The Modern Asset: Big Data and Information Valuation

by

Jacques B. Stander

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

Supervisor: Prof. P.J. Vlok
Co-supervisor: Dr. J.L. Jooste

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.

Date: ....

Abstract

The Modern Asset: Big Data and Information Valuation

J. B. Stander

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

Thesis: M.Eng (Engineering Management)

August 2015

The volatile nature of business requires organizations to fully exploit all of their assets while always trying to gain the competitive edge. One of the key resources for improving efficiency, developing new technology and optimizing processes is data and information; with the arrival of Big Data, this has never been more true. However, even though data and information provide tangible and often indispensable value to organizations, they are not appropriately valued or controlled. This lack of valuation and control is directly related to the lack of a reliable and functional valuation method for them.

This study takes a qualitative and inductive approach to developing Decision Based Valuation (DBV); a proof-of-concept information valuation method. DBV addresses the need to correctly value the data and information an organisation has and may require. Furthermore, DBV is presented with its valuation framework and value optimization and performance assessment tools. These tools address the issue of management and control of information, following in the footsteps of Physical Asset Management (PAM). By using complimentary valuation methods and attributes from PAM in combination with intangible asset valuation methods, DBV is able to capture what is essential to the value of information.

Beginning with a background to Big Data and PAM, their value is made clear to reader. Furthermore, the difficulty and need for a valuation method catered towards information is presented. This will set the stage for the introduction of data and information principles as well as physical and intangible asset valuation methods. These methods are drawn upon for the development of DBV as well as the valuation framework it is based upon. The valuation framework acts as the foundation of DBV and addresses the core principle of information valuation. After detailing DBV in full, proposed value optimization and performance assessment tools are described. These tools are created to assist with the control and management of information. Concluding this study is the validation of both the method itself and the need for it. Combining depth interviews and case studies, the need and importance of a method such as DBV will become clearer to the reader. Furthermore, the success of DBV as a proof-of-concept is illustrated.

The method presented in this study shows that it is possible to create a reliable and generic valuation method for Big Data and information. It sets a foundation for further research and development of the Decision Based Valuation method.

Uittreksel

Die Moderne Bate: Groot Data en Inligting Waardebepaling

("The Modern Asset: Big Data and Information Valuation")

J. B. Stander

Departement Inudstriële Ingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: M.Ing (Ingenieurswese Bestuur)

Augustus 2015

Die wisselvallige aard van die omgewing sake vereis besighede om hulle bates ten volle te benut, maar om terselfdertyd ook ‘n mededingende voordeel te bewerkstellig. Een van die belangrikste hulpbronne om doeltreffendheid te verbeter, nuwe tegnologie te ontwikkel en prosesse te optimeer is data en inligting. Met die koms van die konsep van Groot Data is data an inligting belangriker as tevore. Selfs al verskaf data en inligting tasbare en noodsaaklike waarde vir besighede, word die waarde daarvan nie behoorlik bepaal of beheer nie, wat direk verband hou met die gebrek aan ‘n betroubare en funksionele waardbepalingsmetode vir data en inligting.

Hierdie studie volg ‘n kwalitatiewe benadering en ontwikkel ‘n model vir "Besluit Gebaseerde Waardasie" (BGW) - ‘n konsep inligting waardasiemetode. BGW spreek die behoefte vir korrekte data en inligtingwaarde vir besighede aan. Die metode verskaf die waardasie raamwerk en waarde optimeringsen evaluering van prestasie metodes. Hierdie metodes spreek die probleem van bestuur en beheer van inligting binne die fisiese batebestuur omgewing aan. BGW is in staat om die waarde van inligting te bepaal deur die gebruik van ’n kombinasie van die waardasiemetodes en eienskappe van fisiese batebestuur asook die waardasiemetodes van ontasbare bates.

Om die waarde van beide te verduidelik, word die agtergrond van “Big Data” en fiesiese batebestuur omgewing, waarna die probleme en vraag na ‘n waardasiemetode vir inligting geillustreer word. Dit baan die weg vir die

bekendstelling van data en inligting beginsels asook die fisiese en ontasbare waardasiemetodes. Hierdie metodes dien as fondasie waarop die BGW waardasiemetodes en -raamwerk gebaseer word. Dit dien as die basis vir BGW en spreek die kernbeginsel van die waardasie van inligting aan. Na oorweging van die besonderhede van BGW, word die waarde optimerings en prestasie evalueringsmiddele beskryf. Hierdie middele is geskep om die beheer en bestuur van inligting aan te help. Hierdie studie word afgesluit met die bekragtiging van beide die waardasiemetode asook die behoefte daarvoor. Dit word bewakstellig deur die kombinasie van in diepte onderhoude en gevallestudies. Verder word die sukses van BGW as ‘n waardasiemetode uitgebeeld en bewys.

Die BGW metode bewys dat ‘n betroubare en generiese metode vir die waardasie van “Big Data” en inligting geskep kan word. Dit dien as grondslag vir verdere navorsing en ontwikkeling van die waarde-gebaseerde besluitneming methode.

Acknowledgements

I would like to express my sincere gratitude to the following people for their valuable industry insights they provided.

Riaan Nagel
Manager in Strategy & Operations
PwC South Africa

Brian Williams
Director in Power & Utilities
PwC Middle East

Dean Griffin
Director in Asset Management
Gaussian Engineering

Grahame Fogel
Director in Asset Management
Gaussian Engineering

Dr. Barry van Bergen
Director in Global Strategy Group
KPMG London

Johan Paulsson
Manager Asset Management - Expert Services
Tetrapak Sweden

Ian Beveridge
Asset Management Consultant
Gaussian Engineering

Contents

List of Figures

Nomenclature

Literature Review Variables

FV	Future Value
PV	Present Value
CAC	Contributing Asset Charges
NPV	Net Present Value

Value Variables

V_N	Node Value
V_δ	Average Decision Value Range
V_{max}	Maximum Potential Value of the Decision
V_{min}	Minimum Potential Value of the Decision
Q_f	Quality Factor
V_D	Data Value
V_R	Data Value Ratio
V_p	Processing Value Ratio

Cost and Amortization

C_D	Node's Depreciable Cost
C_{ND}	Node's Non-Depreciable Cost
C_T	Node's Total Cost
N_L	Node's Life Cycle
A_m	Monthly Amortization Amount

Accuracy Variables

I_A	Accuracy Modifier
A_I	Information Accuracy
A_R	Required Accuracy
A_{RO}	Required Accuracy Value
A_F	Accuracy Floor

A_{FO} Percentage Value Lost
A_C Accuracy Ceiling
A_{CO} Percentage Value Gain

Frequency Variables

I_{Fr} Raw Frequency Component
I_F True Frequency Component/Frequency Modifier
F_T Frequency Tolerance
F_N Node Frequency Requirement
F_I Information Frequency

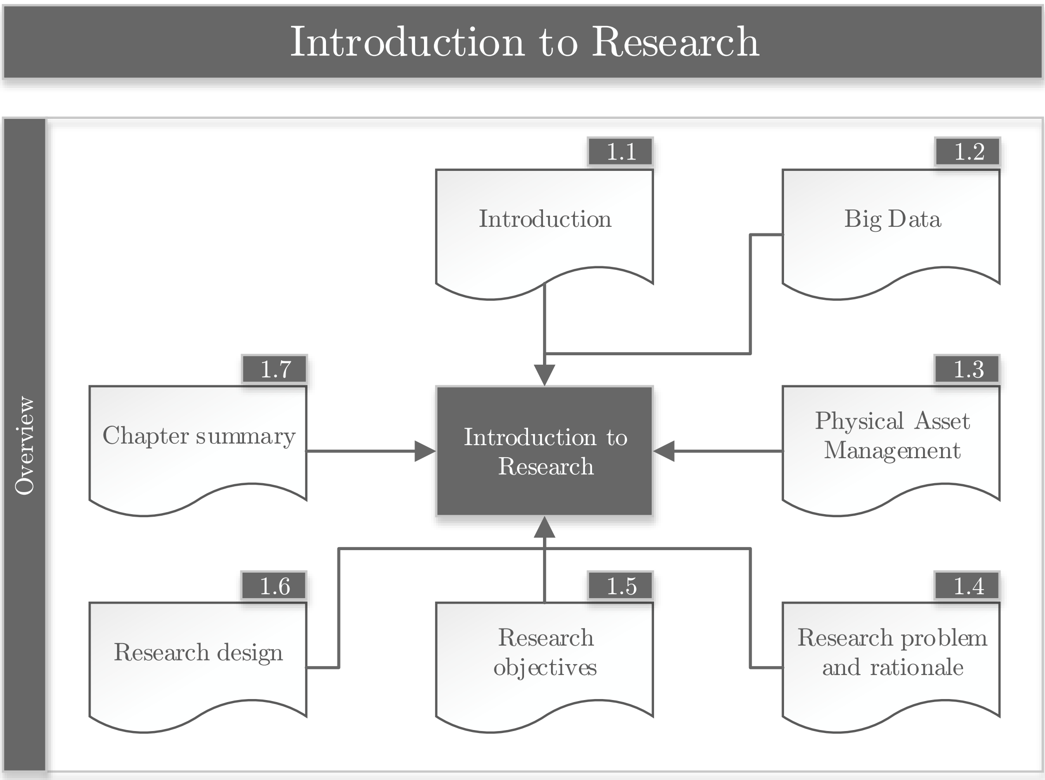
General Acronyms

IAS International Accounting Standards
IFRS International Financial Reporting Standards
GAAP Generally Accepted Accounting Practices
PAM Physical Asset Management
IC Intellectual Capital
IP Intellectual Property
IAS Intangible Asset Register
DBV Decision Based Valuation

Chapter 1

Introduction to Research

Chapter 1 introduces the research topic and gives an overview of Big Data, Physical Asset Management and how these research fields are related to each other. Furthermore, the research problem, rationale, and objectives of this study are provided. The chapter concludes with the research design and methodology where the type of research and the desired outcome of the literature review will be detailed. Lastly, the thesis outline is provided to guide the reader through the study.



1.1 Introduction

This study details the development of a proof-of-concept method that can be used to value Big Data and information. Furthermore, this study provides tools that can be used by organizations to improve their current data and information systems. The aforementioned method and tools are developed to address the growing need in industry to understand the value of data and information, and to make more strategic and informed decisions relating to these resources. This method and toolset will be developed by borrowing aspects of established physical asset valuation and intangible asset valuation methods. The following section provides a background to Big Data and its benefits, as well as the current difficulties organizations face valuing it.

1.2 Big Data

Organizations currently find themselves in an information age, where technology and data play a key roll in both obtaining a competitive advantage (Porter and Millar, 1985) as well as being innovative (Zhu, 2004). Yet, as is the case with most resources, to make effective use of this new technology and data an organization needs to be able to determine the value of information and data. Determining this value is a simple and well defined process for technology, where there is generally an associated performance gain which can be calculated. However, this task is not as simple for data where there is a unclear understanding of the actual value it contributes to an organization's profits. This is an even greater problem with big data systems, due to the sheer volume of data and the difficulty with associating value to it.

The implementation of data and information systems and technology has become more common in industry (Davenport and Short, 2003); with the price in doing so dropping significantly in the past decade. In part this drop in price has been attributed to the reduction in cost of the technology but also due to improved methods such as stated by English (1999). With these systems, organizations have access to previously unobtainable levels of accuracy and variety of information. These new pieces of information derived from big data - and in fact low volume, high quality data - are becoming significant contributors to organizations' income. Yet, these undeniable assets lack the maturity and theory to be utilized to their maximum potential such as their physical assets counter-parts. This lack of maturity means that currently, data and information are unable to be recognized on financial statements as well as them being unable to valued. Both of these issues can be alleviated with a better understanding of how an organization can value their information and big data systems. By being able to reliably determine the value of information assets, organizations will be one step closer to being able to account for them in their balance sheets and other financial statements. Furthermore, if an organization can value its data it is then able to make more effective strategic decisions, especially when those decisions relate to their data and information systems. It is these issues - valuation and accountability - that this study attempts to resolve through the implementation of a new method and toolset.

1.2.1 History of Big Data

In a paper by Coffman and Odlyzko (2002), they argue that Moore's law can also relate to the increase in data traffic each year. In other words; data traffic increases in a way similar to the rate Moore deduced silicon per die would (doubling every two years). It should then be obvious that big data would become a more prominent topic within organizations of all types. One of the first occurrences of big data among professionals was the Very Large Data Base (VLDB) conference in 1975 (Fisher et al., 2012). Here professionals argued that datasets would become too large to handle, with Shneiderman describing big data as a dataset too large to fit on one screen. No matter how it was described back then, it was still apparently obvious at that conference that large datasets would become a thing of the future. Fisher et al. (2012) goes on to state that a problem with big data is that due to its size it cannot fit onto a computer's memory and must be processed from the computer's hard drive instead. This slows down analytics drastically, as memory outperforms hard drives in speed by a significant factor. It should be noted that the ram issue is no longer as big a problem with the development of super fast solid state drives that can replace expensive ram Shah (2013). Borkar et al. (2012) describes big data as being born in enterprises' data warehouses. Where they stored large quantities of their historical business data electronically. This data then needed to be queried for reporting and analysis, subsequently large manufactures and server providers, such as IBM, started catering for these demands. Eventually they created powerful multi-threaded machines to process and deal with these vast quantities of data. Borkar et al. (2012) further states that a significant milestone in the field of Big data occurred in 1986 when Teradata shipped its first parallel database system. Therefore it is safe to assume that the idea of big data started with large enterprises. However, only when the likes of Google and Facebook started giving it attention did it really start to take off.

The field of Big Data is rapidly expanding year on year. Manyika et al. (2011) project that the growth in the global data generated by organizations is 40% per year as opposed to the 5% growth in IT spending globally. They go on to state that in the US, a further 1.5 million data-savvy managers are needed to take advantage of big data. With the recent increase in cloud computing (Agrawal et al., 2011), businesses are able to scale to their data needs more dynamically. Unlike in the past where big data analysis was just left to large enterprises, now Small Medium Micro Enterprises (SMME) and the like are able to take advantage of this new resource. Cloud computing enables enterprises to outsource analysis of their data to another company, leaving just collection to them. This allows SMME's to scale rapidly to new sources of data while still maintaining the ability to process it. Furthermore, expertise in analysis is not needed by these SMME's and therefore adoption of Big Data practices is made easier for them. Bollier and Firestone (2010) attributes the explosion of data to the rapid advancements in: the mobile sector, cloud computing and new technologies. These areas create more data themselves but also enable the processing of larger sets of data. With mobile phones having built-in GPS (Global Positioning Systems), it is possible to geo-tag data which is the ability to attach the location of where the data was collected to the data itself. This gives data another layer of information to investigate and use. Bollier and Firestone state that Google now not only collects what queries get sent their way but also where they originated. This enables them to determine region specific trends such as Flu outbreaks around the world (Pappas, 2014).

Big data, in its most rudimentary form, can be described as a large volume of stored data. However, this definition is too simple to encompass all that is big data. Kaisler et al. (2013) refers to the three V's of big data: Volume, Velocity, and Variety. Volume simply describes the amount of data being stored or processed, velocity describes the speed at which this data is being captured and processed, and finally variety details the different types of data being collected. Kaisler et al.'s three V's gives a better description of big data, yet it is still unable to provide a complete picture of what big data is and how it affects organizations.

This study will not only explore Big Data as a resource, but more importantly as an asset to organizations that can harness it.

1.2.2 Applications and Benefits of Big Data

Big data has an increasingly more important presence in many industries such as marketing and medical. One benefit as mentioned by Michael and Miller (2013) is the ability to stream data from multiple devices, such as mobile phones, which enable companies to target their marketing even better. Brown et al. (2011) mentions many benefits of big data, for instance its ability to augment or completely replace decision makers. Including the ability to simulate possible decisions to arrive at the best one. Big data gives organizations possibilities that were previously never possible, all deriving from masses of data of various types. Brown et al. (2011) also talks about other possibilities that have been made possible through the adoption of big data, such as discovering new business models in previously untapped markets. Jagadish et al. (2014)

describes that big data gains value throughout its lifecycle from acquisition to information extraction all the way to interpretation and deployment. However, Jagadish et al. (2014) say that big data is not without its challenges such as its scale, inconsistency and timeliness. Although, if these challenges can be overcome, then the true value of big data can be harnessed. Another benefit of big data is its ability to solve previously unsolvable problems. Tien (2013) mentions this with the fact that big data solutions are being used to help solve the 14 grand challenges that are required for 10 breakthrough technologies, all of which would lead to the third industrial revolution. These grand challenges would subsequently lead to significant value and benefits to civilization. Cukier and Mayer-Schoenberger (2013) detail how the approach to data has changed since big data practices. One such change is that from seeing causation in data to correlation. Cukier and Mayer-Schoenberger (2013) mention how being able to see correlation can be invaluable, especially to the medical industry where symptoms can be linked to their underlying causes.

