **Research rigour** - Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact.
6. **Design as a Search Process** - The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
7. **Communication of Research** - Design science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

In a later article, A. Hevner (2007) provides another way of understanding design science research. He refers to this as the design science research cycles. This model consists of three closely related and inherent cycles of activity (A. Hevner 2007).

The context and requirements for the research is established in the **relevance cycle**. The acceptance criteria (“research objective”) against which the research outcome will be evaluated is also defined in this cycle. It provides the requirements and does the field testing. Its applicable fields are people, organizational systems, technical systems and problems and opportunities (A. Hevner 2007). The **rigour cycle** provides grounding and a foundation. It looks at past experience and expertise as well as scientific theories and methods (thus the whole of the knowledge base) to ensure that there is constant innovation in the way it is applied. The **design cycle** lies at the heart of design science research. The design cycle is an iterative process of developing and evaluating artefacts. The three cycles are depicted in Figure 3.1.

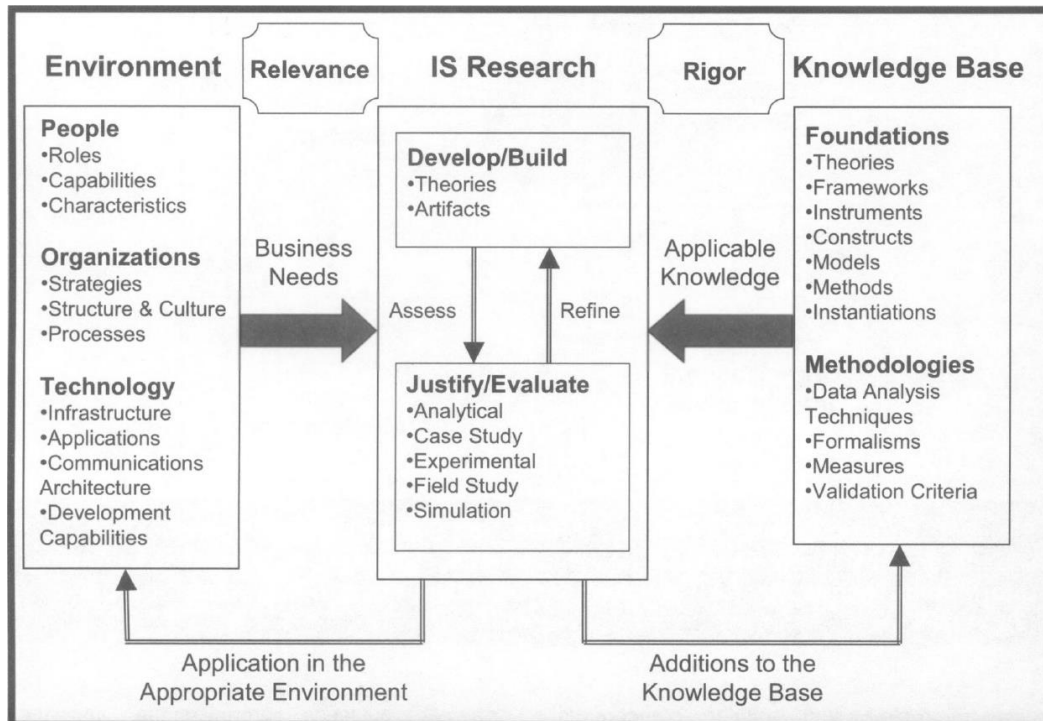


Figure 3.1: Design Science Research Method

The research design for this study shares many characteristics with design science research. The next section, thus, only briefly explains the research design for this study.

3.3 Research Design

The first objective of this study was to identify a problem in practice. This means that the case study is part of the research design from the very start of the study. An initial literature review is then used to provide context to the identified problem and help motivate it.

Once an initial understanding of the problem is achieved, the main research activity becomes an iterative cycle of designing, developing and evaluating a framework for identifying and assessing data quality issues in asset management decision making. Parallel to this design cycle, the literature review and the case study is continued. This means that the research problem, as well as the solution objectives, evolves with the understanding of the problem. The final activity is to reflect on the study and communicate the value from both a scientific and practical perspective. Section 1.5 described how these research activities are documented in this thesis.

The next section expands on the ethical and legal considerations of this

research design.

3.4 Ethical and Legal Considerations

The pragmatic world view requires that research be tested and proven in practice. Therefore, this study not only attempts to describe or explain the world, but also shape it. Ethical responsibility and awareness of relevant laws and regulations are thus an essential part of any study conducted from a pragmatic world view. This section discusses the ethical and legal considerations for this study.

In research, Simmons (2009) argues that the fundamental principle of ethics is to “do no harm”. How this principle is applied in practice, however, is not always clear. Adding to the complexity of ethical issues is that they are tied to the socio-political environment of the study and often arise as a result of conflicting ethical principles. In case study research, these issues often arise in the field. Simmons (2009) argues that possible ethical issues must be considered before engaging in a case study. These considerations include questions of data ownership, data access, obligation of the researcher and purpose of the research.

Design science research implies an ethical change from describing the existing world to shaping it. Even when following this method, one can question the values being followed. For example, one can ask whose values are given preference and whether the dominant world view is being forced onto a situation where it will lead to failure of the project. Research must thus be careful in that it must not only focus on the interests of a dominant group, but must take all interests into account. It is therefore important that when changes are made in an organization, one is aware of the ethical implications it may entail and be wary not to benefit certain groups above others.

Botswana (the country where the case study was undertaken) does not currently have any laws related to personal privacy, but they are in the process of developing such laws (Keetshabe 2012). These laws will be based on the organization for Economic Co-operation and Development (OECD) guidelines governing the protection of privacy and trans border flows of personal data (OECD 2013). This document is a widely used standard which also informed the South African privacy legislation. The four main principles contained in the OECD guideline which are relevant to this study are the following:

Collection Limitation Principle: There should be limits to the collection of personal data [defined as data about an identifiable individual] and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

Data Quality Principle: Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for these purposes, should be accurate, complete and kept up-to-date.

Security Safeguard principle: Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure.

Accountability principle: A data controller should be accountable for complying with measures which give effect to the principles stated above.

These guidelines are similar to the laws related to personal data in South Africa. South Africa's Protection of Personal Information Act 4 of 2013 requires that several principles be followed when collecting personal information. In short, data:

- Section 20: may only be collected and processed by the operator with the knowledge or authorization of the subject;
- Section 27: may only be collected for a purpose related to a function of the responsible party;
- Section 16: must be complete, accurate, not misleading and up to date;
- Section 19: must be safeguarded against unlawful access or processing of information and
- Section 23: must be accessible to the subject on request.

The act requires that the responsible party is accountable to ensure that the above-mentioned central principles are complied with. When recording or acquiring information it is thus necessary that due regard be given to whether the specific information is in fact essential.

It is also crucial to give consideration to how data will be secured. The growing ISO 27000 series of international standards provides guidelines for an information security management systems. Unfortunately, legislation will always lag behind technological development. This is also the case with international standards. It is in this regard that ethics play a central role in that ethical considerations govern the "grey area" between new technological innovations and constantly outdated legislation. All people involved in the processes of collection, processing and storage of personal information have to be aware of the importance of privacy and confidentiality.

Based on the discussion in this section, an attempt has been made to ensure that no personal or sensitive data is made public.

Chapter 4

Framework for Identifying and Assessing Data Quality Issues

The research method for this study (described in the previous chapter) is based on an iterative design process supported by a literature review and a case study. Even though it was clear from the start that data quality was a real-world problem, it took several iterations of the design, develop and evaluation process before the scope, research problem and solution objectives communicated in this document were formulated. This chapter describes how the solution objectives were derived and documents the final design of the framework that satisfies these requirements. References to previous iterations of the design are made only where relevant.

Section 4.1 gives an overview of how the solution objectives were derived, Section 4.2 stipulates the solution objectives and Section 4.3 presents the framework that was developed to satisfy the research objective.

4.1 Introduction

The problem identified in this study deals with the inadequate data available for asset management decision-making and was identified and explored not only in literature, but also in practice at a diamond mine processing plant in Botswana. The literature review was documented in Chapter 2, and the case study in Chapter 5. This section provides an overview of the thought process involved with translating the context of the problem into some of the solution objectives described in Section 1.3.

As discussed in Chapter 1, the problem of inadequate data for decision-making was found to be affecting many organizations around the world. It was also argued that the asset-intensive industry is especially vulnerable to this problem. Four primary fields of study were found to be relevant to this problem: asset management, decision-making, data quality and information systems. These fields and their related concepts were explored in Chapter 2.

This initial identification and exploration phase is depicted in Figure 4.1.

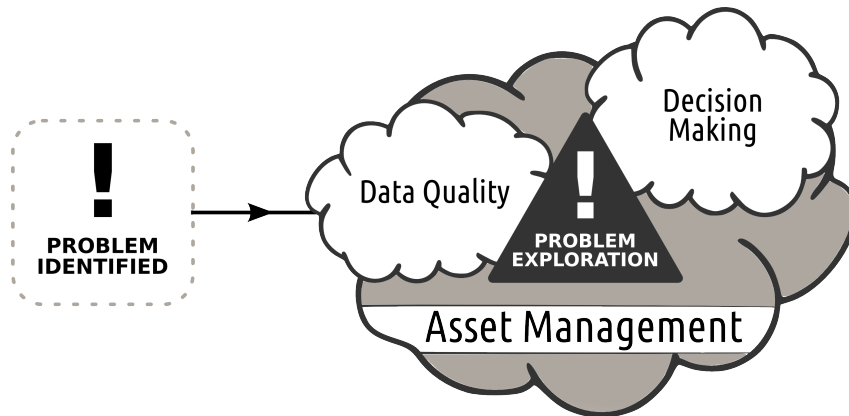


Figure 4.1: The identification and exploration of the problem.

With the publication of the ISO 55000 series of international standards in February 2014, organizations can finally have a standardized asset management system. In the ISO 55000 series, asset management is described as a “risk-based, information-driven, planning and decision-making processes that transform organizational objectives into asset management plans and actions”. This definition of asset management can be modelled as an input (“organizational objectives”) and an output (“asset-related decisions, plans and actions”) yielded from a “risk-based, information driven, planning and decision-making process”. This functional (or “black box”) process model depicted in 4.2 succinctly illustrates the context for this study.



Figure 4.2: Asset management as a teleological “black box” model

The conceptual model of asset management described in Section 2.1.3 breaks open the asset management “black box” from Figure 4.2 by identifying six subject groups and their relations as previously discussed in Section 2.1.3. The core of the problem lies within the “asset knowledge enablers” and its enablement of “asset management decision-making” as depicted in Figure 4.3.

As is the case for data in many other organizations, “asset knowledge enablers” are simply not of sufficient quality to allow confident decision-making.

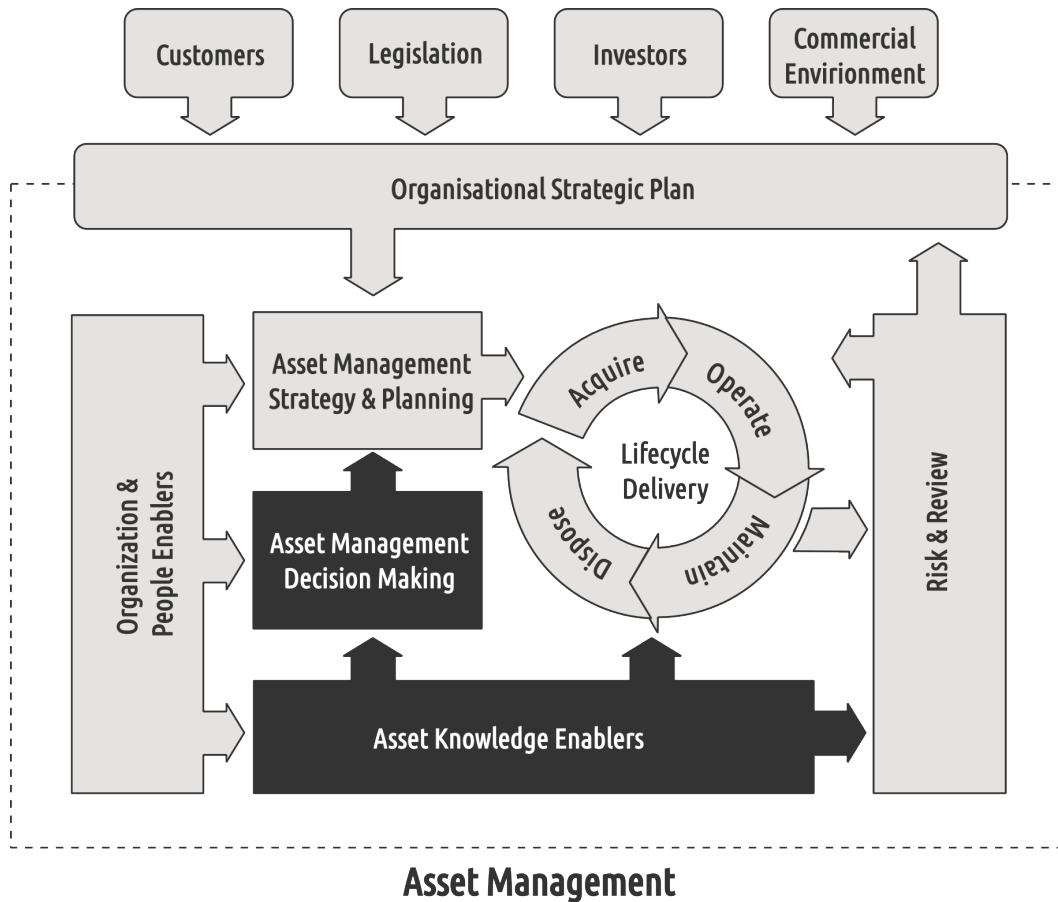


Figure 4.3: Asset management as an ontological “clear box” model with the problem scope in black

Initial iterations of the framework thus attempted to solve data quality issues in its entirety. However, one of the first issues that were encountered during the case study was that departments used different definitions for the same key performance indicator (KPI). When discussing how this problem could be solved, two things became apparent. Firstly, changing KPIs have direct political and financial consequences for both the organization and individuals since employee bonuses and departmental penalties are often directly linked to their KPIs. The second discovery was that decision-makers were either not aware of issues like the inconsistently defined KPIs, or they did not know how to systematically identify these issues. Since attempting to address an issue that impacts something as sensitive as an employee’s bonus introduces several ethical considerations and requires a thorough understanding of the political and financial consequences (which the limited time of the case

study did not allow for), it was decided to focus the scope of this study to address the problem of asset management decision-makers not being familiar with their data pipeline or having a structured approach for identifying and assessing data quality issues.