1.2.3 Valuing Big Data and Information

The majority of publications and articles available detail what Big Data's value is to an organization, such as health care (Moore et al., 2013). In fact, when reviewing literature it soon becomes apparent that there is no defined method on how to attribute a monetary value to Big Data or information. Articles such as those written by Groves et al. (2013), Katal et al. (2013), Villars et al. (2011), Chen et al. (2012), and Brown et al. (2011) go into great detail on what the benefits of Big Data are and how organizations can harness it and extract value from it. However, as previously mentioned, none of the aforementioned articles produce a method that can be used to attribute a financial value to specific piece of information or data.

The need for such a method is documented and discussed by multiple authors. Maxwell et al. (2015) describe how decision makers need to decide on how to spend limited funds, such as whether to spend capital on direct management of new information. To make this decision however, an organization requires a certain understanding of the return on the investment for this business decision. Sakalaki and Kazi (2007) discuss how information is considered a good of uncertain value while paradoxically producing significant value to those who use it. Sakalaki and Kazi (2008) once again refer to this paradox and how information is undervalued and how material estimations are typically used to value them. This highlights the lack of adequate tools and methods to value information, resulting in the use of methods that are not suited to information and data valuation. Cummins and Bawden (2010) analyse companies and how they attempt to assign monetary values to information, further uncovering the link between performance and the successful use of information assets – once again highlighting organizations attempts to value information. El-Tawy and Abdel-Kader (2013) describes an attempt to recognise information as an financially accountable asset and proposes a three stage asset recognition process. This process is an investigation on how to approach the value of information and data and how they can be related to the well defined concept of an asset. The value of knowing what information is worth is not only a recent development, Gorry and Morton (1971) discuss how organizations gaining perspective in the field of information systems is a powerful means of improving effectiveness. Skyrme (1998) concludes that information professionals need to familiarize themselves with intellectual capital and how it contributes to a firm's value.

Therefore, there is a large collection of literature that discusses and provides methods on how to exploit Big Data and information to produce value for an organization. Yet there is one question that remains unanswered, "What is the financial value of the specific data and information being exploited?". There have been attempts made by organizations to determine this financial value, yet by merely adapting established methods, they have been unable to provide accurate and reliable results. Subsequently, there is a great need for the development of a method that can be used to calculate the monetary value of Big Data and information. If such a method were made available to organizations, they would be able to make the important decisions that they are currently struggling to make as noted by Maxwell et al. (2015). The following section introduces the reader to Physical Asset Management and how this more mature field can assist with data and information valuation.

1.3 Physical Asset Management

This section introduces Physical Asset management (PAM) and its core outcomes, after which the benefits of PAM for intangible assets will be discussed and how key insights of PAM can be transferred to handling intangible assets, specifically information.

1.3.1 Introducing Physical Asset Management

PAM is described by Hastings (2010) and Mitchell et al. (2007) as the management of all aspects of assets throughout their lifecycle so that they can perform their required function. It encompasses the entire spectrum of an asset's life from condition monitoring until maintenance, all of which aim to ensure that the asset produces value for the organization. In the past decade PAM has been refined and developed to encompass all aspects of an asset as shown in the PAS 55 and ISO 55000 standards (ISO, 2014). Woodhouse (1997) describes asset management as the simultaneous care and exploitation of assets. In his definition, asset care refers to the maintenance of the asset and risk avoidance, whereas exploitation is its use to achieve operational objectives.

Physical Asset Management helps ensure that an organization's assets are able to perform to their required specifications through the proper implementation of its practices. This can subsequently save an organization lots of time and effort by avoiding breakdowns, meeting production goals, and improving overall reliability. It also helps define the true cost of an asset, specifically taking into account the cost of the assets through its life cycle (Norman, 1990). Baker (1978) states that life cycle costing involves both fixed and variable costs of assets and ties in strongly with financial accounts and decisions. Subsequently, with the introduction of PAM, both the fixed and variable costs associated with assets can be determined at the start of an asset's life; allowing for enhanced financial and strategic decision making. Therefore, PAM does not only have a connection to the physical maintaining and exploitation of assets but also the financial implications thereof.

In summary, PAM can be seen as the combination of activities that both maintain and exploit assets so that they can produce value for their organization allowing it to meet its organizational objectives.

1.3.2 Intangible Assets and Asset Management

Intangible assets stand to gain from increased control and management as physical assets do with PAM. As is the case with physical assets, Chareonsuk and Chansa-ngavej (2010) explains how intangible assets are linked to business performance. Thus proper management of these intangible assets can directly be linked to the performance of a business. Guthrie (2001) explores the management of intangible assets and how it can influence economic performance however, he goes on to state how there is still a need for more research and development.

The management of physical assets is more mature and the lessons learnt from PAM and its PAS and ISO standards can be used to help develop a better set of management tools for intangible assets. However, these two fields need to be linked together for PAM's maturity to be useful for intangible assets and subsequently the valuation of data and information. This linking or correlation between physical assets and intangible assets/information can be seen in these following areas:

1. They both produce value for an organization;

2. Both have distinct lifecycles;

3. Their costs can be precisely calculated;

4. They need to meet certain requirements in order to produce value;
5. Both require a certain amount of oversight and human interaction; and
6. Both lose value at some point.

Subsequently, the way PAM handles these overlapping areas can be studied to determine if the same principles can be applied to intangible assets. The next section details the research problem addressed in this study and the rationale behind it.

1.4 Research Problem and Rationale

This study will address the problem statement that has been identified in the initial literature review.

Problem Statement:

> *The ability to value data and information has become increasingly sought-after by organizations as a means of increasing profitability and operational expenditure. However, existing valuation methods are considered unreliable and inapt for data and information. Consequently, there is a need to develop a new method that can be used to value these resources.*

The above problem can be seen in section 1.2.3 where it was shown that there is both a lack of, and need for, a financial valuation method for data and information. In addition, many articles address the benefits of data and information yet do not present financial valuation methods that can be used by organizations. The need for such a method is widespread and is closely linked to natural business logic; an organization should not spend more on an asset than it is worth. This statement also holds true for information, where organizations should not be spending valuable and often limited capital on a resource that they are unable to accurately value. Due to this lack of understanding and knowledge, organizations tend to collect too much data, the majority of which is superfluous. Fayyad et al. (1996) refer to the flood of data and the need for methods to mine and analyse it. This flood of data is due to the decrease in cost of collecting and storing data and an increased means of processing it, such as the use of cloud services (Greenberg et al., 2008). This had created a change in perspective; organizations now find themselves in situation where they are collecting data unnecessarily. Without knowing how value is generated in the data and information systems, it is difficult to optimize them. Organizations are therefore running ineffective data and information systems and are ill-equipped to change these systems.

Business logic is easily identified with physical assets and the use of PAM, as discussed in section 1.3, which attempts to care for and exploit physical assets for the benefit of the organization. With PAM, organizations fully understand the lifecycle costs of an asset as well as how much value it is contributing to the organization. Furthermore, organizations are able to identify the operational function of assets and adequately control and manage them to achieve their operational goals. The level of understanding and control offered by PAM could assist data and information valuation if some of its core concepts were adapted for intangible assets.

In order to overcome this lack of understanding and knowledge, a new method needs to be developed. This method needs to give organizations the ability to reliably calculate the value of their data and information as well as allow them to understand how its value is generated. The proposed method would have to be a financial valuation method specifically catered towards valuing Big Data and information.

The benefits of the proposed method would include: (1) being able to identify superfluous data, (2) being able to determine the cost versus value performance of data and information systems, (3) reduce excess and redundant data while improving the quality of valuable data, and (4) allow for more strategic decision making while aiding data and information project approvals. The research objectives to address this problem are presented in the following section.

1.5 Research Objectives

The problem statement can therefore be expressed by the following null hypothesis:

> H_0: *Information cannot be valued because it is not an intangible asset.*

This null hypothesis can be expanded into two hypotheses; H_1 and H_2, the first of which is;

> H_1: *Current methods have failed to determine the value of data and information because they were not specifically created to do so.*

Furthermore, the second hypothesis is;

> H_2: *Information that can be valued, can be regarded as an intangible asset.*

Therefore, the aim of this study is to develop an empirical method that can act as a proof-of-concept for calculating the value of information and Big Data – noting that calculating information's value is the first step to calculating data's value. The development of this valuation method will then result in the first hypothesis not being rejected.

Furthermore, during the development of this method, it will also be shown that information can qualify as a financially accountable intangible asset – resulting in the second hypothesis not being rejected. Lastly, if a new information valuation method can be developed, then the null hypothesis can be rejected. Thus the aims of this study and its objects are as follows.

1. To identify established valuation methods for physical and intangible assets and in doing so:
   a) Determine in what why these methods can be incorporated into the valuation method developed in this study, and
   b) Determine why these methods cannot be used alone to value information.
2. To show that information and data can be valued through the development of a new valuation method through;
   a) Identifying where value is lost and gained with information,
   b) Creating a framework that details how to approach information valuation,
   c) Create a new valuation method that can be used on information, and
   d) Create a method to transfer information value to Big Data.
3. To validate the need and success of this method through case studies and interviews by;
   a) Determining if there is need, yet lack of, a valuation method for data and information,
   b) Determining if organizations value data and information,
   c) Determining if organizations have adequate control of their data and information system's costs, and
   d) Determining whether or not the valuation method is able to value real information and data.
4. To show that there are grounds for information to be financially accountable as intangible assets by;

   a) Determining what criteria information would have to fulfil to be regarded as an intangible asset,
   b) Incorporating those criteria in the development of the valuation method, and
   c) Illustrating that information is an asset to organizations.

5. Improve the management and control of data and information's costs and value through;
   a) Creating a set of tools to assess the performance of an organization's data and information systems according to the valuation method being developed,
   b) Creating a set of tools to help organizations extract and improve the value of their data and information systems.

Success of the above objectives will ensure hypotheses one and two will not be rejected while rejecting the null hypothesis. Moreover, it will provide a foundation for future research while address a need within industry. Subsequently, reference will be made to the above objectives throughout the study to determine if they have been met. Furthermore, these aims and objectives will be discussed when finalizing the study to determine if they have been achieved. The research design and thesis outline is presented in the following section.

1.6 Research Design

This research is an exploratory and qualitative study of which the outcome is the initial development of a valuation method. The development of this method will be done inductively and empirically due to the following reasons:

1. The study aims to generate theories and methods to address the research problem;
2. The generated theories and methods apply to existing fields and data, thus need to be empirically tested; and
3. The field surrounding information valuation is not yet mature enough for deductive research.

Once the initial (first iteration) of the method has been developed, future research can iterate it through quantitative analysis of data and case study research.

1.6.1 Research Methodology

This study was chosen to be qualitative and exploratory for the following reason: established methods have been shown to be inadequate in calculating the value of information. Therefore, the study does not test established methods in an effort to develop a previously untested method which approaches the problem from a different angle. That is not to say that these established methods are not considered; the methods that have potential benefits or uses for the newly developed method will be incorporated to a certain extent. Thus it is important that in the investigation of these established methods, the extent to which they can be implemented for information valuation is detailed.

1.6.2 Research Methods

The following resources are used to identify valuation methods and their applicability to information valuation:

1. Peer-reviewed scholarly articles,
2. Legislation, regulations and standards,
3. Consulting and auditing firm's publications, and
4. Interviews.

An important consideration is the aim of the literature review; to identify valuation methods. This aim results in excluding portions of literature that merely speak about the possible value that organizations stand to gain from using information and Big Data. These types of articles and publications rarely address methods that organizations can use to calculate the value of what they are describing and often resort to stating potential gains. Typical examples of such articles and publications that were investigated in the initial literature survey include: Manyika et al. (2011), Luehrman (1998), LaValle et al. (2013), Mahrt and Scharkow (2013), Bughin et al. (2010), Mayer-Schönberger and Cukier (2013), Cuzzocrea et al. (2011), and Agrawal et al. (2011). The majority of research on focusing on topic of information's value, and even more so Big Data's value, are similar to the aforementioned articles. These offer little assistance in developing a valuation method and were subsequently excluded from review. The fact that the majority of literature does not address or describe valuation techniques further validates the need for the new method.

The use of legislation, regulations and standards is an important part of developing the valuation method. Any method that is developed needs to be able to stand up to the rigour of auditing and a government's financial law if it is to be used for financial accounting. Therefore, these standards and laws are used to help guide the development of the method to ensure that it does in fact stand up to auditing and review. These standards are most evident in the classification of intangible assets and their amortization.

The validation of the proposed method will be done through two means: (1) depth interviews with professionals who deal with assets and information regularly, and (2) case studies where the method is applied and tested. The outcome of the validation section should provide sufficient cause for further development of the method as well as prove whether or not the method was successful as a proof-of-concept.

1.6.3 Scope and Limitations

There are a few limitations to the objectives listed in section 1.5 that define the scope of work.

Limitation 1
Due to the enormity of variety when it comes to data and information, it is infeasible to develop standard formulae throughout this study for each data type. As such, this study will only present an initial set of standard formulae and methods for certain data. Subsequently, future research would have to expand upon these formulae and methods. This limitation particularly refers to section 4.1.

Limitation 2
The method developed in this study is only a first iteration and proof-of-concept. As such it may contain flaws that would need to be addressed before being used by organizations. The implementation of the method in the case studies is not expected to yield repeatable and perfect results; however, the method is still expected to determine the value of information and its data. This limitation particularly refers to sections 4.2, 4.3, 4.4, and 6.2.

Limitation 3
Validating the method through case studies is a time consuming endeavour and as such will be limited to two.. This limitation is brought on by the fact that it requires a significant amount of time to arrange and get permission to use a company's data, especially financial data. Furthermore, it is often required that researcher be on site at the company to obtain this data which further limits the case study possibilities. This limitation particularly refers to section 6.2.

1.6.4 Thesis Outline

The following summaries provide the reader with what to expect from each chapter presented in this study. Furthermore, they provide the reader with a brief understanding of the purpose of each chapter.

Chapter 1: Introduction to Research
Objectives: 3a; 3c; 4b
Chapter 1 gives an introduction to both Big Data and PAM, giving an outline of their benefits and how they can be related to each other. Following these introductions, the research problem, objectives, and design are outlined. The research design will describe the overall methodology, the methods used, and the limitations of this study. Lastly, the thesis outline is provided to guide the reader through the study.

Chapter 2: Literature Review
Objectives: 1a; 1b; 2a; 4a
Chapter 2 presents the literature review, it is conducted with the objective to identify establish methods that are used for the valuation of physical assets, intangible assets, and information. Along with these methods, the financial criteria for the classification of physical and intangible assets is investigated. The literature review concludes with an investigation of the amortization of intangible assets and the cost of data, ending with a summary of the findings.

Chapter 3: Valuation Framework
Objectives: 2a; 2b
Chapter 3 introduces the reader to the start of the solution presented in this study. The valuation framework describes the top-down approach to data and information, detailing how data gains its value and, where its costs lie. The information presented in this chapter is important to understanding how Decision Based Valuation (DBV) works but also contributes to the identification of information properties used in Chapter 4. However, this chapter is presented separately to Chapter 4 because it can be used independently to DBV and applies to any data and information valuation tool.

Chapter 4: Decision Based Valuation
Objectives: 2c; 2d; 4b
Chapter 4 picks up after the valuation framework and introduces DBV. First, the classifications of data will be covered; these classifications are proposed in order to differentiate between how certain data gain and lose value. Following classifications, the core concept behind DBV – Decision Nodes – will be presented. After which, the study presents how to determine the properties of information needed for DBV. The end of this chapter is dedicated to the calculations which are used to determine the various values and costs used in DBV.

Chapter 5: Value Optimization and Performance Assessment
Objectives: 5a; 5b
Chapter 5 concludes the proposed solution, detailing tools that organizations can use to optimize and assess their data and information systems. Once again, this chapter is presented separately to chapter 4 as the majority of tools covered can be used independently of DBV. However, even though these tools can be used independently, they cater towards the understanding of data and information given in Chapter 3 and by DBV in Chapter 4.

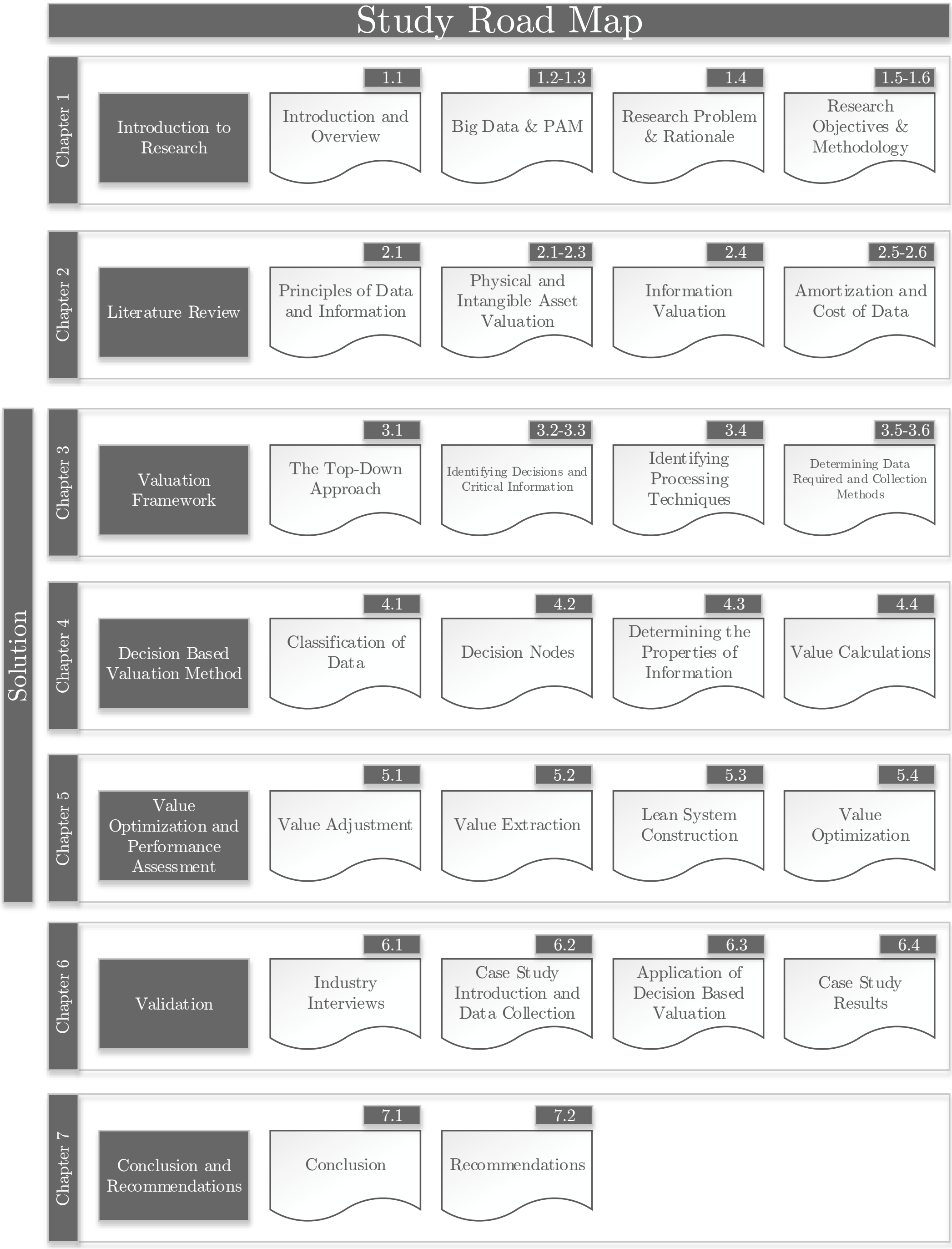
Chapter 6: Validation
Objectives: 3a; 3b; 3c; 3d; 4c;
Chapter 6 presents the validation of both the need and use of DBV through depth interviews and case studies. Statements from industry professionals will be provided with a summary detailing the key points brought up by these statements. Next, the case study choices and data collection will be detailed. Two fully worked case studies are then provided to take the reader through the implementation of DBV. The chapter ends with an analysis of the results and the performance of DBV.

Chapter 7: Conclusion and Recommendations
Objectives: N/A
Chapter 7 concludes the entire study and provides a summary of what was covered and achieved in it. The objectives from section 1.5 are assessed to determine whether or not they were met by the study. Furthermore, recommendations for future research will be provided based on the conclusions from the study and existing deficiencies of DBV.

Study Roadmap

The roadmap (overview) of the study is shown in Figure 1.1. This roadmap should provide the reader with what to expect from each chapter as well as the general flow of the study.

Figure 1.1: Study Roadmap



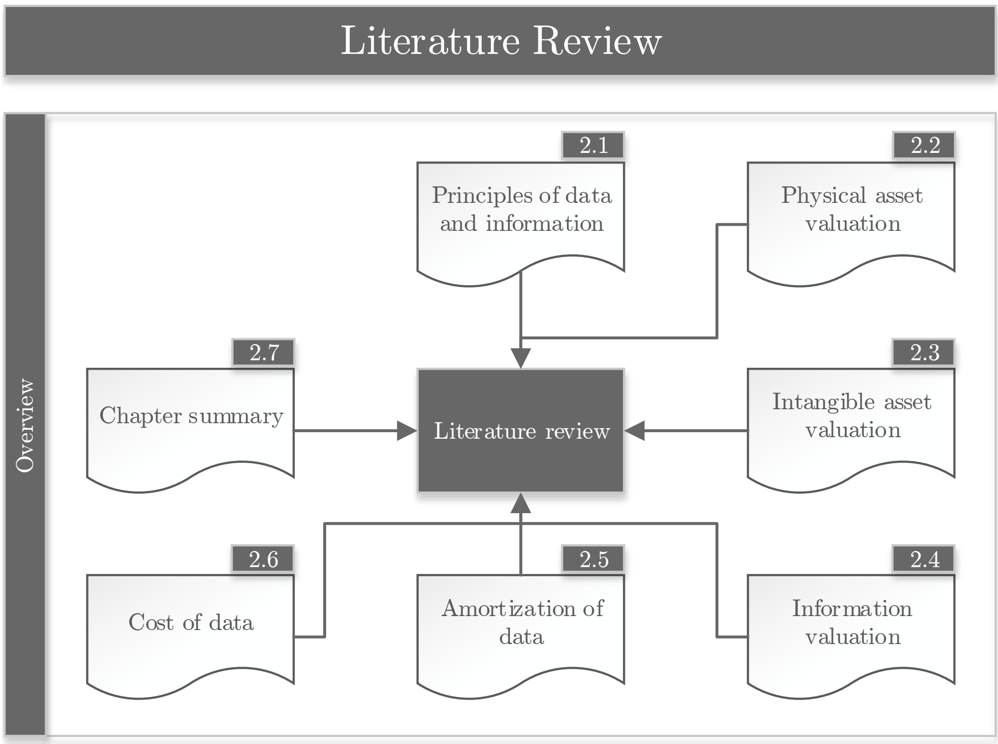
1.7 Chapter Summary

This chapter introduced the difficulty faced by organizations when it comes to valuing data and information, highlighting the need for a new valuation method specifically created for these resources. It is also shown that the development of a new valuation method will require a qualitative and inductive research approach in order to avoid incorporating issues experienced by established methods. Furthermore, this chapter provided a brief description of the limitations of this study together with a summarised outline of the research. The literature review is presented in the following chapter, the aim of which is to explore established valuation methods that can be used for the development of a new method.

Chapter 2

Literature Review

Chapter 2 begins with detailing principles of data and information that are needed to contextualize, and provide understanding to, the information provided in the literature review as well as the study. Following these principles, established methods for the valuation of physical and intangible assets will be reviewed – including the criteria for the classification of both. Furthermore, the theoretical value of information is reviewed to determine what influences it. Ending off this chapter is an investigation of data amortization and its costs and a summary of what insights, and methods will be used for the development of the valuation method in this study.



2.1 Principles of Data and Information

This section includes information that is vital in understanding the methods and concepts presented in this study. It is therefore advised that the principles presented in the section be well understood before continuing with the rest of the method.

2.1.1 What is Data and Information

Data is the raw alphanumeric values obtained from the environment through various acquisition methods. Information, on the other hand, is processed data which has both purpose and meaning. Dretske (1981) states that information is an objective commodity that relates to different events. The thought of information as a commodity is echoed by Eaton and Bawden (1991) who further states that it is a resource to organizations. Israel and Perry (1991) states that, facts carry information and facts are derived from data. Data on the other hand is often seen as a raw resource that needs to be processed before getting value. This is evident by the terminology used, such as data-mining, where both Westphal and Blaxton (1998) and Cabena et al. (1998) use data-mining to describe the process of getting usable data during the early stages of the information age. This terminology is still maintained in the modern era to describe the process of obtaining data as used by Avison and Fitzgerald (2003). Furthermore, Larkin and Simon (1987) and Mayer and Gallini (1990) discuss how processed information, such as diagrams, is worth 10,000 words. This shows how data gains value when processed into a human consumable form.

When comparing data and information, it is natural to use the analogy of the manufacturing industry to show their differences. In such an analogy, data takes on the form of the natural materials, used to manufacture a final product, such as iron ore. Information can be compared to the final product, steel, which has a tangible value and purpose. It will also be apparent that the data (iron ore) still maintains some measure of value before being converted into information (steel). The above analogy conforms to how information and knowledge is viewed, where information is a commodity and data is a raw resource.

2.1.2 Big Data versus Typical Data

The most prevalent difference between the typical data that most organizations and persons deal with and Big Data, is volume (McAfee et al., 2012). The sheer volume of data associated with Big Data causes many difficulties; such as having hardware and software which is unable to handle the volume

(Zikopoulos et al., 2011). This often requires specialized software made specifically to handle Big Data. However, the principles behind Big Data remain the same as well as many of the techniques and strategies. The one requirement is that these data techniques and strategies should be scalable.

Consequently, due to the availability of knowledge, this literature review will be largely based on general data and information. After which the identified methods and strategies will be analysed to determine if they are scalable to the volume demands of Big Data.

Furthermore, it should be noted that even though Big Data has immense volume, the information it creates does not. This concept can be seen through the examples provided by Lohr (2012) where he describes the type of information derived from Big Data. In these examples it can be seen that information needs to consumable by people thus having it in an incomprehensible size makes no sense. Therefore any analysis and processes conducted on Big Data needs to significantly reduce the volume so that its consumable by people and can be regarded as usable information. Subsequently, methods to determine the value of information would be identical for both information obtained from Big Data and typical data, as both sources should produce this consumable information.

Another important consideration is time, specifically for processing Big Data. Demchenko et al. (2013) highlights this time consideration by describing the infrastructure institutions need to process Big Data efficiently. Some methods and techniques may have acceptable operation times when used with typical data however, when scaled for use for Big Data, the operation times might increase exponentially. This fact has resulted in the development of custom software to handle Big Data and process it efficiently (Dittrich and Quiané-Ruiz, 2012). Therefore, methods and techniques should be analysed to determine whether or not they will still be feasible with regards to time, if scaled to Big Data. This becomes an even greater issue when organizations do not have powerful computer hardware to speed up processing or if the method requires significant human analysis. Human analysis refers to tasks which cannot be completed by software and requires the person to conduct the analysis by hand.

2.1.3 The Three V's of Big Data

In the world of Big Data analytics, many tend to focus on the volume aspect of data. However, as Russom et al. (2011) state that there are in fact three key aspects of Big Data which forms its definition. These aspects are called the three V's of Big Data: Volume, Velocity, and Variety. Sagiroglu and Sinanc (2013), Zaslavsky et al. (2013), and Ghazal et al. (2013) mirror this thought by saying that the three V's of Big Data are the cornerstone of Big Data analytics and describe its nature. Moore et al. (2013) also makes use of the three V's to describe Big Data and how they are used to develop strategies and processes for them.

Russom et al. (2011) goes on to give a description of the three V's by giving a few examples of each. In simpler terms; volume describes the size of the data sets in terabytes and storage space. Velocity describes the speed at which the data is collected and processed such as real time streams. Lastly; variety describes the format of the data such as structured or unstructured.

When developing any method or tool for Big Data applications the three V's must always be kept in mind. An effective method will be able to cope with all the aspects of Big Data and not just its volume. This concept is seen in the development of Hadoop - a database software specifically for Big Data - as described by Borthakur (2007).

2.1.4 The Seven Laws of Information

These laws were proposed by Moody and Walsh (1999) as a way to define the nature of information as an asset. These principles or "laws" set out to identify how information behaves as an economic good.

2.1.4.1 First Law: Information is Infinitely Shareable

This law is fundamental to all information; for all information can be copied and duplicated without destroying the original. This trait of information is vastly different to most assets and creates a unique set of possibilities for information. Furthermore, the shared information maintains the same value as the source no matter how many times it has been shared. This is however, contingent on the organization's abilities to realize the information's value. Yu et al. (2001) and Fiala (2005) describe how sharing the same information through a supply chain can benefit each of its members, showing that the same information can be shared multiple times while retaining its value. This fact is true for other industries too, such as credit markets (PAGAON and Jappelli, 1993) and computer security (Gordon et al., 2003).

It should be noted that maintaining multiple duplicates of information within an organization does not increase the information's value, but rather increases its costs.

2.1.4.2 Second Law: The Value of Information Increases With Use

Unlike most assets such as vehicles which depreciate in value the more you use them, information actually increases in value. The core premise behind this behaviour is that information's value is only realized once people use it, and the more they use it the more value they can realize. However, there is a limit to how much value can be extracted from information which depends on the type of information and its area of application.

Moody and Walsh state four prerequisites for the effective use of information:

1. knowing it exists,
2. knowing where it is located,
3. having access to it,
4. knowing how to use it.

It is also possible to view the above prerequisites as barriers to realizing the full value of information within the organization. Therefore, if information is performing sub-optimally within an organization, managers can look at the above four barriers to determine what aspect of their organization needs improvement.

2.1.4.3 Third Law: Information is Perishable

Similar to most physical assets, information depreciates over time and at some point will no longer have value to the organization. The speed at which this depreciation occurs depends on the type of information and its application. Typically, information maintains a short operational useful life. This is largely due to the fact that information is normally gathered and processed for a specific task, but more importantly, for a specific timeframe or window. When this window has passed, the information would have lost most, if not all, of its value. Sale et al. (1997) describes how information can lose its value in relation to the rate of new information being provided. Thus highlighting the perishable nature of information and its link to time but also the frequency of its replacement.

2.1.4.4 Fourth Law: The Value of Information Increases With Accuracy

The general consensus is that more accurate information is more useful and therefore more valuable. Although, in some cases, 100% accuracy is not required, for example; the location of a river for geo-tagging can be off by a couple of meters and still be as useful as one within a few millimetres. Therefore, the accuracy of the information required is highly dependent on the type of information. This concept leads to another trait of information accuracy; each type of information has its own lower and upper bounds where it no longer has value, or no longer increases in value, respectively. These lower and upper bounds can then be extrapolated to form a value versus accuracy curve for each type of information. For example, Burkhauser and Cawley (2008) describe how more accurate obesity measurements can asset the social sciences and enrich its research. Similarly, Poikolainen and Kärkkäinen (1983) discuss how different data collection methods can yield varying degrees of accuracy and how that accuracy affects programs and research.

From a decision making standpoint, it is important to know the accuracy of the information being used. This allows the decision maker to incorporate error margins as well as modify their strategy or approach to the decision based on its risk.

2.1.4.5 Fifth Law: The Value of Information Increases When Combined With Other Information

Moody and Walsh believe that information generally becomes more valuable after is has been compared and combined with other information. The consolidation of information can remove inefficiencies (e.g. duplication) and improve operational use through better understanding and easier access. Furthermore, it can help eliminate errors and inaccuracies through comparing the various bits of information.

Most of the benefits of consolidation and integration can be achieved through the use of standardized templates and processes. Furthermore, identifiers that link different sources of information, and coding schemes, aids in the integration process. It should be noted that achieving 100% integration is neither realistic nor justified, and the Pareto Principle (Juran, 1995) will often yield the most benefits for the least amount of investment. That is, integrating only 20% of the most important information for approximately 80% of the total benefits.

2.1.4.6 Sixth Law: More is Not Necessarily Better

In general practice, the more of a certain resource you have, the better off you are. For example, an organization is better off having more capital and assets versus not. However, this does not always hold true for information. More information often leads to redundancy and an organization's inability to process and handle all the information. This can often lead to an information overload which negatively impacts performance and an organization's ability to realize value from information. However, more information still brings with it more value up until a certain point and thus volume cannot be completely disregarded and seen as a negative aspect of information to be avoided.

Therefore, the volume of information versus its value can be said to follow a negative parabolic curve. Where the value of information increases as its volume does, up until the point of saturation, where after it starts losing value. It should be noted that the information itself does not lose value, rather the value of the information to the organization decreases.