Having narrowed the scope of the study to a manageable context, the research problem (repeated below for convenience) could be formulated as per Section 1.2):

“No practical framework exists for identifying and assessing data quality issues in asset management decision-making.”

To derive the solution objectives that will satisfy the research objective of this study, the asset knowledge enablers subject group was analysed and modelled as a black box process to aid the identification of solution objectives. In this process model, the input is “data” and the output is “reports”. Figure 4.4 depicts the data of uncertain quality that goes into the black box which organizations rarely understand. It is not possible to make judgements based purely on the outputs of the process (in this case, reports), but following the old adage of “garbage in, garbage out” it can be ascertained that if the quality of data is low, the quality of the reports will be low as well.

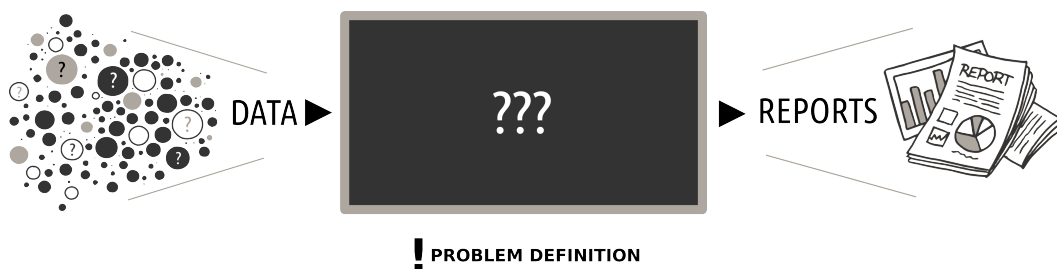


Figure 4.4: Teleological “black-box” data pipeline model indicating uncertainty in both the input and the process.

This means that in order for asset managers to know the quality of their reports (which is an indication of their decision-making capabilities), they must understand all aspects of their organizations’ data pipeline. Components of this “data pipeline” (turning data into reports) have been detailed in the “Data Life Cycle” section (Section 2.3).

The first need of a “structured approach” to understanding an organizations’ data pipeline is a reference model that describes its elements (such as the data sources, storage systems, KPIs and reports) and their relationships. For this reference model to be of practical use, it is further necessary to develop a tool (an information system) that will help asset managers populate their organizations data pipeline model.

The next section continues this thought process and draws from both the literature study and the case study to derive the solution objectives.

4.2 Solution Objectives

From the previous section and the discussion in Chapter 1, it is clear that data quality issues in asset management decision-making can not simply be solved by yet another standard or yet another business intelligence product. Unfortunately, due to the many incompatible systems and standards required for different aspects of an organization's asset management system, it is still necessary to manage data from different systems to allow meaningful reporting and decision-making. As discussed, this process is not well understood by many organizations and the quality of their data is thus not sufficient to inspire confident decision-making. The proposed solution to this problem is therefore a framework to make it easy for asset managers to model their data pipeline and identify issues that affect the quality of their data. The previous section concluded with the three components required of such a framework: a data pipeline reference model, a software tool to make it easy for asset managers to use and guidelines for how to model their data pipeline.

When developing a model, a balance must be maintained between simplicity and ensuring that it can adequately represent the problem domain to facilitate problem solving. At a minimum, the reference model should thus cover the data life cycle and data quality as presented in Section 2.3. However, as discussed in Section 2.4, representing reality in practice often requires "local improvisations". To be successful, the design of the reference model should not prohibit this when implemented. The reference model should also be aligned to the requirements of Section 2.5, in particular the "context specific" requirement. The model should be specifically for the context of asset management decision-making as discussed in Section 2.1 and 2.2.

The purpose of the second component, the tool, is to make it easy for asset managers to build their organization's data pipeline as described by the reference model. In practice, it was observed that asset managers prefer Microsoft Excel to capture data, and if they do not feel comfortable with another tool, they will quickly revert to using Excel spreadsheets. The use of a spreadsheet application, however, is often at the expense of the quality of the data, especially if more than one person is given access to the file. Any tool to help build an organization's data pipeline should thus be as accessible and easy to use as a spreadsheet application, but additionally provide the necessary authentication and authorization for multiple users to use it without compromising the security or quality of the data. The discussion in Section 2.4 identified additional considerations when developing an information system.

The solution objectives are summarized below:

- Develop a reference model that allows organizations to describe their asset management decision-making data pipeline. This reference model should:

- cover the data life cycle from the acquisition stage to the sharing stage;
 - be specific to asset management decision-making;
 - be extensible to accommodate local improvisations; and
 - should not depend on specific technologies or standards.
- A software tool should accommodate the reference model to assist in the modelling of an organization’s data pipeline. This tool should:
 - be easy to deploy in any environment;
 - be user friendly;
 - act as a reference implementation of the model;
 - make it as easy as possible to integrate with other systems;
 - be secure; and
 - should not compromise the quality of the data stored.
 - Provide a description of the process to identify and assess data quality issues to act as guidance for applying the framework.

The next section presents the solution.

4.3 The Framework

This section is divided into three parts, one for each of the three main solution objectives. Section 4.3.1 describes the reference model that allows organizations to describe their asset management decision-making data pipeline, Section 4.3.2 describes the tool built on top of this reference model, and finally, the process and guidelines are presented in Section 4.3.3.

4.3.1 The Model

The first solution objective requires that the framework should allow asset managers to represent their data pipeline in a standards and technology agnostic manner. The reason for this is that, despite many initiatives, no universal data model or system (in either scope or adoption) for asset management exists (see discussion in Chapter 1). The result is that many existing data models that in some way relate to the data pipeline in asset management have to coexist. Examples of these models include:

- A proprietary data model used by an information system for storing PLC data licensed from a commercial business intelligence vendor.

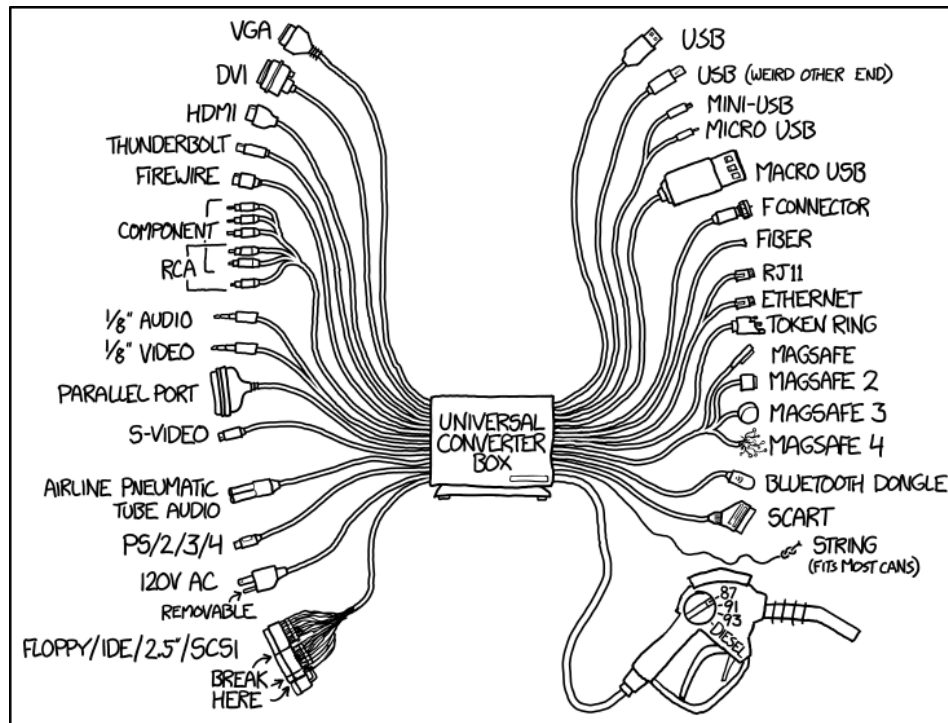


Figure 4.5: The problem of too many systems and standards (Randall, 2014)

- An in-house developed, ISO 14224 compliant information system for storing a petrochemical plant’s asset hierarchy.
- A spreadsheet based information system with no formal data model for storing records of safety incidents.

With so many systems and standards that provide different components of an asset management system, system integration has become a crucial part of asset management. Despite software vendors selling integration suites (“converter boxes”), it is obvious that a “universal” converter box is not a feasible solution. Figure 4.5 light-heartedly illustrates the absurdity of a universal converter box in another field that also suffers from too many competing suppliers and standards.

The purpose of this model is to allow asset managers to model their data pipeline. Since no previous studies were found to address this specific design problem, the model was developed based on general modelling best practice, the data life cycle and data quality discussion in Section 2.3, and practical experience from the case study (Chapter 5).

This section first provides an overview of the core data pipeline model and then discusses each entity and its attributes and relationships. The data pipeline reference model is presented as an ERD diagram with crow’s foot notation in Figure 4.6.

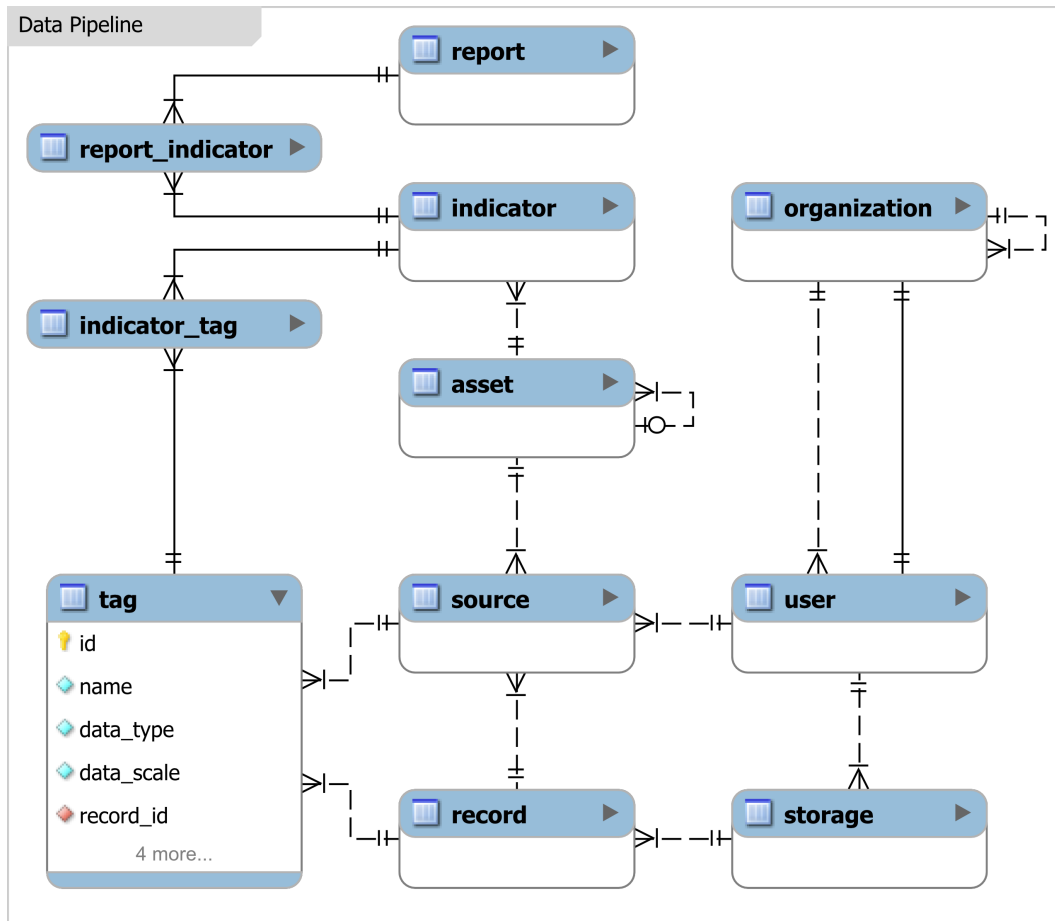


Figure 4.6: Data pipeline model overview

In the early phases of designing and testing the model, it was quickly discovered that a bottom up approach was not feasible: there are just too much data being generated in a modern plant. Another issue was that engineers and managers tended to list what *can* be measured, instead of what *is* measured, resulting in even more data that was difficult to organize. It was thus clear that a structured top down approach was required. Section 2.2 concluded that a decision is the end goal of data. The assumption made in this model is that a “report” is both the catalyst for action (a decision) and the means through which decision-makers update their mental model. It is important to note that selecting the “report” as the starting point for the top down approach means issues such as reports not being aligned with the organisation’s asset management strategy or whether the organisations asset management strategy is aligned with the organisation’s vision will not be directly identifiable. This decision was made to keep the scope of this study from becoming too large.

The top down data pipeline model in Figure 4.6 starts with the report entity at the top. A report shares a many-to-many relationship with the indicator

entity. An indicator, in turn, shares a many-to-many relationship with the tag entity. These relationships are expressed through the “report indicator” and the “indicator tag” junction entities respectively. A tag has a non-identifying, optional many-to-one relationship with the source entity as well as the record entity.

The important entities, relationships and attributes will now be discussed. The complete ERD for the data pipeline reference model can be found in Appendix A.

Tag: Tags (also referred to as key-value or attribute-value pairs) are the building blocks of this model. Ideally, it should represent raw unprocessed data as found at the bottom of the DIKW hierarchy (see Figure 2.7). Raw data, however, is relative: a value aggregated by one department might become a “raw input” for another department. A tag is described by three primary attributes, namely data scale, data type and name.

The **data scale** attribute defines ten high level categories which gives an indication of how the data can be interpreted or used. This attribute is inspired by S. Stevens’ (1946) “Theory of Scales of Measure” and adds additional “scales” for some special cases encountered in practice. The list of data scales that can be assigned to a tag and their explanation is given in Table 4.1.