2.1.4.7 Seventh Law: Information is not Depletable

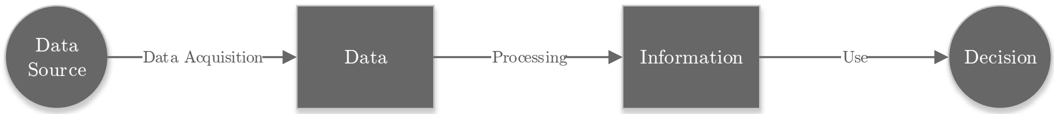
Most resources are depleted over time as you use them - the rate of depletion is dependent on the rate of use. However, information does not get depleted when it is used, in fact using information often creates more information, for example: summarizing data points, creation of presentations and reports, and other derived information. All the while the original information remains intact, leading to information not being depleted and being regarded as an abundant resource.

2.1.5 Data Value Chains and Streams

Value chain is a term often seen in industry to describe where in a product's life does it gain value and where its costs are incurred. Gereffi et al. (2001) speak about the value chains of global business and where companies stand to gain value for their product lines. Similarly, Krajewski et al. (2007) discuss value chains in an operational and processing context and where products get their value from but also where their costs lie. Furthermore, Kaplinsky and Morris (2001) detail the use of value chains and how they are constructed, describing how these value chains capture the value and cost of an item through its lifecycle. Neely and Jarrar (2004) directly applies the concept of value chains to data in an attempt to help organizations extract value from it. Neely and Jarrar (2004) refer to: gather, analyse, interpret, and inform as the core of their data value chain. This interpretation of data's value chain is however not sufficient by itself and can be simplified and extend.

To better understand the method developed in this study, both data value chains, and data streams terminologies are used. Both of these terms refer to similar constructs yet differ slightly. A data value chain describes the distinct stages from data collections to processing and use where value is added and cost is incurred as seen in Figure 2.1. The data value chains in the figures below are a combination of the current value chains implemented in the aforementioned instances. They try to capture the distinct stages where value is added and costs are incurred of information as it progresses from its raw data stage to consumed information.

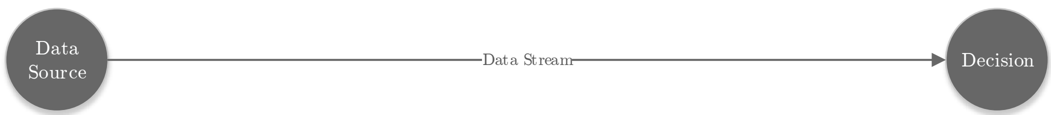
Figure 2.1: Data Value Chain



In Figure 2.1, the value chain starts off with the raw, uncollected data or data source; raw data can be anything from sensor outputs to yet to be captured user opinions. To capture the data from these data sources, an acquisition method needs to be used. This is where the first cost is incurred during the data value chain. Once the data source has been tapped and data has been acquired, it is now stored and considered usable data, at this stage the data will have varying value to an organization and some costs will be incurred for handling and storage. To progress to information, the data first has to be processed and analysed, this requires more investment however it adds significant value to the data. Once the data has been processed into information, it is now in a usable and consumable form, although the information only realises its value during the last stage when it is used for a decision. Therefore, throughout the value chain, data is building potential value which is then realised after it has been used.

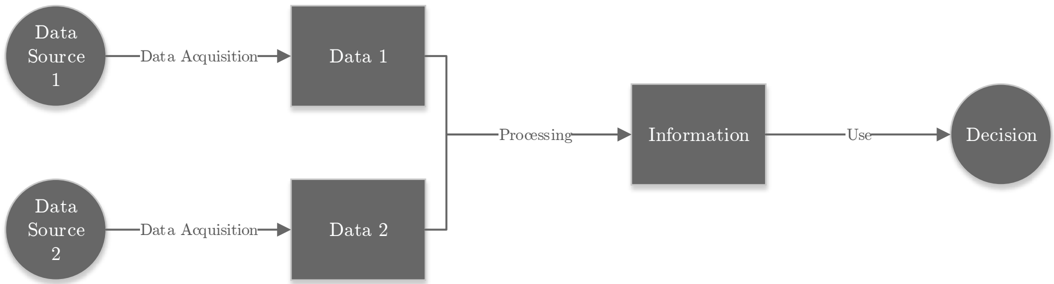
Data streams largely refer to the same thing however, it looks at the data value chain as a single entity without divisions such as processing. Franks (2012) refers to data streams for Big Data, describing them as sources of data coming into a company. This view of data streams is shard by Aggarwal (2007), who describe the use of data streams for data analytics and mining. In fact, as shown in a paper by Gaber et al. (2005), data streams are frequently referred to and used in the field of data mining. Therefore, a data stream is a complete or simplified data value chain for a single source and decision as shown in Figure 2.2.

Figure 2.2: Data Stream



Furthermore, a value chain may contain more than one data stream, shown in Figure 2.3, when information requires data from more than one source to be complete. The concept of clustered data streams is analysed by Aggarwal et al. (2003), referring to them as a combination of multiple single data streams for a single purpose. However, Aggarwal et al. (2003) note that if the too many streams of data arrive at one time, it can have a severe impact on processing performance,

Figure 2.3: Multi-Data Value Chain



In the above example, there is one multi-data value chain and two data streams.

2.1.6 Defining an Asset

The International Organization for Standardization (ISO) describes the fundamental purpose of an asset as, "Assets exist to provide value to the organization and its stakeholders" (ISO, 2014). Bleazard and Khu (2001) and Mitchell et al. (2007) define physical assets as equipment whose functions aligns with an organization's operational goals. Amadi-Echendu et al. (2010) defines engineering assets as equipment whose physical and financial dimensions are linked to their economic value. Hastings (2015) also describe assets as objects that perform certain valuable tasks within an organization. It can then be said that an asset is an object that helps an organization meet its operational goals, subsequently earn a specific economic value.

In financial terms, for an object to be regarded as an asset to a organization, it must in some way generate value for that organization. However, this also opens up the possibility that an object which is considered as an asset to one organization may not be considered an asset to another. This is due to the relativity of the statement of providing value to the organization, not all organizations. Furthermore, there is the case where the object in question may be intangible while still providing value to the organization who owns it.

The International Accounting Standards (IAS) describes an intangible asset as:

> "*Intangible asset: an identifiable non-monetary asset without physical substance. An asset is a resource that is controlled by the entity as a result of past events (for example, purchase or self-creation) and from which future economic benefits (inflows of cash or other assets) are expected.* [IAS 38.8]", (IAS, 2004).

It further states that there are three key attributes of intangible assets, namely:

1. identifiability,
2. control (power to obtain benefits from the asset), and
3. future economic benefits.

These conditions need to be met, in conjunction with those detailed in section 2.3, for an object to be declared an intangible asset according to IAS. However, in practice organizations tend to classify financial assets by their unit prices. Borio and Lowe (2002) describes the relationship between market price and assets, investigating the stability of these prices and how they affect organizations. Taking the classification of assets into consideration, and the fact that asset prices can fluctuate, it is reasonable to assume that certain assets could be declassified as such and visa versa. This shows that the classification of assets is not always once-off and has the potential to be a dynamic classification according to financial terms.

Information on the other hand has never really been able to be classified as an intangible asset. As described by Reilly and Schweihs (1998), information is almost a secondary consideration when valuing intangible assets, and even though those intangible assets could be information, they are not seen as such. Information, in its most basic form, is something which is human consumable and provides insight about one or more topics. In its most basic form, it is difficult to see how information can be regarded as an asset, and in fact this is partially true. However, as with not all equipment being physical assets, nor can all information become intangible assets. Therefore, the question is not whether information can be regarded as an asset, but rather what criteria would specific information have to fulfil to be seen as one? These criteria can be extracted from the definitions of assets, both tangible and intangible, which in its most basic form needs to provide value to an organization who owns it. Consequently, for the method developed in this study to help information become financially accountable, it needs to be able to differentiate between valuable and not valuable information. Furthermore, it needs to describe information in such a way that the common understanding of what an asset is as well as the financial understandings are easily identified.

2.2 Physical Asset Valuation

The Physical Asset Management (PAM) as a field is far more mature than that of intangible asset management. Not only that, but the valuation methods used on physical assets have undergone many iterations and tests and are well defined for their use scenarios. It is not expected that these methods will fit information valuation perfectly however, certain aspects of the methods discussed below can be used to guide the development of this study's method. Furthermore, physical asset valuation can help steer information valuation away from methods that have been proven not to work or require specific circumstances to be used accurately.

2.2.1 Net Present Value

The principle behind the Net Present Value (NPV) is determining the future cash flows of an asset to determine its value. Cash flows refer to the income generated by the asset, typically on a monthly or yearly basis. The NVP method starts by first estimating future cash flows and then determining a discount rate (based of inflation) for those cash flows. The NPV is widely used in industry for investment decisions as referenced by Ross (1995), where the NPV is seen as the quick deciding tool by business school graduates. However, for NVP to be used, a good understanding of the future cash flows of the asset is needed.

The Net Present Value is described in Equation (2.2.1) and documented by Hartman and Schafrick (2004). In the following formula; $-C_0$ is the initial investment, C is the cash flow, r is the discount rate, and lastly T is the time period.

$$\text{NPV} = -C_0 + \sum_{i+1}^{T} \frac{C_i}{(1+r)^i} \tag{2.2.1}$$

There is a flaw with the NPV in that it ignores any flexibility present in real investments, (Myers, 1984), (Pindyck, 1991). Subsequently, Wang (2010) mentions that not only does the NVP method ignore flexibility of investments but, those dynamic features themselves are difficult to impossible to estimate. Thus estimating future cash flows can be problematic and due to the Net Present Value's reliance on cash flows, this can severely impact the accuracy of the valuation.

Applying NPV to information valuation also has its flaws, namely; information rarely produces value in a predictable and regular basis. What is meant by this is that unlike physical assets, it is difficult to determine the future cash flows of information and intangible assets. Moreover, since part of information valuation is to determine what cash flows can be attributed to it, the use of NVP is limited to only after the valuation method has been done.

2.2.2 Decision Trees

Decision trees attempt to model the flexibilities which Net Present Value ignores, (Wang, 2010). In fact, the principle behind decision trees is to be able to model these flexibilities to create a more accurate representation of an asset's value. The decision tree attempts to calculate NPV for the asset for different future outcomes by varying its discount rate, probabilities and cash flows. These options are determined by the practitioner's knowledge of the market and asset. This then leads to a maximum and minimum NPV as opposed to the expected NPV allowing for greater strategic decisions.

However, this method is not without fault; as the investment scenario becomes more complex, the decision tree modelling becomes more complex as well and at a faster rate (Baker and Pound, 1964). This increase in complexity geometrically increases the number of decisions and variables for each future outcome modelled by the decision tree (Trigeorgis, 1996). Furthermore, these variables are difficult to estimate, for instance variables such as market demand are not just high or low but are typically somewhere between. Lastly, the variable discount rate across the decision tree also poses difficulty to the practitioner (Trigeorgis, 1996).

Unfortunately, the issues with NPV and information valuation are still present with decision trees; it is difficult to determine the cash flows of information, especially the probability of them. However, decision trees still act as a useful guide for information valuation. It provides a strategy which can be applied to any single method, highlighting the fact that value isn't constant and changes over time. Therefore, information valuation should be dynamic to a certain extent, where there is a possibility to re-evaluate the information during its lifecycle to determine its true value to an organization. This principle will be incorporated in the development of the valuation method to some extent in an attempt to capture the dynamic nature of value.

2.2.3 Real Options and Lifecycle Costing

Real options modelling gives the practitioner the ability to reassign capital in future options to create a more accurate picture of future scenarios. These options could include such things as downsizing a project or conversely, expanding one. Wang (2010) highlights the fact that real option modelling is a multidisciplinary act. Furthermore, real option modelling layers the investment options of the asset thus splitting up the risks while also producing stages where options are available. These layers or stages can be separated by various parameters associated with the asset, such as net profit or running costs. MacMillan et al. (2006) suggest that the combination of real options and NPV fixes many of its flaws in practice and thus a better valuation can be achieved. The use of NPV in real options is also detailed by Luehrman (1998) and how it affects an organization's investment opportunity. As with the previous two methods, the use of NPV poses difficult for information valuation, and consequently the method cannot be used to its full extent once more.

Real options also sees frequent use in lifecycle costing, where the decision to replace an asset could rely on many factors such as maintenance and operational costs. Cole and Sterner (2000) detail how lifecycle costing attempts to determine the cost of an asset during distinct stages of its life. During these different stages of an asset's life, management would then decide whether or not they should reassign investment or change how the asset is implemented. This directly affect the asset's value to the organization. Gluch and Baumann (2004) discuss the importance of performance metrics when conducting lifecycle costing on assets. The use of performance metrics is one of the main decision tools used in real options, where the goal is to optimize the Return on Investments (ROI) of the organization. ROI is the amount of value generated for the amount of money invested in an asset or venture.

Real option's greatest advantage can also be one of its biggest flaws; the ability to model options requires many assumptions. The difficulty of accurately representing the available options is also problematic. This method also requires an environment where there are options and is steered more towards long term investments, thus the asset itself needs to poses these qualities. Furthermore, how these options are evaluated can change the choices and the “optimal path", such as using Return on Investment. However, there are some attributes of real options that can be incorporated into information valuation and that is the use of performance metrics. Performance metrics help organizations evaluate whether or not an investment in an asset is a sound business decision, one of the most widely used performance metrics is ROI. Subsequently the implementation of ROI in the valuation method would be highly beneficial, especially for future development of the method.

2.2.4 Appraisals

Another option is to determine an asset's value through appraisals that are linked to the current fair market value for the asset. Appraisals are often used during life cycle costing, as noted by Taylor (1981), to determine the best use of an organization's resources. Appraisals are also widely used in different industries such as for forest assets (Xiang-bin, 2007) and even intangible assets (Foster et al., 2003). However, organizations face certain difficultly when there is no current market for the asset or if the market is too small. It often requires a third party to perform the appraisal to avoid bias when used for financial stating such as for balance sheets.

Foster et al. (2003) notes that the need for the appraisal of intangible assets is linked to recently issued accounting standards. This need highlights one of the greatest benefits of appraisals; that it is a true representation of an assets value. The value received from appraisals directly relates to what others are willing to pay for the asset. Insurance claims often rely on appraisals to determine payouts. Thus when its possible, appraisals give a very practical value to an asset that an organization can then use to make strategic decisions with.

Appraisals are not without flaw though as they do not capture the internal value of an asset, that is the value it generates for an organization. Every organization implements and uses assets differently, consequently the assets being used are linked to different cash flows. This results in an asset that could have a substantially lower appraisal value than its internal value. Due to the unique nature of information, any applicable method would have to be able to capture this internal value of an asset.

Clayton et al. (2001) describes another issue with appraisals, the fact that the accuracy of the appraisal is directly linked to which third party does it. An appraisal can be less or more than what the market actually thinks its value is, therefore, special care needs to be taken in who conducts the appraisals. The issue of inaccurate appraisals is echoed by Frentz (2011) who goes on to discuss how the appraisal is linked to the true reporting of certain aspects of the asset being assessed. Subsequently this opens up the method to a certain level of subjectivity, especially by the person conducting the appraisal.

Appraisals are often difficult to do with intangible assets and even more so for information. Without an adequate market to refer to and clear understanding of how value is generated by the asset, it is difficult to provide the asset with an accurate value. There are not many lessons that can be learnt from appraisals for information valuation besides the need for a clear definition of an asset's attributes. Anything that is therefore linked to information's value should be detailed and considered while using the valuation method.

2.2.5 Liquidation Value

The liquidation value of an asset is the value that an organization can obtain from selling it. This value is often determined at the time the asset is being sold, often favouring the shortest route to the sale of the asset. It is most frequently used during the settling of debt as noted by Galai et al. (2007) and Allen and Gale (2000). This method is highly dependent on the current demand of the asset and consequently there is a high variance in the result. Moraux (2002) describes how companies can make use of this ever changing liquidation value to get the most out of their assets, a technique often used in the banking industry. As with appraisals, the liquidation value does not capture its value to the organization and is often well below. Moraux (2002) details the method in which companies can sell their assets at the point when they will get the most value for it. This point is where the internal value is falling behind that of the liquidation value. Therefore, there is a big disconnect between the value of an asset to the organization and the value they are able to liquidate it at the present moment in time. There is also the issue of current markets, without a market that is willing to buy the asset, there can be no liquidation value and its contingent on the asset being able to be sold at that price.

The reoccurring problem with these valuation methods thus presents itself once more; the inability to capture the internal value for an asset, in this case information. This problem once again indicates that the liquidation value is a poor match for use for valuing information. There is also no transferable methodology or strategy from the liquidation value that can be used for information valuation.

2.2.6 Replacement Costs

An internal approach to valuation is determining the replacement cost of the asset. For instance, if an asset were to disappear or break down, what costs would the organization incur to replace it. This method requires the replacement cost to be calculated at the time of replacement, thus the prices may increase or decrease depending when the calculation is done. The calculated cost is inconsistent and acts as a minimum value of an asset to an organization. It should also be noted that the replacement cost does not have to be for the same asset but one that performs the same task/function. Sullivan et al. (2002) note that replacement cost is a valuable tool that can be used to help mitigate maintenance and life cycle costs, this view is also shared by Beichelt (2001). This is because replacement cost works on a function basis and is an internal metric, therefore cheaper but functionally equal assets could be used instead of more expensive ones. Replacement cost thus act as a useful measure of value to an organization but not to a market. Another consideration, as noted by Cheevaprawatdomrong and Smith (2003), is that the replacement of an asset is often linked with a certain performance increase especially when it involves technology. Therefore, the replacement cost as a valuation method can often be inaccurate, since an organization would be valuing the new improved asset and not the replaced one.

The replacement cost of an asset is typically higher than that of its liquidation value. This is often because the replacement cost, as previously mentioned, is for a new and/or improved asset while the liquidation value is that of a second hand asset being sold. Whittington (2008) compares the current value of an asset with its current replacement costs, which is identical to the comparison between liquidation value and replacement costs. He described how these two values differ and how the gap between the two increases over time. This is because as replacement costs increase overtime (due to inflation), the liquidation value of an asset decreases. The once again highlights the dynamic nature of replacement costs, resulting in them often changing.

As was the case with appraisals, there is little that replacement costs can do for the valuation of information. There are cases where an organization can value an intangible asset by how much it costs them to replace it, but this approach is flawed. The main reason being that it is flawed is that there is often no replacement that can be bought or used for the analysis. Note that the replacement cost differs from the cost approach dealt with in section 2.3.2.1.

2.3 Intangible Assets Valuation

In the modern information age, there is a growing need to be able to value intangible assets, such as patents. These methods are a lot less developed than those used for physical assets however, intangible assets share most of the characteristics with that of data and information. There is a strong likelihood that these methods can be adapted to suit data and information used in Big Data environments and organizations in general. El-Tawy and Abdel-Kader (2013) describe a three stage asset recognition process that, for information to be regarded as an asset, it must be:

1. separable in nature,
2. rights based (to control and economic resource), and
3. be a measurable asset.

These recognition criteria are in line with that of International Financial Accounting Board (IFRS) (Board, 2015) and their recognition of intangible assets. Therefore, it has previously been proposed that certain information can be regarded as a financially accountable asset.

2.3.1 Criteria of Intangible Assets

The following criteria for the recognition of intangible assets has been extracted from IFRS. The standard in question is, IAS 38 “Intangible Assets” and has been in effect since is 1 January 2012.

As stated within the standard, an intangible asset will only be recognized if it meets both the definition of an intangible asset as well as its recognition criteria.

An asset is identifiable if it either:

1. is separable; the ability for it to be separated or divided from an entity to be sold or otherwise, or
2. arises from contractual or legal rights regardless whether it is separable or not.

Following these criteria, a distinction is made for intangible assets. Intangible assets are recognized when

1. there are expected future economic benefits that are attributed to the asset for the organization, and
2. the cost of the asset can be measured reliably.

Furthermore, the standard states that the intangible asset shall be measured initially at cost. These costs can be broken into two categories;

1. the purchase price of the asset including import duties as well as purchases taxes, and
2. direct costs attributed to preparing the asset for use.

There are however, separate laws governing the cost of the asset when acquired through the combination of businesses. These laws revolve around fair value for the asset but still require the asset be recognized by the previously mentioned criteria for intangible assets.

2.3.2 Valuation Methods

This section provides the most widely adopted valuation methods for intangible assets as used by industry, according to Anson (2007), Reily and Schweihs (1999) and are described below. Holloway and Reilly (2012) later states that the three core valuation methods for intangible assets are: (1) the cost approach, (2) the market approach, and (3) the income approach. Furthermore, Holloway and Reilly state that in practice a combination of these methods are typically used to value intangible assets. This provides the evaluator with a range of values for the asset, from a minimum to a maximum, thus allowing for a better understanding of the value of the asset as well as improved strategic decision making based on the results.

2.3.2.1 Cost Approach

The premise behind the Cost Approach is that buyers are only willing to pay a specific amount for the asset equivalent to what it would have cost them to develop or obtain. This approach has two distinct methods; Reproduction Cost, and Replacement Cost. "Reproduction Cost" determines the expenditure required to reproduce the exact same asset (often used in patent litigations). Whereas, "Replacement Cost" determines the expenditure required to develop or obtain an asset of similar utility. A important aspect of this method is that expenditures should be determined at the values of the valuation date and not historical values. Furthermore, employee benefits and costs should be calculated based on current practices and not historical values. Overhead and management costs should be calculated pro rata and should be indicative of the true level of involvement in the development process. Lastly, opportunity costs should be added if and only if they were critical in pursuing development or gaining entry to the market. There are some issues with the cost approach.

Lee and Cunningham (2001) expands upon the cost approach to include the cost of not having an intangible asset, in this case goodwill. An intangible asset can be judged not only by the cost incurred by an organization to having it, but also the cost an organization would incur if it were to lose it. Seeing as organizations value intangible assets and information differently, the cost incurred when the asset is removed would also be unique to each organization.

Valuation methods based on the Cost Approach are also suitable in determining the minimum value that should be obtained from an intangible asset. It can be used in litigation and/or when deciding on whether or not an organization should pursue such development itself. However, Holloway and Reilly (2012) mention that some intangible assets are unique and often cannot be replaced, thus making this approach unreliable. Walker and Weber (1984) describes how the Cost Approach is also widely used for make-or-buy decisions. Williamson (1981) refers to is as a transactional cost approach where the emphasis is put on the decision between two options. The make-or-buy decision frequently occurs in practice as noted by Klein (2008) and Leiblein et al. (2002), and occurs equally often with regards to information. Information often sees this make-or-buy decision with the use of consultants who are essentially analysing data and processing it to provide the client with information, especially for Business Analytics consultants.

The Cost Approach is therefore useful for the valuation of information in one main regard, whether or not an organization should invest in obtaining the information itself or if its should be contracted out. Subsequently, the Cost Approach should be implemented to a certain degree in the valuation method to cover this decision. The Cost Approach does not provide any sort of indication of future economic benefit of information and would therefore be best suited for use as a threshold to a yes or no decision within the method. However, as noted by Lee and Cunningham (2001), the Cost Approach can also be used for loss of revenue from not having the asset. The inclusion of loss of revenue in the Cost Approach could give an estimation of the value of information to an organization. As previously mentioned, information's value is unique to an organization, so the best course of action would be to base the method off this relationship.

2.3.2.2 Market Approach

Market value is the price that organizations within that market are willing to pay for the item according to what they deem its worth. Banz (1981) details that he market value of an asset is typically based of what an organization stands to gain if they acquire it. The Market Approach attempts to determine the worth of the item by what people were willing to pay for it, basing its valuation on values derived from previous sales, licensing and transfers of similar assets. Thus, it assumes that the amount others were willing to pay for the asset would be representative of its value. This is however inherently difficult to achieve with intangible assets as they are, in most cases, unique to their organization. Thus it is difficult to find a market and transactions to use as a basis for this method. One of the issues with the market approach is identifying which market it belongs to. Srivastava et al. (1998) highlights this issue in their discussion of market based assets. The issue of identifying a market value for intangible assets is dealt with by Choi et al. (2000), in his research he states that he biggest issue with this approach is accurately identifying future incomes from intangible assets. Gu and Lev (2001) also identifies this issue of accurately predicting the future incomes derived from intangible assets. Without knowing its future incomes, it is very difficult for a market to value an asset. The lack of predictable income is not only an issue for the Income Approach in section 2.3.2.3 but also for the creation of a market for that asset.

However if a market and a list of transactions exist, this approach becomes both intuitive and reliable as it can be seen as a true external reflection of the value of the asset. Certain factors need to be accounted for when determining the equivalent value in comparison to the market counterparts such as geographical layout, payment methods, and time-frame to name but a few. Sullivan Jr and Sullivan Sr (2000) describes a situation where the Market Approach is applicable for intangible assets; for intellectual property. For the case of intellectual property (IP), organizations often stand to gain market share from superior products and subsequently there is a market of information that could lead to improved products. This is easily seen in the lawsuits between Apple and Samsung over intellectual property (Albanesius, 2011) and (Cusumano, 2013) and what they deem they lost in profits from the other using their IP.

Due to the nature of information being unique and how the Market Approach functions, there is very little of this approach that can be used for the creation of a valuation method for data and information. With the issues with identifying and creating a market for the majority of data and information – IP being part of the minority for instance – it would be difficult to create any method for valuation that relies on it.

2.3.2.3 Income Approach

Also called the Income Model as used by Rodov and Leliaert (2002), this approach relies on the estimation of future income streams from the use of the intangible asset. These income streams are then discounted through the use of present value calculations. An important factor in this approach is the separation of value streams derived from the intangible asset and the organization as a whole.

Thus, value of the intangible asset is that of its future earning for the organization, discounted to present value. However, any associated risks with obtaining the full value of the assets income stream need to be factored into the calculation. This approach is widely used due to the availability and accuracy of the information needed, even for intangible assets. However, Andriessen (2004) states that the field of intangible asset accounting is still poorly developed and it still required a lot of standardization. Without this standardization, the accuracy of these predicted incomes can often be poor representations of the true value, thus creating a difference between the intangible assets book value and its actual value. Barney and Barney (2003) apply the income approach for valuing patents as well as other intangible assets by looking at what income the can bring to companies who acquire them. This approach however has one major flaw, it relies almost solely on estimations and forecasts of potential future earnings. Nothing is for certain, and the entrance of a new and improved IP could completely devalue the intangible asset before being used. This is unlike physical assets that perform a function and continue to do so at the level it was valued at even if there are better versions. This fact is described by Poon (2004) and the difference in production capabilities when upgrading technology.

Another way to look at the Income Approach is the excess earnings of the asset as stated by the Canadian Institute of Charted Accountants (2015). This view highlights the fact that the value of the intangible asset is the excess earnings, that is the earnings minus that of any contributing assets. It is obviously required that the earnings of these contributing assets can be determined as well. The generic formula for this approach is shown in Equation (2.3.1). In the following formula; r is the discount rate, t is the expected life, FV is the future value, PV is the present value, CAC's are the contributing asset charges, taxes are the future tax rates, and lastly the tax benefit is the tax amortization benefit.

$$\mathrm{FV} = \mathrm{PV}(r)\sum_{t=0}^{t}(\mathrm{Revenue} - \mathrm{Expenses} - \mathrm{CAC's} - \mathrm{Taxes}) + \mathrm{PV}(r)(\mathrm{Tax\ Benefit}) \tag{2.3.1}$$

The Income Approach is perhaps the most applicable current valuation method for intangible assets. It has some drawbacks such as determining what exactly the contribution of the intangible asset is however, if this can be determined, then this approach gives a good approximation of what the internal value of the intangible asset is. The main principle of this method to include in the development of the new valuation method is the fact that an intangible asset's value lies between the value produced with and without it. Therefore, if the default outcome – without the intangible asset – can be determined as well as if the outcome with the intangible assets can be predicted and later confirmed, then the asset can be valued. Subsequently there needs to be a way to determine the predicted outcome and what influences that value.

2.3.2.4 Relief from Royalty Approach

The premise behind the Relief from Royalty Approach is that an intangible asset can save an organization royalty fees if developed internally. Therefore, to determine the value of the asset, its future earning potential is determined over its usable life then its potential royalty fee is determined. This fee can range from 0.25% of the net profit earned from the asset up to 25% in some instances. The royalty fee is calculated through determining its value to the organization and its contribution to the organization's overall net profit. Once the royalty fee and future earnings from the asset have been calculated, its Relief from Royalty can be determined. Furthermore, it is important to note that future earnings and royalties should take into consideration an organization's ability to realize the asset's value. This includes organizational structure, market share, as well as manufacturing and development proficiencies.

Schiozer and Suslick (2003) details how royalty relief is used in the petrochemical industry between companies and governments. In this case the royalties are linked to government-owned land and have to be paid to the government, highlighting the diverse implementation of Relief from Royalty. Reilly (2008) describes how Relief from Royalty is used for Intellectual Property (IC), namely: patents, trademarks, copyrights, and trade secrets. In his paper, he describes how this method can be used to weigh up the options between development or purchase for these ICs. He further discusses the useful life of these ICs and how that can affect the method. This shows a strong reliance on accurately identifying the useful life of whatever is being evaluated with the Relief from Royalty approach – as it has a significant effect on the outcome of the calculation. King (2002) discusses the value of intangible assets, IC and goodwill, in his paper he described different valuation methods including the Relief from Royalty Approach. However, he notes that this method is only usable in certain instances and often there isn't sufficient information or market to apply the method. The result of which means that this method is only applicable in certain circumstances and cannot be used as a generic valuation method.

In summary, this approach can be seen as a method to determine the maximum an organization should spend on developing an intangible asset if it is available for purchase or lease elsewhere. When there is no purchasing options, this method reduces to that of a cost analysis for the creation of the intangible asset. However, it can also be used to evaluate whether or not an organization should develop an intangible asset currently in the market for sale by the royalties they might receive from it. A generic formula, by the Canadian Institute of Charted Accountants (2015) for the Relief from Royalty approach is shown in Equation (2.3.2). In the Equation (2.3.2); FV is the future value of the asset, PV is the present value, r is the the discount rate, t is the expected life, revenue is the forecast revenue of the asset , and lastly royalty is the royalty rate applicable to the asset.

$$\mathrm{FV} = \mathrm{PV}(r)\sum_{t=0}^{t}(\mathrm{Revenue} \times \mathrm{Royalty}(1 - \mathrm{tax})) \tag{2.3.2}$$

The main take-back from the Relief from Royalty Approach is the significance of the useful lives or lifecycles of the intangible assets being valued. The importance of life cycles is a reoccurring theme in valuation methods and should be incorporated in the method developed in this study.

2.3.2.5 Technology Factor Approach

This approach is only applicable to technology within an organization. However, the idea of technology does not have to be only that of a physical asset. The Technology Factor Approach uses a similar methodology to that of the Relief from Royalty Approach, by determining the asset's value through its contribution to the organization's market value/net profit. Another similarity is determining an organization's ability to realize this value from the technology. When using the Technology Factor Approach, all aspects of the lifecycle of the technology needs to be accounted for, such as: capital required to make use of technology, the size of the potential market and sales margins. These factors, as well as many others should then be weighed and scored according to their importance and impact on the earning potential of the technology. Once this has been done, the upper and lower bounds of the earning potential and value of the technology to the organizations can be determined. These bounds can then be seen as the maximum potential of the technology for the organization (upper bound) and its minimum value to the organization (lower bound). The lower bound should always be greater than the cost of acquiring/developing the technology to the organization for it to be economically viable.

As previously stated, technology does not have to refer to just physical assets, and is often linked to intangible assets too. The development of technology often arrives from IP and patents and other intangible assets. Arrow (2002) echoes this statement by detailing how intangible assets such as Intellectual Property are linked to technology growth yet organizations fail to perceive and control them as financial assets. Leitner (2005) investigates how intangible assets influence Research Technology Organizations (RTO) where the majority of investment is for Research and Development (R&D). R&D is typically comprised of a significant amount of IP and other intangible assets. Lefebvre et al. (1996) describes how intangible assets are at the centre of Advance Manufacturing Technology (AMT). It is evident that technology and intangible assets go hand in hand, giving credence to the use of the Technology Factor Approach. However, as is shown with the provided examples, the use case for this approach is limited to those organizations which rely heavily on the development and use of technology. This excludes many other organizations who may use intangible assets for other purposes, such as brand loyalty or trademarks.

As was the case with the Relief from Royalty Approach, there is little takeback from the Technology Factor Approach. It does however reaffirm the need to incorporate the lifecycle of the asset being valued and to include any and all costs related to the asset. Both of which will be included in the development of this study's information and data valuation method.

2.3.2.6 Real Options

Similar to the approach for tangible assets, real options can be applied to intangible assets using the aforementioned methods while modelling the various scenarios and outcomes. The premise behind real options is the ability to change the operation and use of an asset at some point in time due to a certain trigger event. A trigger event can be the hours of run-time for physical assets or net profit for intangibles. This allows for a dynamic evaluation of the asset as well as generating different worst and best case scenarios. Sudarsanam et al. (2006) describes how real options have been used to assess intellectual capital and how it has helped a firm's growth. Similarly, Bloom and Van Reenen (2002) discusses its use on patents and how it affects a firms performance and how it can be used to value innovation. However, the use of real options on intangible assets is typically limited to those that can be sold to other organizations.