S. Stevens’ (1946) originally defined 4 “levels of measure”: nominal, ordinal interval scale and ratio scale. In Table 4.1, discrete, continuous, percentage, ratio and currency are all instances of ratio scale. The reason for this division is to make it easier to define quality metrics for these scales. For instance, unlike discrete data, continuous data is susceptible to *precision* errors. Another example is currency data that requires additional data such as the exchange rate or inflation rate at the time the data was recorded to be *useful*.

The second tag attribute is the **data type** attribute. This attribute is based on data types as commonly classified in programming languages or relational databases. Some common data types are: string, integer, float, date, date-time and boolean. Some specialist databases such as geospatial information systems (GIS) also define compound data types such as a coordinate. A coordinate can be represented as a tuple of two floats where the two values represent the longitude and latitude of a location respectively. The data type indicates the database column type or how the value is serialized or internally represented in a system. This attribute can be used to determine how much processing is required before a value can be used.

The **name** attribute simply gives the tag a human friendly description. In addition to these primary attributes, tags also have optional or attribute dependent attributes which include “unit”, “format” and “options”. For discrete, continuous, ratio and currency data scales, the **unit** attribute provides additional information to ensure compatibility. Likewise, due to localization, the **format** attribute is crucial for tags with data types like date, date-time

Table 4.1: Tag data scales and their description with examples in italics

Data Scale	Description
Text	Text signifies unstructured text. <i>Delay description: Conveyor slipping due to wet ore.</i>
Nominal	Categorical data with no inherent order. <i>Continent: Africa/Asia/Europe/...</i>
Dichotomous	Special case of nominal scale where there are only two values. <i>Gender: Male/Female</i>
Ordinal	Categorical data with order. <i>Priority: Low/Medium/High</i>
Interval	Values that do not have a fixed zero point. <i>Purchase Date: 2014-06-14</i>
Discrete	Rational, discrete numbers. <i>Near Hit Incidents: 3</i>
Continuous	Rational, continuous numbers. <i>Elapsed Time: 154.37s</i>
Percentage	Rational number expressed as percentage. <i>Engineering Availability: $\frac{AT-DT}{AT} = 72\%$</i>
Ratio	Rational number expressed as ratio. <i>Plant Feed Mass Flow: 1340 tons/hour</i>
Currency	Rational number expressed as currency. <i>Repair Cost: 120 Pula</i>

and float. In addition to these attributes, a fifth attribute, **options**, is used as a “catch-all” attribute to provide some flexibility to the model. Options, for example, can be used to store the list of legal values for a nominal, ordinal or dichotomous tag.

Record: Tags are acquired and stored in the form of records. Records are analogous to paper forms: a single record (page) with many tags (text describing what data must be written to the empty block next to it). The only other attribute of records is a foreign key to indicate where the record is stored.

Storage: The storage entity is intended to allow organizations to capture all the systems where data is preserved. The storage entity has the following

attributes:

- Name
- Type
- Digital
- Interface
- Can Export
- Export Formats
- Accessibility
- Security
- Audit Trail
- Access String

These attributes allow users to store data related to the accessibility, security, and compatibility of the storage system.

Report: A report is a collection of indicators relevant to the decision-maker it must inform. A many-to-many relationship between reports and indicators models the fact that a report contains more than one indicator, but also that a specific indicator can be reported in more than one report. A report can represent indicators textually, in a table, as a graph or in any other human perceivable method. The lights on the control panel in the decision-making example used in Section 2.2 are, thus, a form of a “report” as defined in this model. The basic attributes of a report that affect data quality are the average time it takes to compile, whether it is automated and what the format of the report is.

Indicator: An **indicator**, typically in the form of a key performance indicator (KPI), is a derived value that aggregates or expresses a relationship between other indicators and tags. An indicator has a unit, type, category, source and a formula. An indicator can be any one of three types. It can be an aggregate, derived or target indicator. Aggregate indicators are simply an aggregate of one tag over a certain time frame, for example, the total accidents in past year. Derived indicators are a functions of other indicators and tags. This relationship is described in the “formula” attribute. Aggregate indicators also have an “aggregation period” when linked to a report. Examples of aggregation periods and their duration are shown in Table 4.2. Target indicators are simple decision variables included in a report as historic reference or for comparison.

Source: Each tag should ideally have a source. A source can either be an asset, a user or another tag. The most prevalent examples of data coming from (physical) assets are PLCs and SCADA systems. When a source is a user, it means that the information comes from a person observing and recording an

Table 4.2: Aggregation Periods

Aggregation Period	Length (minutes)
Calendar Hour	60
Calendar Day	1440
Calendar Week	10080
Calendar Month	40320 - 44640
Calendar Year	525600 - 527040
Production Shift	480
Production Day	1440
Production Month	28800 - 44640
Production Year	525600 - 527040
Calendar Month to date	0 - 44640
Calendar Year to date	0 - 527040
Production Month to date	0 - 44640
Production Year to date	0 - 527040
Rolling 7 day week	10080
Rolling 60 minute hour	60
Rolling 30 day month	43200
Rolling 356 day year	512640

event, or filling in a job card. The last case, where the source of a tag is another tag, is a quality issue (duplication of data).

User, organization and asset: The user, organization and asset entities are placeholder entities that attempt to represent the asset and organizational hierarchies in an organization. Even though much more complex asset hierarchies and organizational hierarchies exist, they were deliberately kept as simple as necessary to allow organizations to extend them to their needs.

4.3.2 The Tool

The previous sections in this chapter described the data pipeline model. The second component of the proposed framework is a data quality issue identification and assessment web application tool.

Proposing yet another software tool as part of a solution might seem counter intuitive in the light of the discussion in Chapter 1 (too many systems with too little compatibility). Yet, no tool to address this specific problem exists and relying on asset managers to implement the model themselves will

severely restrict the utility of the framework. Another big concern of relying on self-implemented spreadsheet models for identifying data quality issues is that the collected data will be much more susceptible to data quality issues.

The purpose of this tool is thus three fold:

1. it makes the framework more accessible;
2. it serves as a reference implementation of the data pipeline model; and
3. it allows the framework to be tested in practice.

This section documents the design and development of this tool. The solution objective requires that the framework must make it easy for asset managers to model their data pipeline, identify data quality issues and assess its criticality. This objective implies several sub-objectives. The tool must be:

- easy to deploy;
- user friendly;
- extendible;
- adaptable;
- supportive of quality data; and
- sustainable.

The development of such an information system requires careful considerations on:

- the development methodology that will be followed;
- the legal and ethical implications of a system designed to store data; and
- the technology that will be used.

These considerations had a strong influence on the design of the tool and is explained below.

An agile inspired development methodology was chosen for developing this tool. This choice was based on two primary factors: (1) the discussion on development methodologies and the design-actuality gap that frequently cause information systems to fail in Section 2.4.2 and (2) the explorative (and thus rapidly changing) nature of the pragmatic research method explained in Section 3.2.

The legal and ethical aspects of an information system (and for this study in particular) were considered in Section 3.4. The key requirements from that

discussion that were incorporated in the design of the tool are security, quality and accountability.

For the selection of technologies, a decision tree was created. The technology chosen on each branch were evaluated against the solution objectives in Section 4.2 and criteria presented in Section 2.4.3. The decision tree and selected technologies are summarized in Figure 4.7: red boxes represent a decision branch and green boxes indicate the selected technology. Some alternatives are shown in grey boxes.

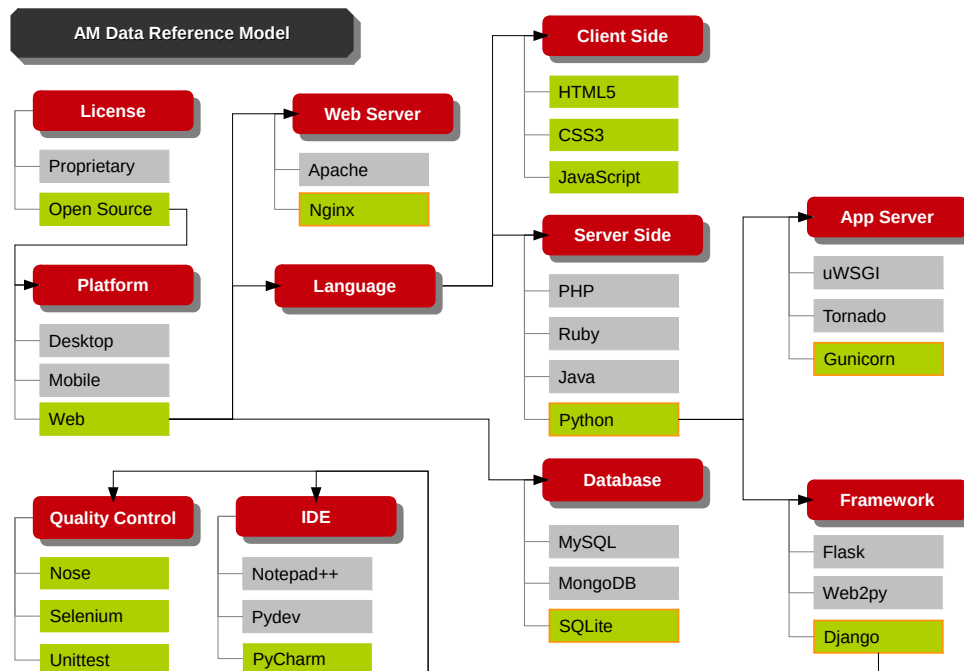


Figure 4.7: Decision tree showing some of the technologies that were considered

The first two fundamental decisions, the choice of license and technology platform, have the biggest impact on subsequent decisions. After evaluating the various objectives and constraints, a decision was made to develop the tool on open source web technologies. This decision minimized the cost, compatibility, deployment, lack of support and sustainability risks. The remaining decisions had lesser impacts on the criteria and were primarily chosen based on familiarity and personal preference.

The resulting tool was a Django-based web application with a server-less relational database (SQLite), served through the uWSGI application server. This configuration allows extreme flexibility for deploying the tool. It can be run from a virtual private server accessible through the World Wide Web, served from within a company's private network or carried around on a memory

stick. Because it is a web application, the first two deployment options allow users to access the tool from any device with a modern web browser such as their phones, tablets or laptops.

The Django framework (Django Software Foundation, 2014) “is a high-level Python Web framework that encourages rapid development and clean, pragmatic design”. Key features that set Django apart from most other web frameworks are its object relation mapper, automatic admin interface, security model, its template system and internationalization. These features allow very rapid development without sacrificing performance or security.

The discussion above explained the objectives, considerations and technology decisions made regarding the development of the tool. As mentioned, the objective of the tool is to make it easy for asset managers to document their current data pipeline and identify data quality issues.

The tool should thus not prevent the user from inputting “wrong” data (for example, the same data is stored twice in different stores). The idea is that asset managers define their “as-is” information flows and decision-making process. After the model is completed, the tool will print out any inconsistencies. Using a predefined set of rules and weights, it will highlight the problem areas that (if addressed) will give the greatest improvement in data quality.

Examples of the scenarios that will cause a “red flag” in the tool is:

- Duplication of data
- Manual collection and processing of business critical data
- Incompatible stores for data that frequently need to be analysed together
- A tag whose source is another tag
- Any other user defined “red flags”

One of the solution objectives was to make the tool user friendly. One of the ways in which this was achieved was to allow drag and drop functionality for building hierarchies. When building hierarchies, a natural way to present them is using an indented list. Using JavaScript and CSS, it is possible to build a widget that allows users to drag and drop parts of a hierarchy. Dragging entries vertically adjusts their position, while dragging horizontally allows users to change the entry’s hierarchical level. Clicking on an element allows users to edit the entry. See Figure 4.8 for a screen shot of the dragging action.

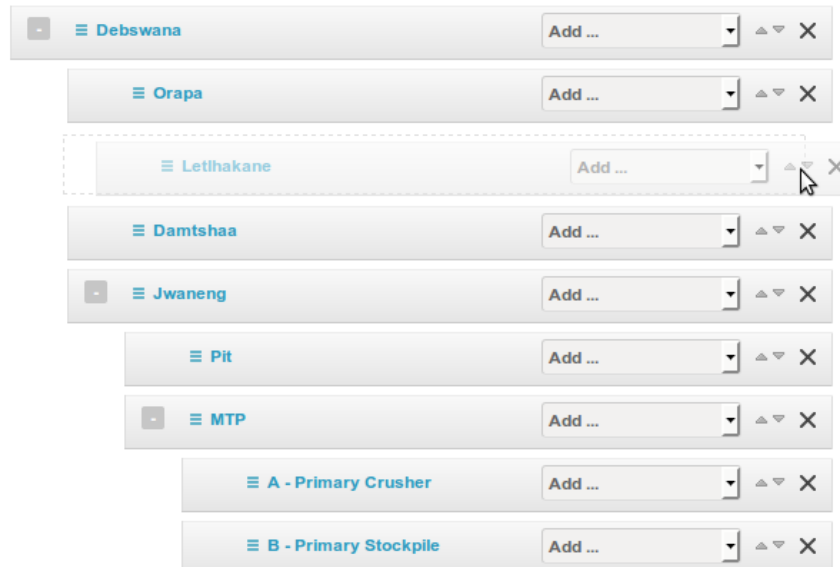


Figure 4.8: User friendly interface for building hierarchies

This section described the tool, how it was developed and how it helps asset managers to build their own data model from the reference model. The next section provides the process and guidelines to identify and assess data quality issues.

4.3.3 The Process

The aim of the process described in this section is not to get one manager's perspective on the data, but rather an "objective" perspective as told by the actual reporting process.

The process guidelines are listed below:

Understand the organization: This includes attending meetings, interviewing managers and getting to know the organization and the decisions they make on a daily, weekly and/or monthly basis. This will require managers on different levels of the organizational hierarchy to be observed and interviewed. Actively participating in the reporting process is also useful.

Deploy or implement the tool: As explained in Section 4.3.2), there are several ways to deploy the tool. If none of them are feasible, the data pipeline can also be build in a spreadsheet like Microsoft Excel, but this is not ideal due to the implications it holds for data quality. Once the tool is deployed or implemented, the data collection process can start.