The implementation of real options for intangible assets does not differ from its use with tangible assets, as such the inclusion of performance metrics in the developed valuation method remains as do the conclusions of the method.

2.3.3 Intangible Asset Valuation Obstacles

There are a few obstacles that make it difficult to value intangible assets with current valuation methods. These obstacles also pose a challenge to any new valuation method and should be considered during the development of the method in this study. These obstacles are: (1) the market value of intangible assets, (2) the terminal value of intangible assets, (3) the separation of income contributions from intangible assets, and lastly (4) the replication of valuation results. These four obstacles with be discussed in more detail in the following sections.

2.3.3.1 Market Value

Market value can be described as the value other organizations are willing to pay for information from the holding organization (Moody and Walsh, 1999). However, very little of the information used and collected has a market value as most of it is unique to that organization. Furthermore, Moody and Walsh (1999) states that information cannot be bought and sold like regular assets and unlike regular assets; information is still retained by the original organization after sale.

Having an inability to define a market value for certain information hinders its accountability. Higson and Waltho (2009) explains a paradox that exists between the competitive nature of information and its uniqueness to its organization. Stating that, if information brings a competitive advantage to an organization, it is usually unique to that organization and therefore it is difficult to measure its market value. Higson and Waltho (2009) summarises this paradox as follows;

> "*So the great paradox of intangibles is that what makes them hard to account for in conventional terms is the direct result of what makes them valuable, which is their uniqueness.*"

Not being able to accurately determine the market value of intangible assets prevents the use of many valuation methods such as appraisals and the similarly named market approach. Both of which give reliable estimates of the value of intangible assets when an active market is present. Hall et al. (2007) analyse the difference between the private value of R&D and patents in comparison to the market value for the same. In their analysis they also note the difficulty in finding a market value and the difference between the market and private values. Once again this shows the difficulty of identifying a market value for intangible assets. Bosworth and Rogers (2001) illustrate that with competition and development of IP to achieve similar goals, a market can be formed for intangible assets. This provides evidence that markets can eventually form over time, therefore the methods that currently rely on active markets could become viable in the future.

2.3.3.2 Terminal Value

Also referred to as an asset's retirement value (Alexander and Hiner, 2001), or its residual value as noted by Whitworth (2003), the terminal value of an asset is the value it retains at the end of its useful life. However, the terminal value of intangible assets isn't as one-dimensional as that of their physical counterparts. As noted in section 2.1.4.7, information is not depletable, therefore it can be shared and used by many organizations without degrading it as an asset. The only change is that the value this information hold towards an organization differs from one to the next. The fact that an organization can get all the value out of information then sell it afterwards to the next, opens up a new dynamic relationship for information. Even though the information no longer holds much or any value to an organization after being used, it may still hold its full value to the next organization who has yet to use it. A simple example would be information for optimizing a program or removing a security flaw. This results in information being able to have multiple terminal values and those terminal values being dependent on the industry and market (if any).

Due to the nature of terminal value, it is a important factor in the valuation of intangible assets, King (2002) uses terminal value as a key step in the three valuation methods he describes. However, it is often not that simple to determine what the terminal value of intangible assets will be, especially those such as information. The unique nature of information means that it is often produced for a certain purpose or decision. However, this decision may change and no longer require the same information, rendering what was produced moot. Unlike physical assets which can be re-purposed or resold with relative ease, information is normally locked to a specific organization's needs. Subsequently, an accurate terminal value becomes paramount for the effective valuation of intangible assets. This concept of terminal value needs to be adapted to suit the needs of information.

2.3.3.3 Separation of Intangibles

The separation of intangible assets is an important step in the valuation of them. Steenkamp and Kashyap (2010) comments that out of the survey they conducted, managers reported the most difficulty found was in separation of these intangible assets into different components. Furthermore, he states that the process of separating intangible assets is problematic, commenting on their entangled nature. Gröjer (2001) highlights the importance of separation for the accurate reporting of intangible expenses and assets. Steenkamp and Kashyap (2010) note both the benefit and difficulty in separating intangible assets and Intellectual Capital (IC) and refers to it by; "Problematic issues of framing and separating IC". Kanodia et al. (2004) describes how separating intangible assets is important to determining economic trade-offs and investments, referring to the fuzzy nature of operating expenditures and investments in intangible assets.

Not only is the separation of intangible assets an important step in determining their value, it is also a mandatory one imposed by the IFRS Board (2015). An inability to separate the income derived from a collection of intangible assets can hinder their acceptance as financial assets. Qi et al. (2006) also states the need to be able to separate the value of assets for corporate understanding and decision making. Therefore, being able to separate intangible assets is one of the first obstacles that needs to be overcome to have a better understanding, and value, of them.

2.3.3.4 Replication of Results

Another key aspect for valuing intangible assets and assets in general is the ability to independently replicate the results of any cost/value calculation by a third party. This is to ensure the accuracy of what is reported as well as benefits to the organization's strategic planning through accurate evaluations. Power (2003) and Dittenhofer (2001) reference the need for auditors to be able to replicate the financial calculation and declarations documented by organizations. Dando and Swift (2003) state that there needs to be a certain level of transparency in organization's financial accounting methods so that they can be tracked to see if they meet financial regulations. However, not being able to replicate financial calculations do not only impact accuracy but can also cause heavy financial penalties. Carcello and Nagy (2004) discuss fraudulent financial reporting and how being unable to replicate results can lead to further investigation and often ending with financial penalties.

The problem with replication of results for intangible assets arises when the value of the asset is unique to the organization leading to difficulty in third party evaluations. As previously mentioned, information and data's value is inherently unique to its organization. Therefore, the valuation done relies on the honesty and accuracy of the organization whose data it is. To allow for replication of results, the organization in question would be required to permit a lot of transparency in their calculations as well as how they valued it. The methods and assumptions made would need to be detailed for inspection. Any method developed thus needs to account for this by having both transparent and standard calculations and assumptions, as well as an easy-to-follow valuation method.

2.4 Information Valuation

Economist Gould (1974) defined the value of information in terms of risk avoidance for the decision maker. Gould further states that some information has no value and that information's true value is only ascertained by how it will be used.

Gould's expression for the value of information is as follows. Suppose that $s_1, s_2...s_k$ are all possible states within the world and the decision maker's prior knowledge of these states is listed as probabilities $p_1, p_2...p_k$ where $p_i = Pr(s_i), i \in k$. Let t be the decision maker's chosen variable, if the world is in state s_i then the payoff function for that state, considering $t_i = f(p_i, \alpha)$ (α is a set of decision metrics), is $H(t|s_i)$. Now suppose that decision maker knew that the world was in state s_i and would thus choose $t = t^*$, where

$$H(t_i^*|s_i) = \max H(t|s_i). \tag{2.4.1}$$

However, if the world's state is unknown, the decision maker would choose $t = t'$ which is its best prediction for t on average. That is to say,

$$\sum_{i=1}^{k} p_i H(t'|s_i) = max \sum_{i=1}^{k} p_i H(t|s_i). \tag{2.4.2}$$

Thus, Gould's formula for the value of information is the difference between the outcome with perfect information (t^*) about state s_i and the best prediction on average, namely

$$\sum_{i-1}^{k} p_i H(t_i^*|s_i) - \sum_{i-1}^{k} p_i H(t'|s_i). \tag{2.4.3}$$

Implication

Due to the fact that Equation (2.4.3) is non-negative, information obtained at no cost never makes the decision maker worse off. Furthermore, there is a strong correlation between probability of the state and the payoff function.

Similarly to Gould, Stephens (1989) adopts the view that the value of information lies in its ability to reduce variance/deviance of the decision maker's decision and the optimal decision.

Contemporaries of Gould, Shannon and Weaver (1949), state that the value of information increases as the number of equally likely outcomes increases. This directly relates to the Shannon measure which is described thus. If n is the number of mutual exclusive states (read "future states") of the world with the respective probabilities $\pi_1, \pi_2, ..., \pi_n$, then information that says that state i will be obtained with certainty has the value

$$\log(\frac{1}{\pi_i}) \tag{2.4.4}$$

The above equation illustrates that the less likely the state, the more valuable the information of the certainty of the state is. The expected information is then

$$\sum_{i=1}^{n} \pi_i \log \frac{1}{\pi_i} = -\sum_{i=1}^{n} \pi_i \log \pi_i \tag{2.4.5}$$

It is easily shown that Equation (2.4.5) has the value of 0 when the probability $\pi_j = 1$ for some j, which intern means $\pi_i = 0 for i \neq j$. Furthermore, it can be shown that (2.4.5) is maximized when $\pi_i = 1/n$ thus reducing to Equation (2.4.6).

$$\sum_{i=1}^{n} \frac{1}{n} \log n = \log n \tag{2.4.6}$$

Implication

The above formula highlights Shannon's conclusion that an increase in equally likely states increases the value of the information as previously stated.

Higson and Waltho (2009) assert that the value of information, as an asset, is the difference between the value of the organization with and without the information. This method of valuation is termed the *Economic Value* (EV) and the value added by the information is the difference between its cost and its economic value.

2.5 Amortization of Data

It is generally accepted that depreciation refers to physical assets while amortization refers to intangible assets. However, these terms can be used interchangeably and represent that same process, although for the purpose of this study, the convention will be kept.

2.5.1 Intangible Assets

Knowing how an organization's assets lose their value is often just as important as knowing how they generate value. Thus being able to determine the amortization rate of intangible assets is a vital tool for organizations of all types. This sentiment is matched by Høegh-Krohn and Knivsflå (2000) who state that to properly match future benefits, intangible assets need to be capitalized then amortized over their useful lives. By doing so, the value relevance and informativeness of financial statements and reports are improved, irrespective of their type. Penman (2009) however affirms that even when historical costs are identifiable, the amortization schedules of intangible assets is typically quite speculative. This creates a problem where there are significant benefits to be gained from amortization, yet great difficulty in determining how to accurately amortize. In addition, Penman (2009) highlights the fact that reporting on fuzzy and speculative numbers can damage the informativeness of financial reports. This would also reduce their strategic value to organizations.

Schenk (1966) states that specific criterion for the depression of intangible assists has not been given to tax payers by governments or courts. Schenk (1966) gives an example of how intangibles fit the mould of amortization where patents and copyrights for instance are only valid for a certain period of time after which they lose their value. These forms of intangible asset fit the mould of amortization ideally. However, to amortize intangible assets, the organization is required to prove that it has a determinable useful life and that the intangible assets exhaust or undergo some form of wear. This can be seen as one of the key factors in determining if and how an intangible asset should be amortized.

Choi et al. (2000) suggests that the amortization be based off the assessed uncertainty of the value and timing of future benefits derived from each intangible asset. This is a completely different approach to amortization. Choi et al. (2000) reasons that this method would be better suited to financial reporting as markets insignificantly regard amortization expenses and that amortization expenses of intangible assets are not significantly related to stock returns of organizations. This lack of correlation adds more uncertainty to financial reports.

Being able to determine an amortization schedule for an intangible asset is a vital aspect for it to be financially accountable and represented on financial statements. This means that part of making information financially accountable is being able to amortize it reliably and according to Generally Accepted Accounting Practices (GAAP). This requires an understanding of the various GAAP depreciation methods available to tax payers.

The three most widely used and accepted depreciation methods will be investigated to evaluate their applicability. Wakeman (1980) affirms that accelerated methods – Sum of Years Digits and Double Declining Balance – dominate the straight line method in terms of positive discount rate. In fact, in most cases it is generally accepted that accelerated methods theoretically outperform the straight line method. However, while Berg et al. (2001) accepts that theoretically, accelerated methods are superior, there are cases where straight line depreciation is favourable. According to Berg et al. (2001) the straight line method favours scenarios where there is a stable and growing future cash-flow as well as where there are high probabilities of negative reported income. This means that it could be the favourable method with the uncertainty behind income derived from information.

2.5.2 Information Technology

Technology and computer assets are the backbone to information systems and even more so to Big Data systems. These assets form a large portion of the costs associated with data and information systems and thus should be represented on financial statements. Furthermore, their costs should be depreciated as is the case with typical assets.

Tam (1998) mentions the fact that computer assets have a steeper than average depreciation curve in comparison to typical assets. Furthermore, according to Bott (2000), software both purchased and developed internally can and should be regarded as intangible assets and depreciated accordingly. Bott (2000) further states that the useful life of software is often short, using two years in his example. Together with the steep depreciation of technology assets, the data and information systems will have a rapid depreciation schedule. This view is mirrored by Antonopoulos and Sakellaris (2011) where he argues that even though the physical deterioration of computers is low, the economic depreciation is massive. He also mentions that the rate of depreciation is often linked to the release of newer technology and the performance leap of the newer generation. He highlights this fact by showing that there is statistical significance in processor speed where it is always positive in regression analysis. Stating that for personal computers, the CPU speed has the largest influence on pricing among all other characteristics. Therefore, the rate of technology progress influences the rate of depreciation of technology and computer assets.

2.6 Cost of Data

In their paper, Briggs and Gray (2000) makes an opening comment that in a cost based analysis; if the benefits of a project exceeds its costs then it should be implemented. This is indicative of most business decisions which often hinge on whether or not the benefits of certain projects outweigh their costs. Thus it is important to understand the costs of data. Deelman et al. (2008) states that the costs of doing science in the cloud (off location central storage and processing) depends on the compute, storage, and communication resource demands. This is true for all data and information systems, where their costs are linked to the demands of the data and information, mainly volume and velocity. That is, how fast the data needs to be communicated and processed as well as how much of it needs to be stored.

Jydstrup and Gross (1966) highlights an important fact when it comes to attributing costs to information, which is particularly significant when relating to data value chains. He mentions that there is a lot of downtime, especially for labour, where salaries are being paid for unproductive work. For instance, even though a data processing task should take three man hours, it may take five instead due to unproductive activities. Furthermore, he states that general overheads should also be included such as security as well as general overheads such as space, lighting, and electricity. Jydstrup and Gross (1966) further states that down the line there are more costs associated with information handling, those being informal communication between employees and clients. For example, if an employee takes an extra hour to explain information to another employee, that is an extra man hour associated with that data.

Bekenstein (1981) highlights another cost associated with data; the energy costs of information transfer. Bekenstein goes into great mathematical detail determining the energy cost of data according to bit rate, however this level of detail would be unnecessary for most organizations. In saying that, the principle still holds, there is a tangible and sometimes noticeable energy cost of data transfer that needs to be accounted for. This is especially true for large networks with lots of data transfers happening. These costs will often appear in the utility costs of the hardware such as the electricity consumption of the processing computer. However, things such as WiFi routers and stations are often plugged into grids around the organization and are often neglected. Therefore, it is important to trace the entire flow of data to determine its true energy costs. As mentioned by Jydstrup and Gross (1966), it is also important to determine the operation time of these devices to calculate how much of the hardware energy consumption is associated with a certain data stream.

Another important cost is software as mentioned by Sklavos (2010) in his analysis of Crypto-Processor architecture. Software can often cost more than the hardware it runs on, especially on smaller systems, and thus cannot be neglected in cost calculations.

2.7 Chapter Summary

This chapter aimed to meet the objectives: 1a; 1b; 2a; and 4a as described in section 1.5. These objectives centred around the following aims: *To identify current valuation methods for physical and intangible assets*, *Identifying where value is lost and gained with information*, and *To show that there are grounds for information to be financially accountable as intangible assets*. The literature review was successful in meeting these objectives, a summary of which is provided below.

The overall theme of the literature review is; that accounting and valuation principles for tangible assets is far more mature than that of intangible assets. This accounting immaturity is even worse for information, with poor understanding on how to reliably calculate its value. Methods such as the Market Approach (section 2.3.3.1), Appraisals (section 2.2.4), and Liquidation Value (section 2.2.5) all require an existing market to function effectively. This is a major flaw with information as noted multiple times throughout this chapter. Moreover, it difficult for organizations to have a similar value for information as its value is unique to each. Thus it is also difficult to form a market for information where there can be a current price to evaluate a company's information against.

Methods that rely on an organization's future incomes such as Net Present Value, Real Options, and Income Approach require an organization to accurately predict and or know the contribution of value derived from their assets. This can be done reliably for tangible assets, less reliably for intangible assets, and becomes increasingly difficult for information. However, as an overall approach, determining the value of information through a standard method for its organization has the most potential. This takes into consideration the unique nature of information's value. For such an approach to be usable for accounting purposes, it would have to be standardised and the same result should be obtainable by third parties – if they are evaluating the information from the same organization.