Document the as-found data pipeline: It does not matter if the managers do not know the reporting process, what is important is that the reports they use in their decision-making are recorded. This might prove difficult since few decision-makers read a report and then make a decision based on it. Rather, as described in Section 2.2, decision-makers rely on reports to update their mental models on which they base decision-making.

After a few records have been added to the model, the next step is to split reports into indicators and their metadata. Indicators need to be classified into aggregate, derived and target indicators. By identifying derived indicators' formulas, they can be iteratively reduced to tags. Once the reports have been reduced to tags, time can be spent to identify the source of each tag. Keep in mind that not all tags will have easily identifiable source.

Care must be taken not to populate the organization or the asset hierarchy as a separate exercise. Instead, rely on the sources of the tags (which can either be an asset, a user or another tag). This will enable an organization to see their asset hierarchy as seen from the perspective of the reporting process.

Identify data quality issues: The assessment is twofold: First, as per ISO 55000, merely collecting data can be a great insight, but structurally storing the data allows direct comparisons and identification of issues. This combination of basic issues being automatically identified by the tool and the experience of working with the data enables the organization to much better identify and assess their data quality issues.

Address the data quality issues: Based on the identified critical data quality issues and their improved understanding of the reporting process, asset managers should now be able to make a much more informed decision regarding their issues.

4.4 Summery

This chapter began with an overview of the thought process that lead to the solution objectives documented in Section 1.3. The first solution objective required that a reference model be developed to describe an organization's asset management decision-making data pipeline in order to identify and asses data quality issues. This was achieved by ensuring that the relevant entities and attributes related to each stage of the data life cycle and the possible data quality issues (both described in Section 2.3) could be captured. The result is that all stages, from data acquisition (data source) to sharing (reports) can be represented along with attributes that might indicate data quality issues such as duplication of data (indicated by the source of a tag being another tag) or reports that are manually compiled.

The second objective was to develop a tool to serve as reference implementation and allow the framework to be of practical value. The tool was built on open source technologies to allow easy deployment on a wide variety of platforms without worrying about licensing fees. Since the reference model was designed to be independent of specific technologies or standards, the implementation had to add additional entities for user management, data security, logging changes and a wiki. Enabling users to graphically extend or adapt the model for their organization (“local improvisations”) were deemed to be too difficult to implement. Instead, the wiki component was added to give users the necessary flexibility to verbally describe what they could not capture using the entities and their relationships described in the reference model.

Finally, Section 4.3.3 provides a simple process description to guide users on how to approach the process of identifying and assessing data quality issues in asset management decision-making. The next chapter documents the application of this process to solve a subset of the original, real-world problem of inadequate data that prompted the undertaking of this study.

Chapter 5

Case Study: Diamond Treatment Plant

“The pragmatist knows that doubt is an art which has to be acquired with difficulty” – Charles Sanders Peirce

Chapter 4 described a framework for identifying and assessing data quality issues in the asset management decision-making process. As mentioned in Chapter 1, the problem that this framework addresses was identified in a case study executed at a Debswana owned and operated diamond mine. This chapter documents the “data integrity project”.

Debswana is a partnership between the De Beers Company and the government of Botswana. The board of directors of Debswana acknowledged the importance of asset management in their 2012 report to their stakeholders. Due to reduced output (“production slowdown”), the board initiated an “Asset Management Improvement Plan” to focus on asset management in their operations. This production slowdown was a result of “challenges with earth-moving equipment and treatment plants”. They also acknowledged the importance of data by reporting that a “considerable focus was placed on developing key performance indicators to drive improved performance in all stages from drilling to recovery” (Debswana 2012).

As part of this focus, the Debswana head office, which relies on operational and engineering reports from their mines, initiated a pilot project to investigate and improve the quality of the data coming from their mines. The scope of the pilot project was limited to data coming from the main treatment plant of Jwaneng (one of their diamond mines) and limited to 100 days.

The length of this “100 day data integrity project”, and its limited scope, made it an ideal project for exploring the problem through a pragmatic research method - in parallel with a strong literature review - and developing a framework (a knowledge “artefact”) for addressing the issues. Thus, the data integrity project not only provided the opportunity to translate a real

world problem to a research problem, but also (partially) to design, develop, demonstrate and evaluate a framework *on site*.

This chapter presents the results from the 100 day data integrity project to demonstrate the value of the framework (described in Section 4.3). The structure of this chapter follows the process as defined in Section 4.3.3.

5.1 Understanding the Organization

The Debswana Diamond Company (Debswana) is a fifty-fifty partnership between the Government of the Republic of Botswana and De Beers, a subsidiary of Anglo America. Debswana, originally known as De Beers Botswana Mining Company, was established in June 1969. Debswana employs more than 4 000 people, making it one of the largest employers in Botswana, second only to the state. Revenue derived from diamonds mined by Debswana accounts for 33% of Botswana's GDP and 50% of Botswana's public revenue. Figure 5.1, from the 2012 stakeholder report, shows the structure of Debswana. (Debswana 2012)

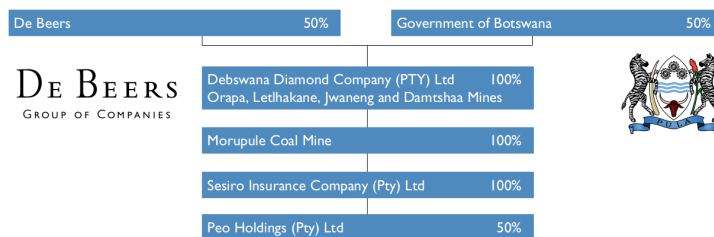


Figure 5.1: Debswana ownership and subsidiaries

The Debswana Corporate Centre in Gaborone, Botswana's capital, acts as the headquarters for their diamond mining operations located at Orapa, Letlhakane, Damtshaa (OLDM) and Jwaneng. They also own and operate the Morupule Coal Mine.

The data integrity improvement project was piloted at Jwaneng, the second newest of the four mines operated by Debswana. The Jwaneng diamond mine began operations in 1982 and is located in south-central Botswana, about 120 kilometres west of Gaborone (see Figure 5.2). Jwaneng is an open pit mine that moves 37 million tons of waste rock and processes 9.3 million tons of ore per year to produce approximately 11 million carats (2,200 kg) of diamonds.

Jwaneng has the highest carats per ton ratio of all Debswana's mines with an average of 1.25 carats per ton. The high rate of diamond extraction, combined with high quality diamonds, make the Jwaneng diamond mine the richest diamond mine in the world by value of recovered diamonds. Jwaneng Mine employs approximately 2100 people and is known for its safety record, winning

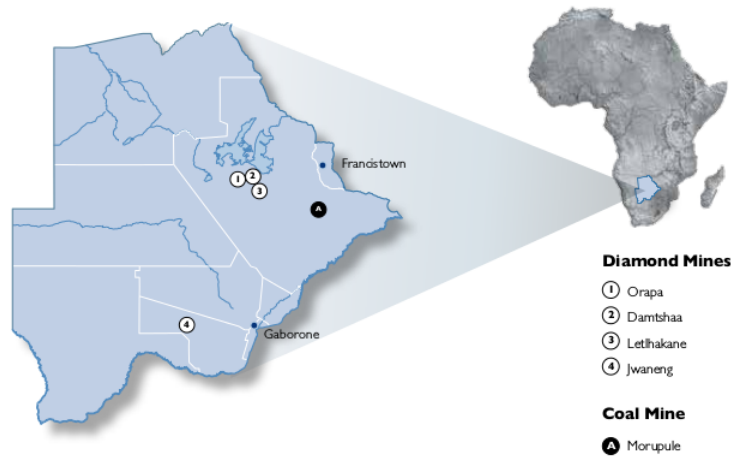


Figure 5.2: Location of Debswana's mining operations

multiple national and international safety awards. The mine maintains an OSHAS 18001 certificate for safety and an ISO 14001 certificate for environmental compliance (Debswana 2013).

The Jwaneng value chain consists of six components: mining and blasting, crushing, treatment, recrush, recovery and sorting. Each of these components is discussed in more detail below.

Mining and blasting: Jwaneng mine is an open cast mine and thus requires cutbacks of increasing depth into the surrounding earth to access the kimberlite (diamond bearing ore) deposits. Blasting is used to break up the kimberlite, which is then taken to the primary crusher. The waste material is hauled to the waste dumps.

Crushing: Ore is tipped by dump trucks into the primary tipping bin (MAD-20-09). It is then fed into the primary gyratory crusher (MAD-20-20) by means of an apron feeder (MAD-22-28). A boom hydraulic rock breaker (MAD-25-19) is used to reduce oversized lumps and aid material flow from the tipping bin. Material is crushed to a nominal particle size of 200mm. Crushed material exits from the gyrator crusher and passes through a discharge chute (MAD-20-07) onto a sacrificial conveyor MAX-20-18. The product discharges through discharge chute MAX-04-07 onto the primary stockpile feed conveyor MAX-28-18. Tramp metal is removed from MAX-20-18 with belt magnet MAX-20-45. Several other dust extraction and suppression systems, spillage pumps and raw water reticulation pumps support the primary crushing process.

In addition to the gyratory crusher, Jwaneng also owns a mineral sizer. The mineral sizer (MAF-01-20) is semi-mobile installation that serves as a

standby primary crushing unit. The mineral sizer is fed by front-end loaders and is designed to achieve a feed rate of 2000 tons per hour.

Material from both the gyratory crusher and the mineral sizer are transferred by conveyor to the primary stockpile.

The second phase of preparing the ore for treatment is the scrubbing phase. Material is withdrawn from the stockpile through two extraction streams, each fitted with two apron feeders. Each apron feeder is capable of feeding a maximum of 2000 tons per hour of (dry) ore. Both streams have belt magnets for removing tramp metal and transfers ore onto the Plant Feed Conveyor. The Plant Feed Conveyor can discharge into the scrubbing surge bin of the first scrubbing stream or onto a shuttle conveyor which can then deliver ore to the next scrubbing stream. This pattern is repeated so that ore can be delivered to any of the four streams. An automatic bin filling system assigns priorities to the four bins based on the bin level, whether the bin is feeding the scrubbing module and whether the stream is being maintained. The designed capacity for each stream is 550 tons per hour.

Oversize material (material that does not fit through a 25mm grid screen) from the scrubbing and screening section is fed onto the Scrubbing Oversize Conveyor (MCX-10-18) and transferred to the Secondary Crushing Silo. The Secondary Crushing Silo feeds the secondary crushers (Cone Crushing Section). Cone Crushing Oversize Conveyors return oversized material to the Secondary Crushing Silo for another round of crushing. Fine material (material below 25mm diameter) is delivered from the scrubbers and cone crushers to the feed preparation stockpile.

Treatment: Once the ore has been crushed to a manageable size, it is put through a series of washing and screening processes. Next, it is mixed with slurry to separate diamond bearing kimberlite ore from waste particles. This process is known as “Dense Medium Separation” (DMS) and it relies on the fact that diamonds are heavier than the material it is typically surrounded with.

Recrush: Since the process described above is not very sophisticated, the “waste material” from the DMS plant is sent to the Recrush plant. In the Recrush plant, ore is again crushed using a more refined process. This plant helps extract smaller diamonds not recovered in the original treatment process.

Recovery: The “streams” coming from DMS and the Recrush plant are fairly diamond rich or “concentrated” by this stage. To further concentrate the stream, three more properties of diamonds are exploited: they repel water, stick to grease and emit light under X-rays. When mixing the stream with water and running it over a grease belt, diamonds will thus stick to the grease where laser sorters can pick them out. This process forms part of the

Completely Automated Recovery Process (CARP), which is one of the two components of the “Aquarium” project at Jwaneng.

Sorting: The other component of the Aquarium Project is the “Fully Integrated Sort-House” (FISH). FISH is completely automated and is thus a secure and efficient technology for sorting, cleaning, packaging and weighing diamonds. Jwaneng’s Aquarium also treats scheduled deliveries of concentrated diamond streams from the three other mines operated by Debswana.

The packaged diamonds are then sent to the Botswana diamond trade centre in Gaborone where it is sorted into more than 15 000 different categories before being sold and distributed for cutting and polishing.

As described above, the Jwaneng value chain stretches all the way from the pit to the packaged diamonds. The data integrity project, however, was restricted to data originating from between the primary crusher and the feed preparation stockpile. This area will be referred to as the “Main Treatment Plant” (MTP) in this chapter. The MTP is responsible for “crushing”, the second component of the Jwaneng value chain. An aerial photograph from Google Maps with annotations of this area is shown in Figure 5.3.

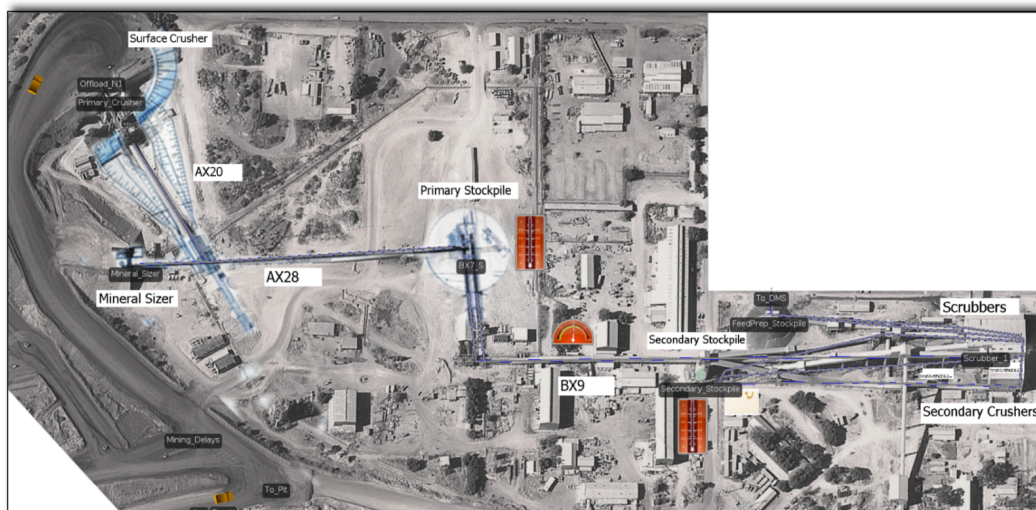


Figure 5.3: Aerial view of Jwaneng’s Main Treatment Plant

The left-hand side of Figure 5.3 shows the gravel road going past the mineral sizer to the primary crusher. In the middle of the figure is the primary stockpile with an overlay indicating its maximum diameter. The right-hand side shows the building containing the secondary crushers, scrubbers and screens. Ore from the primary stockpile is delivered to a bin at the top of this building. From there, the four parallel scrubbing and crushing streams descend vertically. After scrubbing, screens redirect oversized ore for secondary

crushing. Fines (ore with diameter below 25mm) are sent to the feed preparation stockpile. Ore that is still too big after secondary crushing is recirculated until it is the desired size.