There are set criteria for determining if an intangible object is in fact an asset, these criteria are set by IFRS. If these criteria are met, then there are grounds for information being handled and accounted for as an intangible asset. Furthermore, information has been seen to share similar attributes to assets such as: a finite lifecycle, a defined function and purpose, reliably and calculable costs, future incomes, and require certain upkeep and human interaction – these are seen in sections 2.1.4, 2.6, and 2.5. Subsequently, the idea of using Physical Asset Management techniques are valid to the extent that they relate to these attributes. This also shows that information should more often be treated as an asset and not just an expense.

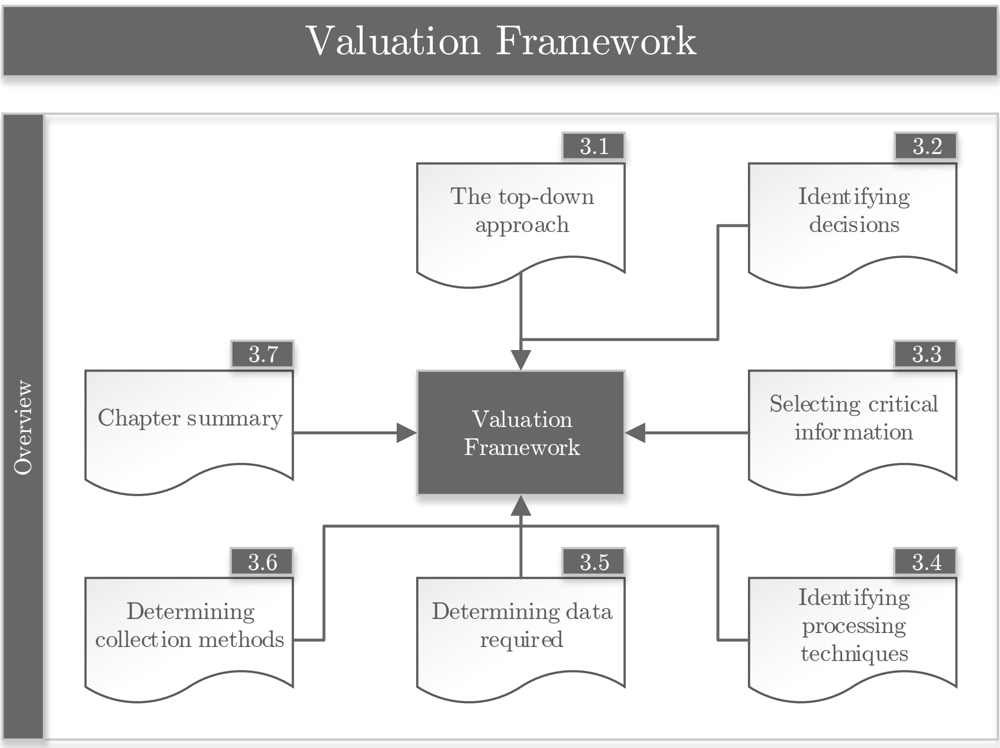
It was illustrated that the value of information is also dependent on the organization and how it used. Subsequently, any method developed to value information must be able to capture these aspects. The value of information scales as its accuracy is increased up until a saturation point. Perhaps the most reliable aspect of information to calculate is its cost, which would also need to be incorporated into any method and can be used for a Cost Approach as shown in 2.3.2.1.

In summary, there is an overall need for a more reliable and well defined approach to determine the value of information. This approach or method's development has been stifled by some prominent issues relating to information's attributes which are needed by established valuation methods. However, there are still aspects of information that can be used to value it as well as handle it as an intangible asset. The next chapter presents the valuation framework based on what was discovered in the literature review. This framework is the first part of the solution provided by the study, although many aspects of the framework can still be applied independently from the rest of the solution.

Chapter 3

Valuation Framework

Chapter 3 presents the top-down approach; the fundamental principle and foundation of Decision Based Valuation (DBV). Furthermore, it describes how an organization should construct and optimize their data value chains. The aforementioned principle is important in understanding DBV and can be used independently from the other methods presented in this study. By implementing the framework described in Chapter 3, an organization is able to extract additional value from their data and information systems. This value extraction is achieved through properly identifying the costs and value of data and information so that these resources can be optimized.



3.1 The Top-down Approach

The need for the simplification of Big Data and information systems is apparent when observing their complexity. It has become the norm within industry to collect data first and then analysis thereafter to determine its worth. However, this approach creates large volumes of data that often hold no value for the organization collecting it. There is also the fact – identified in the literature review – that data is unique and should be linked to the organizations needs. Thus, the question of what makes data unique to and organization needs to be answered and that answer is decisions; since organizations have to make unique decisions according to their circumstances and situation, the information they require is inherently unique as well. Therefore, the crux of the problem with valuing data is also its solution.

By adopting a top-down approach; starting at the decision – its uniqueness – a valuation method is able to capture the true potential value of the data and information. Chareonsuk and Chansa-ngavej (2008) uses a top-down approach to help identify intangible assets through a four step method starting with the company's vision, but more importantly, the function of each intangible asset. In a similar sense, the function of a intangible asset relates to information and its function; to make a decision.

The action of looking at a decision to value its information has another benefit; if there are no decisions relying on a specific piece of information, or if there are alternatives available that do a better job, then that information has little to no value. This is evident in the description of a data value chain in section 2.1.5, where value is added during distinct stages of the value chain but, more importantly, the value is only realised once the information is consumed and used. This is the most fundamental principle to the valuation framework presented in this study; *information only has value once it is used*. The valuation framework is therefore the building block for the valuation method developed in this study and determines how it values data and information.

The top-down approach builds upon aforementioned fundamental principle and directs organizations to, "value first then collect later". If an organization knows that there is a decision that can use the information, then there is a mechanism where information can realise some potential value. If there is no decision that can use the information, then it becomes significantly more difficult to realise its potential value. Therefore, by approaching the data value chain from the top of chain – the decision – to the bottom of chain, the likelihood of realising the information's value is significantly increased. Furthermore, this allows for value to be filtered through the different stages of the value chain so that each stage can be valued. The result of which is being able to value the data used for the information, as well as determining the contribution of value added by processing. This understanding of how value is added provides organizations with greater control over their data and information systems. This greater control arises from being able to make strategic decisions on what data to collect and how to process it. Moreover, by only investing in information that has potential value that can be realised, organizations can simplify their data and information systems and cut back on waste. This simplification and removal of wasteful data and processes saves both time and money for organizations. However, the benefits do not end with the removal of wasteful processes and data, it also allows organizations to focus on quality; specifically the quality of valuable data. Now an organization is not only saving money and time but they are also generating more value out of the modern asset that is data and information.

In summary, the top-approach to data specifies that data should not be collected without a fully worked out value chain. That value chain is the decision down to its data; capturing all of the points where value is lost and gained. Relying on the fundamental principle of how information's value is realised to eliminate worthless data and attribute value throughout the value chain. It should be noted that this approach to data is not always the optimal approach, especially in the scientific and Research and Development (R&D) fields. In these fields, the organization is attempting to find correlation or hints about things that are currently unknown. Therefore, the norm of collecting all the data possible is in fact the best solution. However, for most organizations, especially businesses, this would not be the case and a top-down approach would yield better results.

The following sections will detail how an organizations would apply the top-down approach to a data value chain by looking at its distinct stages. The main costs and important aspects of each stage are also identified to allow organizations to be prepared for the application of Decision Based Valuation presented in Chapter 4.

3.2 Identifying Decisions

Identifying decisions is the first step in constructing a data value chain and dictates the rest of the inputs. The top-down approach is also only applicable for decisions that require certain information to make. Consequently, quick decisions that are made "off the cuff", especially those made from experience or previous insight, can be disregarded. It is also more beneficial for an organization to identify the most critical decisions first for analysis as this will yield the greatest results. This may not be applicable for all organizations as some will have sufficient finances and labour to collect information for all important decisions.

Therefore, when presented with a set of decisions to analyse, the follow methods can be used to determine which subset of decisions should be looked at first when resources are limited.

3.2.1 Decision Utility

After identifying the most critical decisions within the organization, the next step is to determine which of these decisions have substantial information requirements; excluding those that merely require the decision maker to make a judgement call or using his or her experience. If this new list of decisions is still too large to analyse, then simple utility function can be applied to decide which subset of this set of decisions would be the most beneficial to analyse. This utility function is simply described below in Equation (3.2.1) where the value of the chosen subset ($v(S^*)$) needs to be greater or equal to any and all other subsets (S_i) of the set of decisions (N). Furthermore, this subset is limited in size by financial, time, and labour restrictions. Thus, depending on those restrictions, the subset will be expanded or reduced. Note that these restrictions refer to conducting the analysis and not the decision itself; for instance an organization might only be able to afford to put an employee on the task of creating value chains for one week which could equate to five decisions.

$$v(S^*) \geq v(S) \quad \text{for all} \quad S \in N \quad \text{where} \quad S \subset N \tag{3.2.1}$$

Thus, an organization will want to make a subset of decisions that; meet the financial and labour restrictions, and whose utility is greater than or equal to all other subsets'. This will assist to maximize the possible return on investment on the organization's big data system. Equation (3.2.1) can be made into a rule for a program to automate the process of selecting the subset of decisions. Doing so would obviously require some initial development to have the database of decisions and the formula to determine the best subset. A simple version of this program can be done in Microsoft Excel, where the decisions can act as line items with their value and other attributes populated.

The concept of decision utility often does not provide a detailed enough picture of the decisions to adequately choose between them. Therefore, the concept of volatility is introduced as is discussed in the following section.

3.2.2 Decision Volatility

Decisions are often not only ranked on their utility, there is always a certain aspect of risk associated with them. This risk can be derived from many aspects however, in the context of this study, the risk arises from: (1) the range of the decision's outcome, and (2) the rate of change of the decision. In simple terms; the more volatile the decision, the greater the need for the top-down approach. Therefore, depending on the resources of the organization, they would analyse the top x volatile decisions. Since volatility includes the value of the decision, it will still favour decisions with high financial rewards such as the decision utility approach.

3.2.2.1 Outcome Range

The range of the decision's outcome is simply described as the range between highest and lowest potential gain after making the decision. If the information for the decision is perfect, then it is expected that the decision will yield its highest potential gain. Conversely, if the information is imperfect to some extent, the decision will yield a potential gain between that of its highest and lowest values. If the range of the decision's outcome is large, it is said to be volatile.

3.2.2.2 Rate of Change

The rate of change of a decision describes the frequency at which new information is required for the decision. For example, a decision being made on a weekly or daily basis would have different frequencies. This inherently affects the rate of information required to make an accurate decision. If the decision is made more frequently (e.g. daily versus monthly), it is said to be more volatile.

3.2.2.3 Volatility Reward Equation

To describe the volatility reward of a decision, Equation (3.2.2) was created. In Equation (3.2.2), V_{max} and V_{min} are the maximum and minimum of the decision's value range and the variable RoC is the rate of change of the decision in days. For the basis of this research, it is assumed that the decision will not change more frequently than daily however, it is possible. If the decisions are made more frequently, the equation can be adapted to suit an hourly or less rate of change.

$$V_\delta = \frac{V_{max} - V_{min}}{2} \times \frac{\text{RoC}}{12} \tag{3.2.2}$$

This definition of volatility reward yields the average potential value per month for which the decision is worth if accurate and reliable information is provided. It should be noted that the specified range should be linked to the rate of change provided. For example, if the rate of change is every five days, then the range should be the potential value of making that decision every five days.

This volatility reward value (V_δ) is then used to rank the decisions. Thus, the top x volatility reward values would then be chosen to apply the full top-down approach to.

3.3 Selecting Critical Information

Once a subset of decisions has been identified and chosen, they will need to be analysed. This analysis is to determine the information that is needed to make an effective decision. Subsequently the first step is to determine the list of information that is required for the decision, this includes for instance the units used to describe the information. The next step of the process is to determine how the information should be presented by considering how it affects: (1) the users understanding of the information, (2) the ease of selecting the optimal decision while using the information, and (3) validating and defending of the decision with the provided information. When information is poorly presented, even if it is accurate, it can often lead to misinterpretation and human errors. Thus organizations must always be clear on exactly how they want the information to be presented for optimal use.

Another important aspect of selecting information is to ensure that all decision parameters are covered with what is chosen, while selecting too many parameters can also lead to confusion and difficulties. Thus, the most critical parameters need to be covered. A *critical parameter* is one that if not considered, can lead to significant loss in value through waste, damages, and loss of revenue.

Example

To accurately select where to place oil drills, the organization requires detailed information on where the oil deposits lie as well as the terrain type. The desired output is a map that is divided into coloured zones of one square meter blocks. The colour of these blocks should represent the distance to the oil reserves underneath the ground. Furthermore, they require each of the grid lines to show elevation. Lastly, they require a secondary translucent overlay that can be placed over the first map which hides all the blocks that cannot be drilled due to terrain difficulties. Thus the organization requires these two maps to make the most optimal decision.

3.4 Identifying Processing Techniques

After the determining what information is needed, processing and analytics methods that have the ability to transform data into the desired information need to be investigated. It should be noted that there are often different methods which all yield the same results however, some may require resources which are unavailable to the organization or specialized software. Furthermore, it is always important to think of minimizing costs where possible. For example, if there are two processing methods available but the first choice requires investment from the organization, it may be wise to select the second best option.

The processing method can be a variety of things, from specialized hardware and software to human capital. When a typical processing environment is assessed, it is normally the employee responsible for processing the data that has the most influence on the value added. Therefore, selecting the right persons to do the processing can be just as important as choosing the right tools. These methods will also add the most significant costs to the data value chain.

3.4.1 Selecting Optimal Methods

When selecting the method(s) to be used to process the data into the desired information, an organization has to choose a certain method. This methods should have the following characteristics in descending order of priority:

1. The ability to convert the data into the desired information output, including how the information should be presented,
2. Is the most efficient and accurate method available to the organization, and
3. Is the fastest method available to the organization.

The above characteristics should help to guide an organization in selecting a processing method.

3.4.2 When to Invest in New Methods

An organization should invest in new methods - such as specialized software or hardware - when the performance benefits are proportionally better to the cost increase. Some key questions to be asked by an organization when investing in new methods are:

- Is this for a once off decision that will likely not need to be made again?

- Will these methods be able to process data for other decisions, and if so would they perform better the what is currently available?
- Is the cost of this method covered by the value of the time saved or quality of the information?
- What is the likelihood that the new method may require further training of employees, thus extending the time before it is operational?
- Are there no methods currently available to the organization that can meet the desired information output and presentation?

The answers to the above questions will help an organization decide on whether or not its worth it to invest in new methods.

3.5 Determining Data Required

Collecting the right data is perhaps the most important step in the entire value chain. No matter what technique an organization uses, if the data you need isn't there, your information will be ineffective or worse. Therefore, care needs to be taken to ensure that the data being collected will be able to meet the requirements of the information once processed.

The next important aspect of choosing what data to collect is deciding between multiple options. There may be cases where an organization can process different data into the same information. A simple example would be that of the motion calculations of an object, where there are multiple combinations of data that can generate the full picture of its motion. In such cases, it is important to determine which data will yield the greatest accuracy while still being cost effect. Thus, there are few key points to consider when selecting what data to collect.

- What is the available resolution and quality of the data;
- What is the difficulty of obtaining the data;
- Is there alternative data that can be manipulated to provide the same results;
- Can the data be used for other data streams;
- How reliable is the data;

There may well be cases where there is alternative data that can be manipulated to suit the requirements as set by the information output. Furthermore, this data may already be collected, or be in the process of, to be used for another information output entirely. In such a case, it would be highly beneficial and cost effective to use the same data for the current information output. There is however one restricting factor, even though the alternative data can be manipulated to suit the current output, the quality of that data still needs to meet the quality requirements of the information. This aspect becomes even more significant when the alternative data is from older systems or systems that use less effective capturing devices. Therefore, before an alternative data option is used, the organization must ensure that the data can be manipulated to at least the lowest required level of accuracy as is needed by the information output.

If no alternative data options are currently being used in the organization, the next question that should be asked is, "are there any other information outputs that would be able to use this data?". This question could be very involved and it may be too lengthy a process to determine all the possible uses for the data to be collected. To avoid unnecessary hours of analysis, it is best to ask the question, "can this data be used for at least two different outputs?". This is by no means a complete picture of the uses of the data, but it does introduce the possibility of future information outputs using the same data while being significantly less time intensive.