Besides the crushers, scrubbers, screens and stockpiles already mentioned, conveyors also form an important part of the process. Figure 5.4 shows the primary components of the MTP process and their equipment codes.

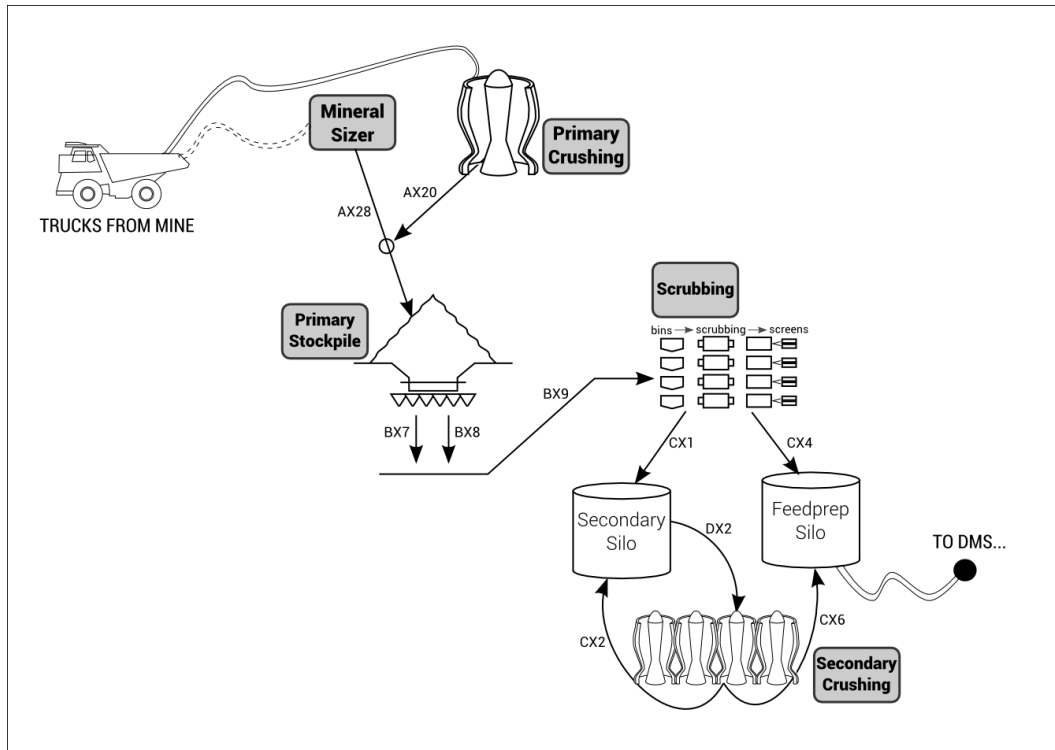


Figure 5.4: Jwaneng MTP process flows

The MTP is operated 24 hours a day in three shifts of eight hours each. Due to the pervasiveness of programmable logic controllers (PLCs) in industrial equipment, the whole MTP can be controlled in real-time using a SCADA (Supervisory Control And Data Acquisition) system. At Jwaneng's MTP, operators log into the SCADA system from the Central Control Room (CCR). The CCR is located to the left of the MTP in Figure 5.3.

Ideally, most of the operators' time should be spent *monitoring* the plant, rather than *controlling* it. This is because the various actuators and sensors connected to the PLCs actually allow the plant to be semi-automated. Analogue signals from various devices are used for continuous measurement of power, level, flow, mass, pressure, temperature and speed. There are also a number of digital signals which indicate switches for position, flow, temperature, pressure, storage level and alarms.

Several process and safety interlocks prevent equipment from damaging itself or overloading the plant. For example, if a sensor detects that the feed preparation stockpile is becoming too big, the system will automatically turn off one or more crushing and scrubbing streams. Similarly, if the recirculation rate is becoming too high, the system will reduce the feed rate of ore from the primary stockpile. The SCADA system used by Jwaneng provides an interactive graphical interface which will alert operators when any interlock becomes active. This interface also displays real-time process and equipment information as reported by the sensors.

Figure 5.5 shows the SCADA graphical interface for “Scrubbing Stream A”. The red bin at the top indicates that the actuator that opens the bin to allow ore to flow into the scrubber has an active interlock. This is likely due to the crushing stockpile being at 90.5% capacity (as indicated by the yellow label on the right-hand side of the figure). The crushing stockpile is a silo that provides a buffer for the crushing and scrubbing streams. Various other labels showing data such as the feed rate of conveyors (in tons per hour) or the rotation speed of the scrubber can also be seen in 5.5.

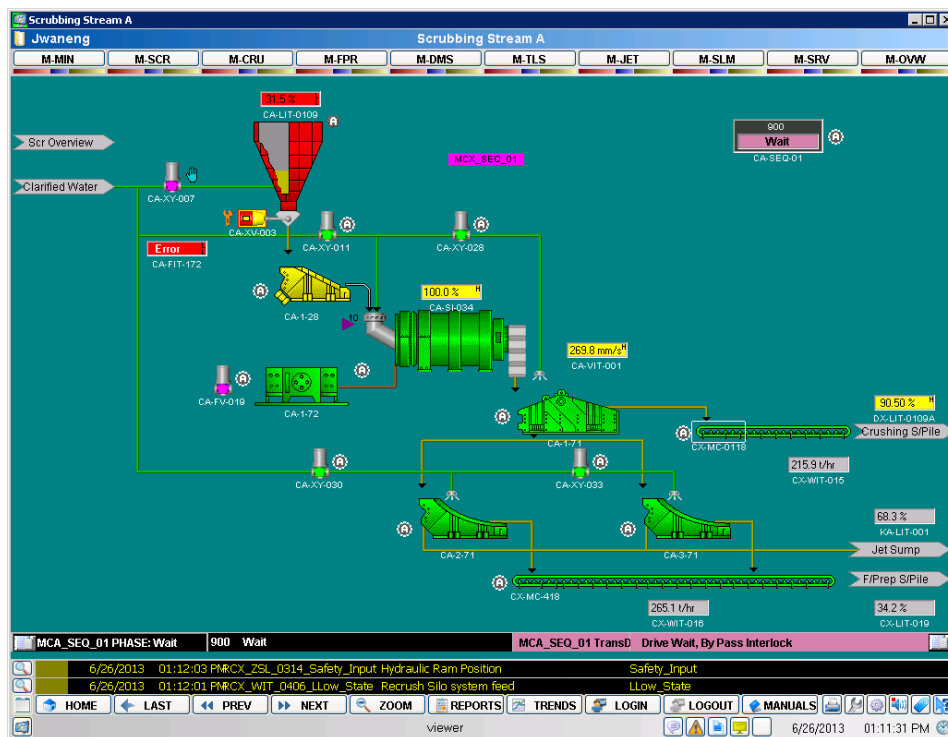


Figure 5.5: Example view of graphical interface for SCADA

As mentioned, the primary duty of operators is to monitor the SCADA system for any potential issues. Operators in the CCR are in constant radio contact with workshops and foremen in the plant. When an interlock auto-

matically engages, CCR sends someone to investigate in the plant. They also ensure that safety interlocks are in place and give clearance for artisans to perform inspections and maintenance work.

Maintenance jobs are coordinated by the various workshops spread throughout the plant. Each workshop has a foreman, a planner and scheduler and a team of artisans. Artisans are specialists in various technical fields such as boilermakers, mechanics, electricians and control and instrumentation technicians.

Planners meet several times per week and a maintenance schedule is worked out. Planned shutdown of the plant to perform maintenance happens every second Wednesday. Every week one stream is also stopped to perform maintenance.

The work process for maintenance and repair jobs at Jwaneng is well-defined. When an artisan (or anyone in the plant) identifies an issue, they can either contact CCR (over radio) or their foreman or workshop planner to create a notification (IW21 in SAP). The foreman in charge of the shift will review the notification and decide if it is an emergency (which requires immediate attention) or if it can be scheduled for later attention. If it is an emergency and not the night shift, the foreman issues a job card (IW22 in SAP). The CCR then has to issue permits and put safety interlocks in place to allow the jobs to be safely completed. If it is during the night shift, a maintenance job ticket is completed and only converted to a job order on the following shift. After any job is completed, the job order should be updated.

Another duty of the operators in CCR is to keep track of production metrics. Every hour they copy the production values to a logbook (paper). These values include metrics such as tons per hour for various conveyors, number of truckloads delivered and from where, amount of flocculants used and water usage.

Every two hours they will also compare the feed rate of the main conveyor (head feed) taking ore from the primary stockpile to the crushing and scrubbing streams. If the feed rate is below the target (for the period of the project this target was 1450 tons per hour), operators must log what the reason for this was. Since feed rate lower than 1450 tons per hour translate to “lost time”, these delays are recorded in minutes. Thus, if the head feed was only 1000 tons per hour, and it was determined that the delay was due to metal detected on the conveyor (which caused the conveyor to stop until the metal is removed), an entry for 20 minutes with delay description of “metal detected on BX9” will be recorded.

All issues are recorded in a logbook which is then entered into a spreadsheet every two hours. All safety, health and environmental incidents are also logged by CCR. Lastly, there is also a “walkabout” action log in which any additional comments affecting the plant are recorded.

The operator entering the logs to Excel is also responsible for assigning several categories to the delay description. These categories include location

such as “Primary Crusher” or “Scrubber” and responsibility such as “engineering” or “operations”. These sheets are all linked to a “daily report” which summarises the previous day’s operational, safety, health and environmental metrics. Every morning, representatives from the mining, engineering and process department meet to discuss this report. An important function of this discussion is to allow representatives to dispute delays incorrectly booked against their department. If issues are found, it should be corrected in both the report as well as the Excel source document. This is important, since quite a few of Debswana’s KPIs is derived from whose responsibility a delay was.

For weekly, monthly and ad hoc reporting, the process department relies heavily on this report, but will sometime also retrieve data from Wonderware Historian. The engineering department also relies on the daily production summary, but include data from SAP PM, which is the CMMS used at Jwaneng.

These documents, processes and information flows is documented in Figure 5.6

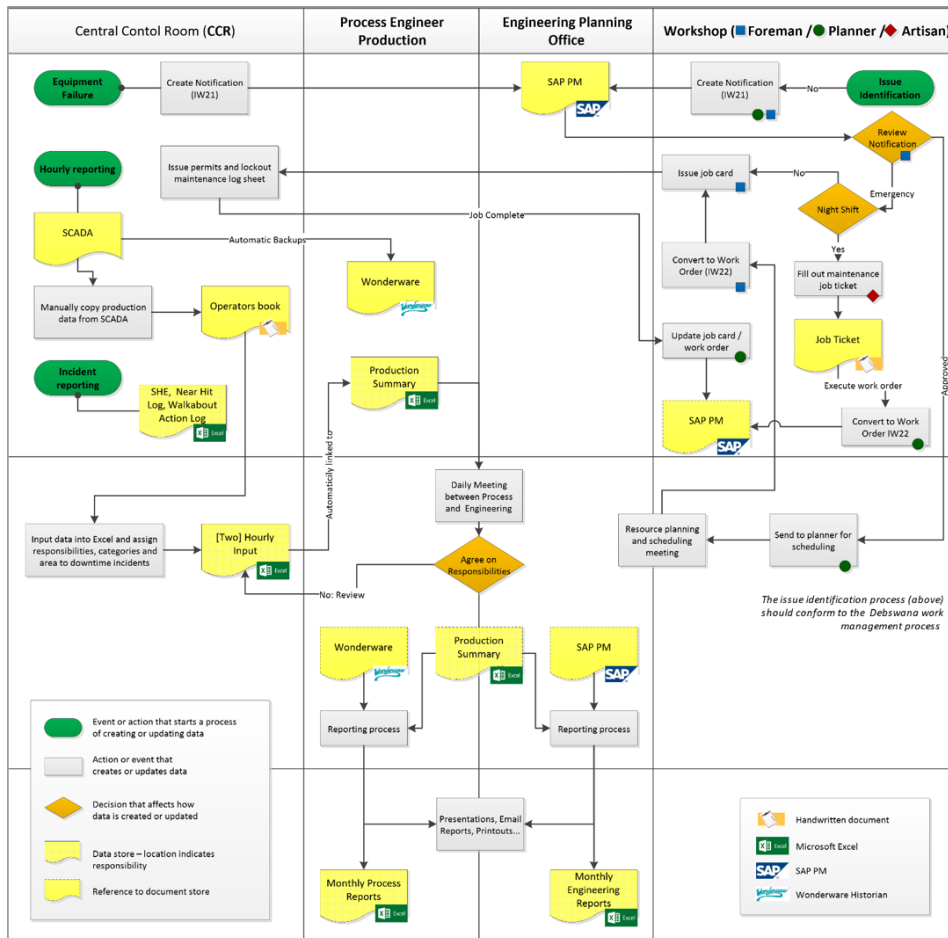


Figure 5.6: The as-found information flows

These reports are then emailed to superiors and/or head office, used for meeting presentations and printed out for filing.

This section provided an overview of Debswana, Jwaneng, the diamond value chain, the extraction process, the work order process and information flows for reporting in the MTP. This data was collected from various reports, documentation, informal interviews, and practical experience. In the next section, the reports, data storages and sources first introduced in the process and information flow model (Section 5.1) are formally documented using the data model developed in Section 4.3.1.

5.2 The Data Model

The model presented below is a snapshot of the author's interpretation of the data pipeline used at Jwaneng. This "as-found" state was documented through informal interviews with various managers, engineers, operators, planners and IT personnel. Some data was also documented by walking the plant, observing operators and participating in the acquisition, processing and sharing stages of the data life cycle. The initial data was used to shape the reference model described in Section 4.3.1, which was then used to collect more data as the model evolved. This iterative process of designing, developing and evaluating was repeated for the duration of the data integrity project. The data in this section has been sanitized to remove all personal and sensitive organizational data as discussed in Section 3.4.

The data model entities are presented roughly in the order they appear in Figure 4.6. The first entity discussed is thus reports.

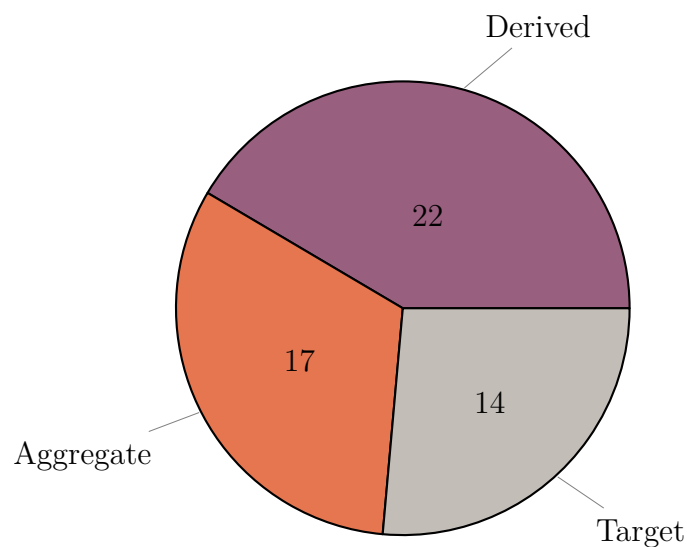
Reports are the final product of the data pipeline. As defined in Section 4.3.1, reports are what informs the decision-making process. Jwaneng, however, did not have a "report registry" or any formal documentation defining the types of reports, their layouts or content. The reports discussed below were discovered through informal discovery process described in Section 4.3.3.

Throughout the data integrity project, 14 reports were encountered. Eight of these were ad hoc reports for a specific project, meeting or presentation. Four were specialist reports with limited audience. Of the four, three originated from the condition monitoring laboratory and one from contractors working on the conveyor belts. The remaining three were regular interval reports with a fixed layout and wide distribution. Senior asset managers at Jwaneng recommended that the demonstration and evaluation be based on these three reports. The three reports are: (1) the Month-end Production Report from the Ore Processing Department, (2) the Weekly Engineering Report from the Engineering department and (3) the Daily Production Report from the CCR. Table 5.1 shows the records for the three reports as stored in the report model.

Table 5.1: Report Table

Name	Daily Production	Weekly Engineering	Monthly Production
Interval	Daily	Weekly	Monthly
Format	Excel Dashboard	Emailed Graphs	Excel Data Table
Automated	TRUE	FALSE	FALSE
Source References	TRUE	FALSE	FALSE
AverageTime	-	3 days	1 week
Target	Planning Office	Head Office	-

The next step was to break each of the three reports into its indicators and link them together. In total, 53 unique indicators were extracted. Of this 53, 22 were derived, 17 were aggregates and 14 were targets (see Figure 5.7).

**Figure 5.7:** Distribution of indicator types

Of the 17 aggregate indicators, six were aggregates of delay categories expressed in hours and five were counts of events. Since indicators are defined independent of a time frame or scope, many are reused in reports with different scopes. For example, Engineering Availability can be sliced in terms of any arbitrary combination of a time frame (monthly, weekly...), plant section (scrubbing, crushing...) or profession (electrician, boilermaker...) amongst others. At Jwaneng, the top five indicators (as measured by the number of occurrences in reports) are:

- Engineering Availability
- Equipment Availability
- Engineering Utilization
- Overall Plant Utilization

- Tonnes Treated (Wet)

There are many ways to categorize indicators. Figure 5.8 is a graph showing a distribution of indicator categories. Although not an indicator of data quality, it does provide an interesting overview of the type of information that is reported and sparks discussion on whether this is what *should* be measured.

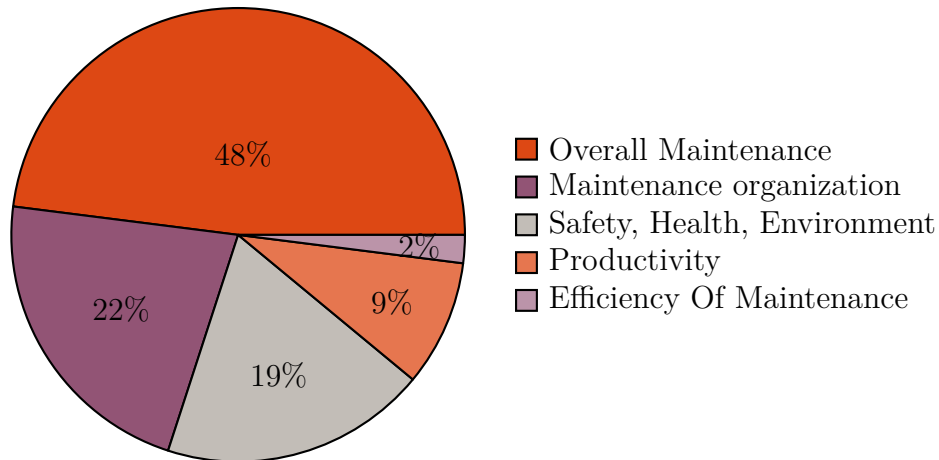


Figure 5.8: Distribution of indicator categories

When attempting to compare indicators, it was initially erroneously assumed that the period for which an indicator is reported is fixed. From the analysed reports, 17 periods used for aggregation were found. It was also discovered that not all these periods were of fixed length. Thus, the data pipeline model was extended to also record the aggregation periods as shown in Table 4.2.

Indicators can either be a tag, an aggregate of a tag, or a function of several tags and/or other indicators. Indicators, tags and records, their definitions and their relationships have already been discussed in detail in Section 4.3.1. Due to space restrictions, not all indicators will be discussed in this section. Instead, one indicator will be traced in full to illustrate the process. The indicator is engineering availability. Total delays are reported or used in almost all reports encountered at Jwaneng.

Engineering Availability (ENA) Debswana defines **ENA** as:

“Engineering availability is a measure to determine the percentage of time the Engineering function requires to maintain equipment, both scheduled and unscheduled.”

ENA measures the degree to which equipment is in an operable state at the point in time when it is needed. It is a measure of maintenance effectiveness.

The equation is:

$$\mathbf{ENA} = \frac{T100 - D100 - D200}{T100} \times 100 \quad (5.2.1)$$

T100, D100 and D200 is defined below:

- **T100** = Controllable Time; Available equipment time attributable to any internal factors under the control of the operation that impacts production
- **D100** = Unscheduled Maintenance; Downtime as a result of maintenance work not included in the confirmed weekly maintenance plan
- **D200** = Scheduled Maintenance; Downtime as a result of maintenance work included in the confirmed weekly maintenance plan

T100 is determined by the production schedule and is stored in the production calendar (an Excel workbook updated once a week). D100 and D200, however, are not measured directly. How they are derived will now be discussed.

There is a single conveyor belt (MBX-09-018, or simply BX9) transporting ore from the Primary Stockpile to the MTP (see Figure 5.3). The weightometer on this conveyor is used as an estimate for the ore processing rate of the entire MTP. Each year the annual tonnage call is fixed. Since the plant only closes for a few public holidays throughout the year, the total manned hours, or “controllable time” (T100), is then used to derive the hourly call:

$$\text{Hourly call} = \frac{\text{Annual Call Tonnage}}{\text{Annual Manned Hours} \times \text{Target Overall Utilization}} \quad (5.2.2)$$

Annual Call Tonnage, Annual Manned Hours and Target Overall Utilization are all “decision variables”. Their values are decided upon based on historic data and strategic objectives and is usually fixed for the year. Trying to capture them in the data pipeline model, however, proved to be difficult as they did not have a designated record or storage. Their values were thus acquired through informal discussions with managers and engineers.

The hourly call value for the year 2013 was 1450 tons per hour. Total delays (in minutes) for a given period, is defined as a function of actual ore produced and manned hours (MH), as these are both values readily available:

$$\text{Delays} = \text{MH} - \frac{\text{Estimated processed tons}}{\text{Hourly call}} \times 60 \quad (5.2.3)$$

For example, if at the end of a regular, eight hour shift the BX9 totalizer shows 10150 tons passed into the MTP, it means there were $(8 - \frac{10150}{1450}) \times 60 = 60$ minutes of delays.

Thus, to calculate delays for a given period, only the scheduled production hours and the estimated amount of ore processed is required. Since Jwaneng's MTP is scheduled to produce 24/7 (except for some public holidays), MH is simply calendar-time. Estimated processed tons is available through a SCADA interface that graphically represents digital and analogue PLC signal values in human readable format and is updated in real-time. It also provides basic reporting functionality to allow operators to, for example, view the totalized values of a weightometer for a selected period. These values can be exported to a comma separated (csv) file.

To calculate ENA, however, delays must be categorised in order to differentiate between Unscheduled Maintenance (D100) and Scheduled Maintenance (D200). To this end, CCR operators check their SCADA interface for the average flow rate of several weightometers every hour. These values are typed into an "hourly production log" Excel workbook. If the average flow rate for BX9 is below the hourly call, CCR operators are required to determine the cause and write a description of the delay in the corresponding two-hour interval in their production logbook.

Throughout the shift, the manually written delay records are retyped into the "two-hourly production log" Excel workbook. The flow rate for the corresponding two-hour interval is retyped from the "hourly production log". Equation 5.2.3 is then used to display the delays in minutes for that two-hour period. At the end of the shift, the delay descriptions and the delay minutes are again manually retype back into the "hourly production log" Excel workbook. These "daily delay summary" records consist of seven tags:

- Description
- Shift
- Duration (minutes)
- Plant Area
- Section
- Responsibility
- Type

Two more tags (Date and Shift manager) are derived from the file in which the record is stored. The last five tags' values (Plant Area, Section, Responsibility and Type) are selected from a drop-down at the operator's discretion. The operator selects these values based on the description and might radio a workshop or foreman to get more information. Even though it is the same Excel workbook, it was added as a separate record in the model. Figure 5.9 shows screen-shots of the process described above.

The "responsibility" tag from the "daily delay summary" record finally provides a tag that can be used to differentiate between D100 and D200 downtime so that ENA can be calculated. The responsibility tag can have one of five possible values:

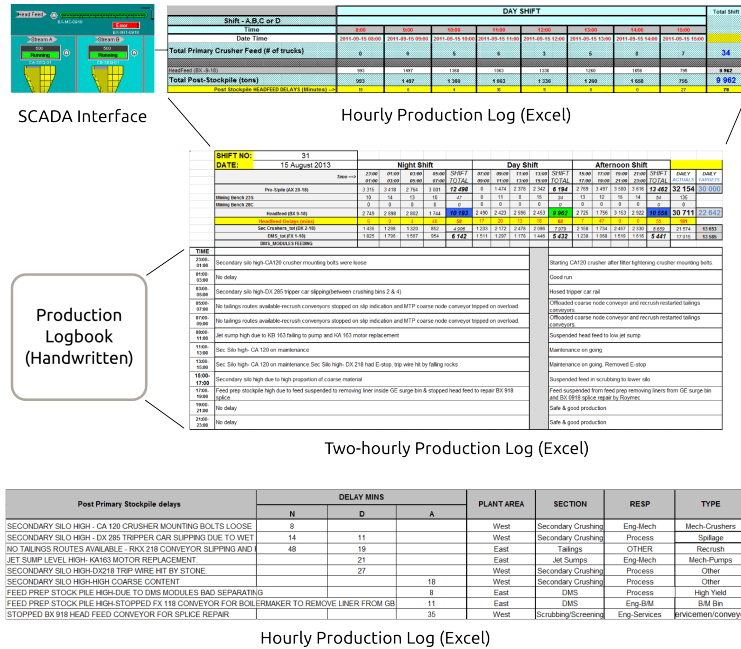


Figure 5.9: How delays are recorded

- Maintenance
- Engineering
- Process
- Mining
- Other

When a delay record’s responsibility is classified as “engineering” it contributes to D100 and when it is classified as “maintenance” it counts as D200.

Finally, ENA’s formula can be represented in the data pipeline model as the following SQL:

```

SELECT ALL ((SUM(T100) - SUM(D100) - SUM(D200))/(SUM(T100)))*100 as ENA FROM (
    SELECT record_id,
           SUM(CASE WHEN responsibility='Engineering' then 1 END) as D100,
           SUM(CASE WHEN responsibility='Maintenance' then 1 END) as D200,
    FROM tag
    WHERE record_id = 3          -- the 'daily delay summary' record
    GROUP BY production_day;
) delays,(
    SELECT record_id, production_day, SUM(manned_hours)
    FROM tag
    WHERE record_id = 2          -- the 'production calander' record
    GROUP BY production_day;
) manned_hours
WHERE production_day BETWEEN [start] AND [end];
    
```

This is the SQL generated by the tool described in Section 4.3.2.

From Figure 5.6, four types of storage systems were encountered: handwritten documents, Microsoft Excel workbooks, SAP PM and Wonderware Historian. Even though production data is available through Wonderware which is an automated, real-time reporting system, the source of the delay data is not Wonderware, but the Excel workbook which is a static copy of the production data.

The next section looks at these data quality issues in more detail.

5.3 Data Quality Issues

The previous section demonstrated how the framework was applied and used ENA as an example of how an indicator that was used in several reports can be reduced to its tags and their sources. As mentioned, the framework aids asset managers to identify data quality issues in two ways. Firstly, a simple and extendible rule-set allows issue to be automatically identified as the data pipeline model is populated. Secondly, the act of creating and using a tool as described in the framework “can stimulate and improve organizational knowledge and decision-making” (ISO 55000 2014, Section 2.5.2 (a)).

First, each of the storages were analysed for intrinsic design quality issues. SAP PM and Wonderware are proprietary systems with proper authentication and security, with backup and auditing. These systems did not have any intrinsic data quality issues. The Microsoft Excel based systems, however, does not have proper security, no auditing and is only backed up (indirectly) to magnetic tape, once a day. These systems are not complete either. When delays must be categorised, dropdowns are provided, but these values are not a comprehensive list of all root causes. The “other” option also causes garbling of input since it is not possible to map “other” back to a meaningful real world state.

When assessing the larger reporting process, many more data quality issues were identified. The most critical issues were:

- Disconnected/incompatible data sources
- Unprotected data input: entering data into Excel is prone to corruption and accidental changes/deletions.
- Slow, manual reporting process
- Inconsistent Ontology (assets have multiple codes used inconsistently across the various systems)
- Two indicators with same name (engineering utilizations), but connected to different reports have different formulas.

- Few indicators define an aggregation period (e.g. monthly and when the month starts)
- Of the 48 584 automated records, only 23 are used in reports
- The data is acquired multiple times through different sources (production rate and delays)
- Storage systems are frequently used as manual source for tags (data is retyped instead of dynamically linked).

To validate the framework, these issues were solved to evaluate whether the framework was useful. The next section evaluates the framework.

5.4 Evaluation

Based on the critical issues identified in the previous section, a simple reporting tool was developed. This automated reporting tool is described below.

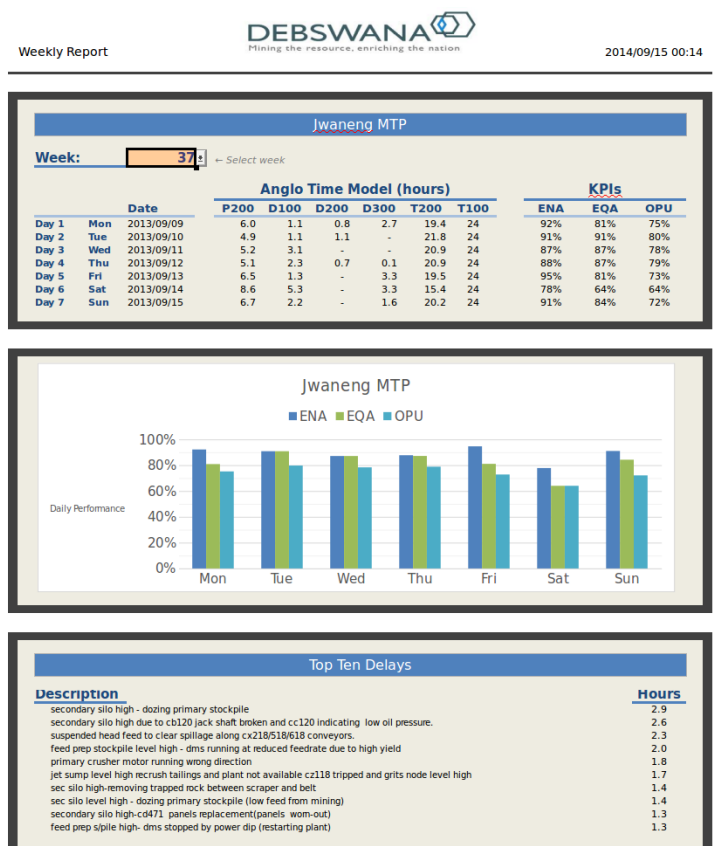


Figure 5.10: The reporting tool

The reporting tool aggregates data from Excel sheets across the network drive after each shift. The collected data is cleaned and tagged. Semi-structured data such as the delay descriptions are analysed and categorized. The resulting data set is stored in a database file that can easily be imported into Excel for further analyses. It also makes important data that is otherwise trapped in Excel available for future dashboard projects. A protected Excel sheet makes use of this database file to report KPIs “live” (every time the file is opened, the latest data is retrieved). This automated dashboard is displayed in Figure 5.10. Using an equipment code map, it is possible to retrieve all data for specific equipment from SAP, Wonderware and Excel. Figure 5.11 shows this improved process flow.

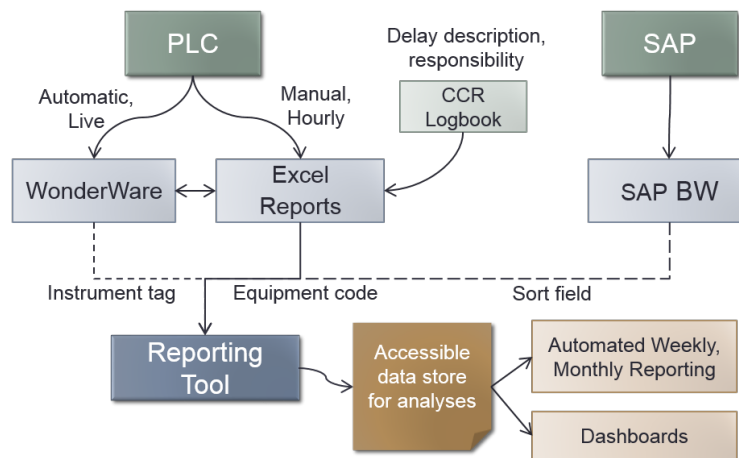


Figure 5.11: The improved process flow

Most quality issues identified were due to human error. The reporting tool eliminated almost all human interaction and, thus, drastically increased the quality and speed of the reporting process. Yet, a better understanding of the social and cultural context might have revealed additional insights into the data reporting process. Baum and Vlok (2013) suggested that Social Network Analysis is the ideal tool to map and understand social relations in an asset management context.

From the demonstration of the framework in this chapter, it can be concluded that the primary research objective has been achieved. The framework is practical and was useful for identifying critical data quality issues. Figure 5.11 shows the new simplified and automated reporting process that was implemented to address the critical data issues identified by the framework presented in this study. Its usefulness has also been affirmed by feedback from asset managers at both Jwaneng Diamond mine and Debswana’s head office.

Chapter 6

Closure

This concluding chapter reflects on the methodology, experience and usefulness of the development of the framework and makes recommendations for future studies in Section 6.1. The primary and secondary contributions (both practical and scientific) are presented in Section 6.2 and the study is concluded in Section 6.3.

6.1 Reflection

The issue of inadequate data available for asset management decision making was identified in practice. When this problem was translated to a research objective, it was decided to keep a narrow focus so that the solution might be of immediate value for the diamond mine processing plant where the case study was conducted. This was also the case: when applied, the framework identified several critical issues that could then be solved by automating the data collection and cleaning process.

A key component that contributed to the framework being useful (and thus satisfying the objective of this study), was the parallel interaction of the literature review and the case study to inform, guide and evaluate the development process. The study was not just tested in practice, but originated from practice. This ensured that the study stayed relevant, while the comprehensive literature review provided the theoretical credibility required of a research study. Running these activities in parallel, however, did mean that the study went through many changes in the primary research objective and scope of the problem. This means that the documented framework developed in this study is just one solution to one aspect of the much bigger problem of inadequate data quality in asset management decision making (described in Chapter 1). It must also be noted that by having adopted a pragmatic world view, this study cannot make claims of value beyond the specific case study. Yet, a significant part of the solution was also derived from the literature review and an attempt has been made to not include design decisions that seemed too

specific to the case study.

Still, more studies are required to evaluate the framework in different contexts. Since the study had a very narrow scope, there are a lot of knowledge gaps that must still be covered before a complete solution for fixing data quality issues can be developed. Significant questions that were not addressed by this study include:

- How do you determine what data an organization *should* process and report on? The framework only identifies issues with an organization's existing data pipeline (such as data recorded and processed, but not used in any decision-making). It doesn't indicate what data is *not* being processed and reported on.
- How do you address data quality issues stemming from human error, both accidental and/or deliberate? When the framework identifies issues such as duplication of data, it is mostly trivial to fix. This is not the case for user related data quality issues since they frequently stem from lack of skills, motivation and/or job satisfaction.

Another advantage of the pragmatic research methodology was that by constantly switching between the literature and the case study, discrepancies between reality and the literature were easier to identify. For example, literature recently began describing asset management as a comprehensive and strategic management system, but in practice asset management was still perceived as maintenance biased towards "fix it when it breaks". The development could thus take the literature into account, but also make adjustments to reflect the reality observed during the case study. The result, as previously mentioned, is a framework for identifying data quality issues affecting asset management decision making that has been successfully applied in practice.

However, not all components of the framework were equally successful. The automatic assessment of data quality issues, in particular, proved to be much less useful initially than hoped. The suspected reason for this is that if you are able to develop the logic for detecting an issue, you know what the issue (or at least the symptom of the issue) is. Thus, most issues were not identified through the automated logic, but through the process of building the model. Another aspect of the tool that could benefit from further study is its ease of use. Developing an interface that is both user friendly and complex enough to allow it to be customized by the user turned out to be too ambitious for the scope of this study. There is, thus, also opportunity for improving the framework components themselves, especially the tool for building the data pipeline model.

6.2 Contribution

The primary contribution of this study is a framework, consisting of a reference model, a tool and a process guide for identifying and assessing data quality issues.

Secondary research contributions include:

- Insight into the asset management; decision-making; data quality and information system domains and their interfaces.
- A data pipeline reference model suitable for any organization with an asset management reporting process.
- A methodology and technology selection process for developing simple tools that can be deployed in almost any corporate technology environment.

Practical contributions include:

- A small script to make dispersed Excel sheets more robust.
- An automation tool for analysing failure descriptions.
- Automated dashboard built in Excel.
- This document, which is suitable for asset managers and can help them understand their data and data processes much better and learn about the complexities and possible solutions of data quality issues.

6.3 Conclusion

This study was undertaken in response to the poor state of data observed in a Southern African diamond mine. Chapter 1 introduced the problem context and motivated that the identified problem was not an isolated observation. The key issues that were identified are the pressure for improved asset performance from global markets and the increasing volume of data. Chapter 1 concluded that asset managers were not sufficiently equipped to respond to these issues. The many systems and standards required today for managing assets optimally, the general incompatibility of these systems and the lack of widely adopted technical standards make addressing data quality issues very difficult. In light of this, the primary research objective was formulated so as to focus just on identifying and assessing data quality issues in asset management decision-making.

To achieve the stated primary research objective, a comprehensive literature review was conducted and documented in Chapter 2. The literature addressed the primary topics relevant to this study: asset management, decision-making, data quality and information systems. Highlights from the literature

review include the long anticipated rise of asset management from its maintenance roots; the dual role of data in decision-making and how decision-makers update their mental models and the inseparability of data and information systems. The data life cycle and the many dimensions of data quality have also been discussed.

Chapter 3 presented the research methodology used in this study and discussed the legal and ethical considerations of such a pragmatic approach to research.

The proposed solution is a framework for identifying and assessing data quality issues in asset management decision making and was documented in Chapter 4. Chapter 5 documented the application of the framework in a Southern African diamond mine which demonstrated the usefulness (which is the measure of value in the pragmatic world view) of the framework. Based on this case study, it can be concluded that the study achieved its primary research objective.

The modular nature of the framework allows future studies to be carried out to integrate the framework with various other disciplines to not only identify data quality issues, but also systematically address them. The hope is that this framework will eventually become part of a larger, pragmatic approach to allow asset managers to implement an ISO 55001 compliant asset management system.

Appendices

Appendix A

Reference Model

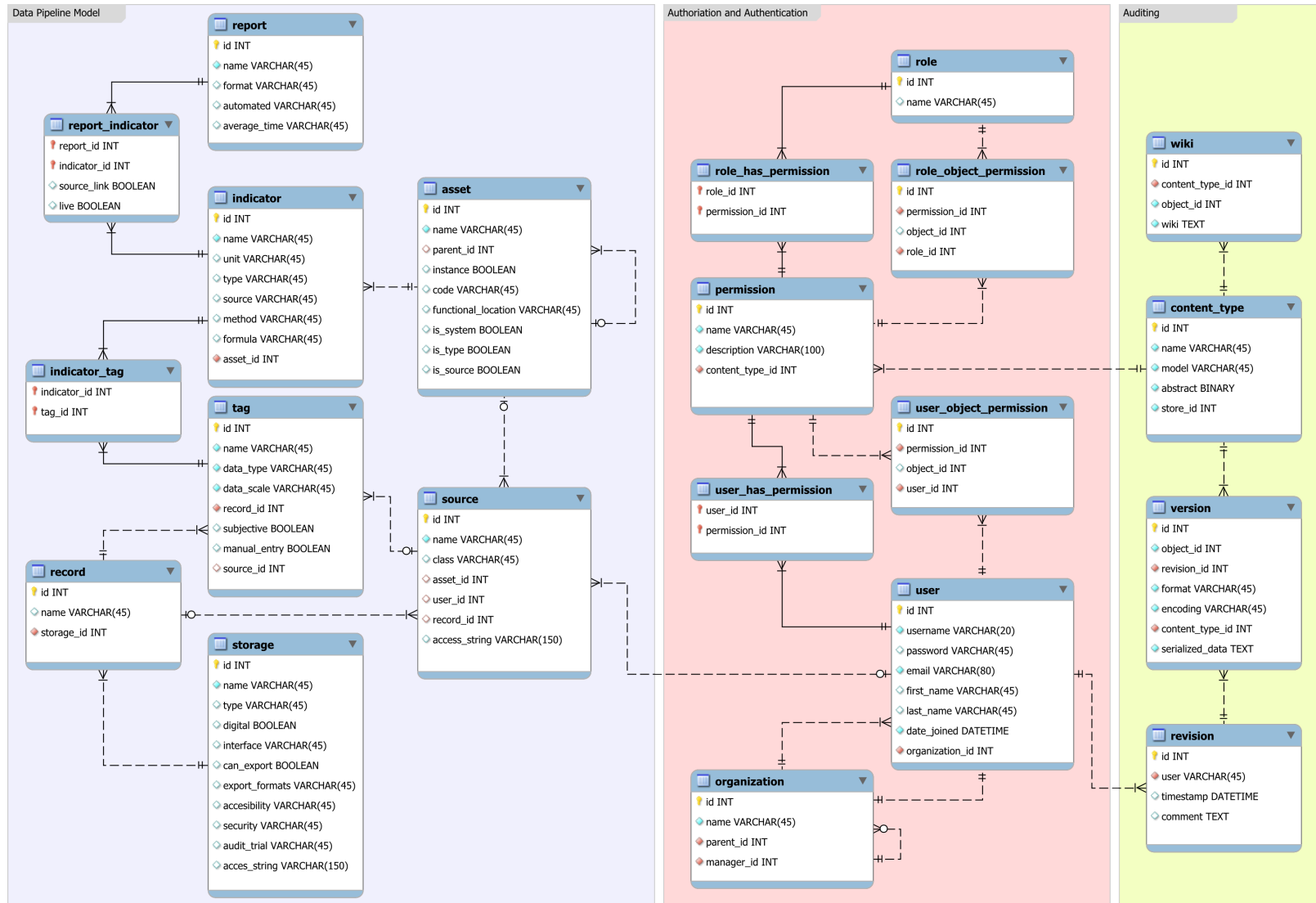


Figure A.1: The complete reference model

Appendix B

Examples of Serialized Data

XML:

```
<person>
  <firstName>Rossouw</firstName>
  <lastName>Minnaar</lastName>
  <academicRecords>
    <academicRecord type="B.Eng Industrial" date="2012-12-09" />
    <academicRecord type="M.Eng Industrial" date="" />
  </academicRecords>
</person>
```

YAML:

```
firstName: Rossouw
lastName: Minnaar
age: 2

academicRecord:
  - id: 1
    type: B.Eng Industrial
    graduation: 2012-12-09
  - id: 2
    type: M.Eng Industrial
    graduation:
```

JSON:

```
{
  "firstName": "Rossouw",
  "lastName": "Minnaar",
  "academicRecord": [
    {
```

```
    "id": 1
    "type": "B.Eng Industrial",
    "graduation": "2012-12-09",
  },
  {
    "id": 2
    "type": "M.Eng Industrial",
    "graduation": "",
  },
],
}
```

CSV:

#FILE: Student

```
id, firstName, lastName
1, Rossouw, Minnaar
```

#FILE: AcademicRecord

```
id, studentID, type, graduation
1, 1, B.Eng Industrial, 2012-12-09
2, 1, M.Eng Industrial,
```

References

- Ackoff, Russell L (1989). "From Data to Wisdom". In: *Journal of applied systems analysis* 15, pp. 3–9.
- Argyris, Chris (1977). "Double loop learning in organizations". In: *Harvard business review* 55.5, pp. 115–125.
- Argyris, Chris and Donald A. Schon (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, Mass: Addison-Wesley. 356 pp.
- Baum, J. and P. J. Vlok (2013). "Mapping primary constraints in physical asset management strategy execution, using social network analysis". In: *South African Journal of Industrial Engineering* 24.2, pp. 47–58.
- Bazerman, Max H. and Don A. Moore (2012). *Judgment in Managerial Decision Making*. 8 edition. New York: Wiley. 256 pp.
- Bird, Katie (2010). *ISO 9001 certifications top one million mark, food safety and information security continue meteoric increase*.
<http://www.iso.org/iso/news.htm?refid=Ref1363> (accessed on 08/18/2013).
- Brown, P. and M. Sondalini (2012). "The Evolution of Maintenance Practices". In: *Asset Maintenance Management - The Path toward Defect Elimination*.
- Bryman, Alan (2006). "Paradigm peace and the implications for quality". In: *International Journal of Social Research Methodology* 9.2, pp. 111–126.
- Burnett, S. and P. J. Vlok (2014). "A simplified numerical decision-making methodology for physical asset management decisions". In: *South African Journal of Industrial Engineering* 25.1, pp. 162–175.
- CEOS WGISS (2011). *Working Group on Information Systems and Services - Data Life Cycle Models and Concepts*.
- Creswell, John W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications. 305 pp.
- DAMA (2014). *About Us - DAMA International*.
<http://www.dama.org/i4a/pages/index.cfm?pageid=3339> (accessed on 08/21/2014).
- David Berger (2010). "Ensure CMMS data is good enough to trust". In: *Plant Services*.
- Debswana (2012). *Report to Stakeholders*. 2012. Gaborone, Botswana: Debswana, p. 60.
- (2013). *About Jwaneng*. Jwaneng.
<http://www.debswana.com/Operations/Pages/Jwaneng.aspx> (accessed on 09/14/2014).

- Fogel, Graham (2013). "Physical Asset Management course presented at University of Stellenbosch". Lecture. Post graduate course in Physical Asset Management. University of Stellenbosch.
- Friedman, Ted (2010). *Findings From Primary Research Study: Data Quality Issues Create Significant Cost, Yet Often Go Unmeasured*.
<https://www.gartner.com/doc/1455915/findings-primary-research-study-data> (accessed on 09/09/2014).
- Friedman, Ted and Smith Michael (2011). *Measuring the Business Value of Data Quality*.
- Gabor, Andrea (1990). *The man who discovered quality: How W. Edwards Deming brought the quality revolution to America*. New York: Random House.
- Gao, Jing and Andy Koronios (2010). "Snap-on Data Quality Enhancement and Verification Tools (DEVA) for Asset Management". In: International Conference of Information Quality. Arkansas, United States: University of Arkansas at Little Rock.
- Gartner (2007). *'Dirty Data' is a Business Problem, Not an IT Problem, Says Gartner*. Sydney, Australia: Gartner.
- (2014a). *Gartner Says Worldwide Business Intelligence and Analytics Software Market Grew 8 Percent in 2013*. Stamford: Gartner.
- (2014b). *Magic Quadrant for Business Intelligence and Analytics Platforms*. G00257740. Gartner.
- Hammond, John S., Ralph L. Keeney, and Howard Raiffa (2002). *Smart Choices: A Practical Guide to Making Better Decisions*. New York: Crown Business. 256 pp.
- Heeks, Richard (2002). "Information systems and developing countries: Failure, success, and local improvisations". In: *The information society* 18.2, pp. 101–112.
- Hevner, Alan (2007). "A Three Cycle View of Design Science Research". In: *Scandinavian Journal of Information Systems* 19.2, pp. 87–92.
- Hevner, Alan R. et al. (2004). "Design Science in Information Systems Research". In: *MIS Quarterly* 28.1, pp. 75–105.
- IEEE (2014). *IEEE Publications & Standards*.
https://www.ieee.org/publications_standards/publications_standards_index.html (accessed on 08/21/2014).
- International Organization for Standardization (1997). *Friendship among equals: recollections from ISO's first fifty years*. Genève, Switzerland: ISO Central Secretariat.
- ISO 55000 (2014). *Asset management - Overview, principles and terminology*. International Organization for Standardization.
- ISO 55001 (2014). *Asset management - Management systems - Requirements*. International Organization for Standardization.
- Johnson, R. B. and A. J. Onwuegbuzie (2004). "Mixed Methods Research: A Research Paradigm Whose Time Has Come". In: *Educational Researcher* 33.7, pp. 14–26.
- Kahneman, Daniel (2013). *Thinking, Fast and Slow*. Reprint edition. New York: Farrar, Straus and Giroux. 512 pp.
- Keetshabe, M Abraham (2012). "Development of ICT Legal and Regulatory Framework in Botswana". General Counsel. Office of the President, Gaborone, Botswana.

- Kendall, Kenneth E. and Julie E. Kendall (2014). *Systems Analysis and Design*. 9th ed. Prentice-Hall. 552 pp.
- Klein, Gary A. (1998). *Sources of power: How people make decisions*. MIT press.
- Kopáčková, Hana and Markéta Škrobáčková (2006). “Decision support systems or business intelligence: what can help in decision making?” In: *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration. 10 (2006)*.
- Lee, Yang W. et al. (2002). “AIMQ: a methodology for information quality assessment”. In: *Information & Management* 40.2, pp. 133–146.
- Lin, Shien, Jing Gao, and Andy Koronios (2006). “A Data Quality Framework for Engineering Asset Management”. In: *Engineering Asset Management*. Ed. by Joseph Mathew et al. London: Springer, pp. 473–482.
- Madnick, Stuart and Hongwei Zhu (2006). “Improving data quality through effective use of data semantics”. In: *Data & Knowledge Engineering*. Including: Sixth ACM International Workshop on Web Information and Data Management Sixth ACM International Workshop on Web Information and Data Management, in conjunction with the 13th International Conference on Information and Knowledge Management (CIKM 2003) 59.2, pp. 460–475.
- Marchant, Gary E, Kenneth W Abbot, and Braden Allenby (2013). *Innovative Governance Models for Emerging Technologies*. Edward Elgar Publishing.
- Marglin, Stephen A and Juliet B Schor (1991). *The golden age of capitalism: reinterpreting the postwar experience*. Oxford University Press.
- Mesel, Terje (2013). “The necessary distinction between methodology and philosophical assumptions in healthcare research”. In: *Scandinavian journal of caring sciences* 27.3, pp. 750–756.
- MIMOSA (2014). *What is MIMOSA?*
<http://www.mimosa.org/?q=node/3> (accessed on 09/08/2014).
- Minnaar, Jan Rossouw, Wilandri Basson, and Pieter-Jan Vlok (2013). “Quantitative methods required for implementing PAS 55 or the ISO 55000 series for asset management”. In: *The South African Journal of Industrial Engineering* 24.3, pp. 98–111.
- Moore, Ron (2011). *A Reliable Plant is a Safe Plant is a Cost-Effective Plant*.
http://www.lce.com/A_Reliable_Plant_is_a_Safe_Plant_is_a_CostEffective_Plant_435-item.html (accessed on 08/19/2013).
- Munroe, Randall (2011). *xkcd: Standards*.
<http://xkcd.com/927/> (accessed on 08/21/2014).
- NRF (2014). *Retail Technology Standards*. Retail Technology Standards.
<https://nrf.com/resources/retail-technology-standards-0> (accessed on 08/21/2014).
- OASIS (2014). *About Us - OASIS*.
<https://www.oasis-open.org/org> (accessed on 08/21/2014).
- OECD (2013). *Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data (2013)*.
- Panel for the review of the European standardization system (2010). *Standardization for a competitive and innovative Europe: a vision for 2020*.

- PAS 55-1 (2008). *Asset Management - Part 1: Specification for the optimized management of physical assets*. The Institute of Asset Management.
- PCA (2014). *About POSC Caesar Association*.
<https://www.posccaesar.org/wiki/PCA/About> (accessed on 09/29/2014).
- Percy, Glen (1980). *The Story of OSHA*. United States of America: Occupational Health and Safety Administration.
- Perdicoulis, Anastássios (2012). “Systems Thinking and Decision Making in Urban and Environmental Planning”. In: *Journal of Regional Science* 52.3, pp. 503–506.
- Pinola, Melanie (2012). “Future-proofing your data: learn the habits that will make your data last for the long haul”. In: *Popular Mechanics* 189.12, pp. 114+.
- Powell, Thomas C. (2001). “Competitive advantage: logical and philosophical considerations”. In: *Strategic Management Journal* 22.9, pp. 875–888.
- Power, Daniel J (2007). “A brief history of decision support systems”. In: *DSSResources* 4, pp. 1–15.
- Rochat, C (2002). “Possible solutions to information overload”. In: *SA Journal of Information Management* 4.2.
- Rokstad, Marius Møller (2012). “An Approach to planning Data Collection for IAM in VMW”. Norway: Norwegian University of Science and Technology. 133 pp.
- Rowley, J. (2007). “The wisdom hierarchy: representations of the DIKW hierarchy”. In: *Journal of Information Science* 33.2, pp. 163–180.
- Rud, Olivia Parr (2009). *Business intelligence success factors: tools for aligning your business in the global economy*. Vol. 18. John Wiley & Sons.
- Saša Baškarada, Jing Gao, and Andy Koronios (2006). “Agile maturity model approach to assessing and enhancing the quality of asset information in engineering asset management information systems”. In: *BIS*. Centre for Integrated Engineering Assets Management, pp. 486–500.
- Satzger, B. et al. (2013). “Winds of Change: From Vendor Lock-In to the Meta Cloud”. In: *IEEE Internet Computing* 17.1, pp. 69–73.
- Simmons, Helen (2009). “Whose Data are they? Ethics in Case Study Research”. In: *Case Study Research in Practice*. London, United Kingdom: SAGE Publications, Ltd.
- Smith, Anthony M. (1993). *Reliability-centered maintenance*. McGraw-Hill New York.
- Stanovich, Keith E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.
- Stevens, S. S. (1946). “On the Theory of Scales of Measurement”. In: *Science* 103.2684, pp. 677–680.
- Supply Chain Council (2014). *SCOR 11.0*.
<https://supply-chain.org/scor/11> (accessed on 08/21/2014).
- The Open Group (2014). *About Us | The Open Group*.
<http://www.opengroup.org/aboutus> (accessed on 09/08/2014).
- Towler, Michael and Sarah Keast (2009). *Rational Decision Making for Managers: An Introduction*. 1 edition. Chichester, England ; Hoboken, NJ: Wiley. 352 pp.
- USGS (2014). *USGS Data Management - Data Lifecycle Overview*. USGS Data Management.
<http://www.usgs.gov/datamanagement/> (accessed on 10/07/2014).

- Velleman, Paul F. and Leland Wilkinson (1993). “Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading”. In: *The American Statistician* 47.1, pp. 65–72.
- Vigon, Bruce W. and Allan A. Jensen (1995). “Life cycle assessment: data quality and databases practitioner survey”. In: *Journal of Cleaner Production* 3.3, pp. 135–141.
- Von Petersdorff, Hagen Alexander and P. J. Vlok (2013). “Identifying and quantifying maintenance improvement opportunities in physical asset management”. Stellenbosch: Stellenbosch University.
- W3C (2014). *About W3C*.
<http://www.w3.org/Consortium/> (accessed on 09/08/2014).
- Wand, Yair and Richard Y. Wang (1996). “Anchoring Data Quality Dimensions in Ontological Foundations”. In: *Commun. ACM* 39.11, pp. 86–95.
- Wilhelm, Alex (2014). *Oregon Sues Oracle*. TechCrunch.
<http://techcrunch.com/2014/08/22/oregon-sues-oracle/> (accessed on 09/18/2014).