3.6 Determining Collection Methods

Once the data to be collect has been chosen, it becomes a matter of selecting the collection/acquisition method that will meet the accuracy and precision requirements. This stage is perhaps the simplest of the process as its just a matter of matching the data requirements to capture device. There are still some consideration to keep in mind:

1. The availability of devices within the organization that meet the requirements;
2. The interface of the data acquisition device;
3. The flexibility of the device:
    a) can it be used to capture other forms of data?
    b) is the precision variable?
    c) can the device have multiple inputs?
    d) is the device portable or not?
4. The cost of the device;

5. The support for the device.

The flexibility of the device is perhaps one of the most important considerations as the data it collects could be destined for a once-off decision. In such a case, it would be a wasted asset if it wasn't able to be used again for other data acquisition purposes. Furthermore, the organization needs to ensure that the method used to obtain the data from people meets the accuracy requirements of the information. Lastly, the cost of the devices or methods used to acquire the data are the first costs incurred by the data value chain but also the first creation of value.

3.7 Chapter Summary

Chapter 3 met the objectives: 2a and 3b as stated in section 1.5. These objectives were: *Identifying where value is lost and gained with information*, and *Creating a framework that details how to approach information valuation*. The top-down approach described in this chapter highlighted the fact that information gains it value from its decision and only after it has been used. Moreover, the top-down approach showed that the costs of that information are spread throughout its data value chain and at each stage in the chain there is an addition of value and cost. This directly relates to the first objective, 2a, listed above.

The valuation framework also provides the approach that Decision Based Valuation (DBV) will take to value information. Understanding that information only gains it value once used, results in DBV using the decision as a starting point to value information. Furthermore, it provides insights into how to determine the cost of information as well as attribute value gains throughout the data value chain. The understanding and down approach detailed in this chapter are not limited for use only by DBV; they are useful to all organizations with data and information systems. Thus this chapter was presented separately to the core DBV method.

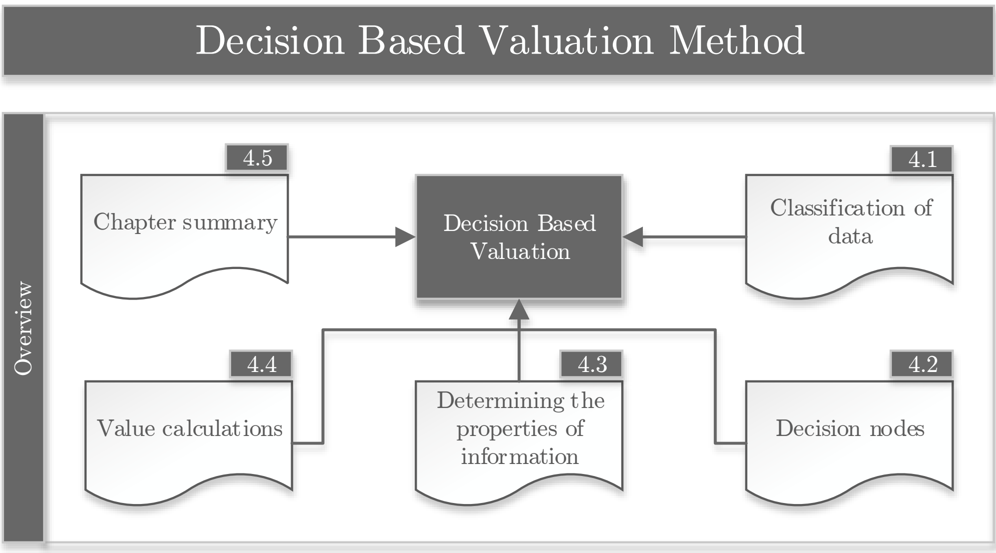
In conclusion, the top-down approach provides a superior way to look at data and information when the ultimate goal is to value and optimize them. However, there are fringe cases where applying a top-down approach to the collection of data is not the most effective strategy. Institutions such as universities or R&D labs often require large quantities of data to be minded to inadvertently come across new and unplanned insights and discoveries; one of the strengths of Big Data. Subsequently, when the aim is to discover something that is currently unknown to the organization, then the top-down approach might not yield the best results. For most other cases, it provides valuable direction and control over the cost and value of information.

Therefore, the valuation framework should provide the reader with a better understanding of the lifecycle of data and information. This understanding can then be used to effectively apply Decision Based Valuation as well as understand its approach to information valuation. The Decision Based Valuation method is presented in the next chapter where the reader is taken through its core concepts of the valuation method – Chapter 4 is the second part of the solution provided by the study.

Chapter 4

Decision Based Valuation Method

Chapter 4 describes the application and details of the Decision Based Valuation (DBV) method developed in this study. The DBV method uses the valuation framework of the previous chapter as a foundation to understanding how information should be valued and how to determine its costs. The chapter begins with the creation of standard classifications for DBV; these classifications are based off data value chains and the valuation framework. The classifications distinguish between different data types and how they gain their value differently throughout their lifecycle. Following the classifications, one of the core principles of DBV, Decision Nodes, is described. Decision Nodes are created to replicate the attributes of physical assets to a certain extent. These Decision Nodes act as the criteria that information needs to fulfil for it to gain value. After describing Decision Nodes, details are given on how to determine the attributes of information and the different formulae required. The chapter concludes with providing the calculations required to determine the value of information and subsequently its data.



4.1 Classification of Data

The classification of data serves three purposes: (1) to standardize the method developed in this study, as well as future studies, according to specific data types, (2) to determine the type of data that can, and should, be valued, and (3) to provide specific formulae catered towards certain data types used by DBV. The standardization allows aspects of the method to be improved independently of the rest. For example, how to set up the required information according to the data it needs – the information and its data are identified by using the valuation framework. Furthermore, the valuation process is length and thus organizations would choose to only use it on specific data. The practical implication of these classifications is to be able to differentiate between how certain data types gain and lose value throughout their value chains – as noted by purpose 3. This is an important consideration as DBV is required to distribute information's value to its data, how that value is transferred is dependent on the type of data and how it was processed. These classifications are presented in the following sections.

Note: All of the classifications provided in this section were created for use by the Decision Based Valuation method and are proposed classifications that require refinement with iteration of the method. Furthermore, these classifications do not refer to existing classifications such as primary and secondary data (Glass, 1976) or technical classifications such as file formats (Guck, 1999). The classifications described below are based off the data value chains presented in section 2.1.5 and, as previously mentioned, propose classifications to differentiate between data that gain and lose value differently through their value chains.

4.1.1 Data Types

There are many classifications for data; whether it be programming based classification, Zandbergen (2015), or the statistical classification of quantitative and qualitative. However, for this study four unique classification are used to distinguish types of data which predetermines how they are handled within the Decision Based Valuation method. The four data types are: Type A, Type B, Type C and Type D, each of which is discussed in the following sections.

4.1.1.1 Type A Data

Type A data, or *Operational Data*, is data that is used on a regular basis by an organization at a frequency determined by its application. This data has a short life span and is continually refreshed. Furthermore, it is very time sensitive and can be rendered useless if not used on time.

Type A data has the following attributes:

1. It has a predetermined frequency;
2. It has a short usable life span proportional to its frequency;
3. It is very time sensitive and its value depreciates rapidly after its time-frame; and
4. Its value is quickly realized through its use.

Example:

A prime example of Type A Data is the monthly consumer demand for a certain product. This information is highly time sensitive as it is only relevant for the month it was collected for, after which its value diminishes almost completely. Furthermore, its has frequency of one month; new data is required once every month. Lastly, its value is easily realized as it is use is almost immediate.

4.1.1.2 Type B Data

Type B data, or *One Time Decision Data*, is data that is used for infrequent and often once off decisions within an organization. This type of data often precedes new projects or business ventures and can take quite some time to acquire and process. It is also very accuracy sensitive with a high potential value.

Type B data has the following attributes:

1. It does not have a predetermined frequency;
2. It has a usable life span of the length of the decision;
3. It is often not very time sensitive;
4. There is a high emphasis on accuracy;
5. It has a high value but often a high cost as well; and
6. Its value takes a long period of time to be realized.

Example:

A good example of Type B data is mineral surveys of certain pieces of land. This data carries a high value since, if a large mineral deposit is identified then the organization stands to earn a high profit. However, once the data has been used, its value almost completely disappears. Furthermore, it will take time to extract the minerals from the earth and determine how accurate the data was, this means the true value of the data is only realized some time after its use. Lastly, the cost of acquiring and processing the data can be almost negligible to its potential value, especially for rare minerals such as diamonds.

4.1.1.3 Type C Data

The third type of data is Type C, or *Legal and Safety Data*. This data often yields no value to an organization other then regulatory and legal indemnification or prevention of injury and damages. An organization will most likely be required to acquire and store a large variety of this data, and will most likely never be looked at by the organization itself. Furthermore, the value is only in the event of something happening thus, if the data is collected and used properly, in the case of safety data, no economic value will be generated. However, not collecting this data can lead to serious penalties or even loss of life if something does happen.

Type C data has the following attributes:

1. It has almost no value to the organization;
2. The true value of the data can only be estimated on the probability of its need;
3. It often has to be stored for a minimum of five years if not longer;
4. The organization is legally required to collect and store it;
5. There can often be a legally required frequency; and
6. It can have an indefinite life span.

<u>Example:</u>

A simple example of Type C data is customer credit checks. They add no value to the organization's business model however, organizations are legally required to be obtain and store their customer's credit checks. Furthermore, these credit checks will likely only be opened and analysed if there is a dispute, for example when reclaiming bad debt.

4.1.1.4 Type D Data

The last type of data is Type D, or *Research and Innovation data*. This data has the innate attribute of being high risk high reward. There will be many cases where Type D data will end up not yielding any income for an organization. However, once in a while it may contribute significantly to the organization, for example through new innovations or process improvements.

Therefore, this data should not be neglected even though, for most organizations, it should not be prioritized.

Type D data has the following attributes:

1. It has a very low chance of providing future incomes;
2. It is difficult to predict whether it will provide value;
3. It has a large value range; and
4. It has a long life span;

Example:

An example of Type D data could be the frequency readings on the housing of electric motors. These readings could fall into another category if they were being measured for a specific purpose of decision; however for this example organization X had spare sensors and decided to deploy them as mentioned above. In this scenario, the data being collected has no inherent value. However, it can obtain value if anomalies or patterns in the data could be identified thus leading to new developments such as improved housings. This type of data has a low chance of yielding value, although there is the possibility it may bring significant development which can result in a net positive value. In the above example, improving the housing could lead to reduced maintenance costs or reduced maintenance times.

4.1.2 Value Growth Rate

Section 2.1.5 describes the value chain for data and how value is added, and costs are incurred, throughout this value chain. Value growth rate focus on the first of these value gains; the value data gains according to its data source and acquisition method.

4.1.2.1 Type 1

Often the value of information increases with its accuracy. Accuracy is born from various aspects from data collection to the quality of the analysis. The one aspect which is of interest for Type 1 is the increase of accuracy due to the volume of data. In certain cases, a relationship can be derived showing how the accuracy of certain information increases as the data it was derived from expands. It is also apparent that this increase in accuracy has a limit and that after a certain point having an increased volume of data does not bring with it any more value. Moreover, after this *saturation point*, value is actually lost as processing times increase as analysts need to wade through the excess data.

Most of the Big Data collected today can be categorized as Type 1 data, where the accuracy of the information depends highly on the volume of data.

Example

A simple example of Type 1 data is traffic data. Collecting this data doesn't require any significant technology thus, the accuracy of this data isn't restricted by the instruments used to collect it. The accuracy is in fact linked to the volume of the data; how many hours of the day were recorded, how many intersections and so forth. Therefore, the greater the volume of data collected – from multiple sources – the greater the overall accuracy of the information. This greater volume of data can lead to new information, such as seasonal or weekly trends, which would not have been possible if only a small sample was taken. This illustrates a relationship between volume and accuracy and how volume directly translates to value; town planners and civil engineers can make more informed decisions on the construction of roads for instance.

4.1.2.2 Type 2

Type 2 value growth rate describes information whose accuracy is derived from the resolution of the data, typically from better measuring instruments. This differs from Type 1 data where accuracy can be improved by merely increasing the amount of data. This results in two main outcomes: (1) it is significantly quicker to improve the accuracy of the data if an organization so wishes, up until the point that technology allows, and (2) the cost to improve the accuracy of the data is not proportional to the current operating costs unlike Type 1 data.

Example

A simple example of Type 2 is the increase in resolution of mining maps. The greater the resolution of the maps, the more accurately operators can drill. This results in less wasted resources. However, it should also be obvious that in this example, there is the limit to the value of the accuracy. If the map is provided to within half a meter there would be a tangible gain over, say, within five meters. Providing an accuracy of millimetres however would not improve the value of the map as the machines cannot drill that accurately. Increasing the accuracy of these maps would require improved technology and instruments, the accuracy of this information is obviously capped at the accuracy the current technology is able to provide.

In summary, Type 2 refers to data sources and their corresponding information outputs which benefit in accuracy through an increased resolution or quality of data. The value of this increase is proportional to the increase in value of the information output when accuracy is improved.

4.1.2.3 Type 3

Type 3 value growth rate refers to information that has a weak relationship between accuracy and value. All information requires a certain level of accuracy to be valuable, but some information's value hardly changes when the accuracy increases. Type 3 data also has a high tolerance to accuracy variation or consistency. Certain catch phrases could be indicators that the data an organization is dealing with, is Type 3 data; such as "ball park figures" or "educated guess". This category of information is often the easiest to supply due to its lax accuracy requirements.

Example

An example of Type 3 data is knowing the amount of alternative solutions that are available for selection for a project. If the number provided is not completely correct, it doesn't affect the outcome of the project as long as there were sufficient alternatives that it would warrant certain responses. These responses could range from further investigation of alternatives to just ignoring them. As can be seen with this example, this information has a high tolerance to accuracy changes and is only weakly related to the level of accuracy.

Another example of Type 3 data is weather information. For most applications, knowing what the temperature is within 100 square meters is unnecessary and adds little value to the general populous. Knowing the temperature for an entire town on average is sufficient for most. Furthermore, knowing what the temperature will be every five minutes also adds almost no value, where often only two to four readings per day would be sufficient. This is a good example of Type 3 value growth rate, where once the volume and velocity requirements are met, the information becomes saturated and value no longer grows. Note that in some cases this may not be true for scientific purposes, subsequently classifications should always be organization specific.

4.1.3 Value Transfer

The next significant value gain of data occurs when it is processed into information. For the Decision Based Valuation, two options were created: the first type is H - high value retention - and the second type is L - low value retention. These two types of value transfers will be discussed below, their general formulation for use in the value calculation will also be provided.

4.1.3.1 Type H

Type H value transfer, or *high value retention* transfer, describes data that holds most of its value as data, with only a small portion of it being added once processed into information. This often occurs when the processing of the data is both cost effective as well as fast. This leads to almost no downtime when processing the data into its information output – to be used for its desired decision. Type H value transfer is the most desirable of the two, as it allows organizations to quickly capitalize on data it collects. Furthermore, Type H data often requires less specialized hardware and software which further reduces its processing costs.

<u>Example</u>

A simple example of Type H transfer data is that of operational yields; such as tons per hour. This data is quickly and efficiently processed into performance metrics such as Key Performance Indicators (KPI's) to be used by operations management to adjust their strategy. As is evident, there would be very little downtime when processing the above data into various performance ratios and thus it can be said to have a *high value retention*.

<u>Formulation</u>

The value gained through processing is dependent on the type of value transfer, in this case Type H data gains little value when processed. This data typically gains between 5% and 10% value after processing.

$$V_P = 1 + 0.05 \quad \text{to} \quad 1 + 0.1 \tag{4.1.1}$$

4.1.3.2 Type L

Type L value transfer, or *low value retention* transfer, describes data that requires a significant amount of processing before any real value is realized. The majority of Big Data is Type L value transfer, where there is significant downtime before the data can be processed into its output information. Another characteristic of Type L value transfer is that this data often require specialized software and hardware, especially for Big Data systems. However, Type L normally yields information of higher value than Type H, meaning it should not be avoided in favour of Type H. Furthermore, Type L usually holds its value once processed for a longer period of time as compared to Type H.

Example

An example of Type L value transfer is high resolution sonar data of an ocean bed. There is a long processing downtime before an actual map can be generated. However, once generated, this map holds a lot of value. Furthermore, this map can be re-used for many different application further increasing its value.

Formulation

Type L data gains a significant amount of value during processing. This value obtained is proportional to the difference between data acquisition time and data processing time as described in Equation (4.1.2).

$$V_P = 1 + \frac{t_P}{t_A} \tag{4.1.2}$$

4.1.4 Overview

When auditing an organization's data and information systems, classification should be the first step in Decision Based Valuation. Beside each value chain should reside a Type identifier such as *A-2-H*. This information is vital in performing the DBV. Furthermore, it helps the organization keep track of all its data which allows it to optimize its ratio of data types.

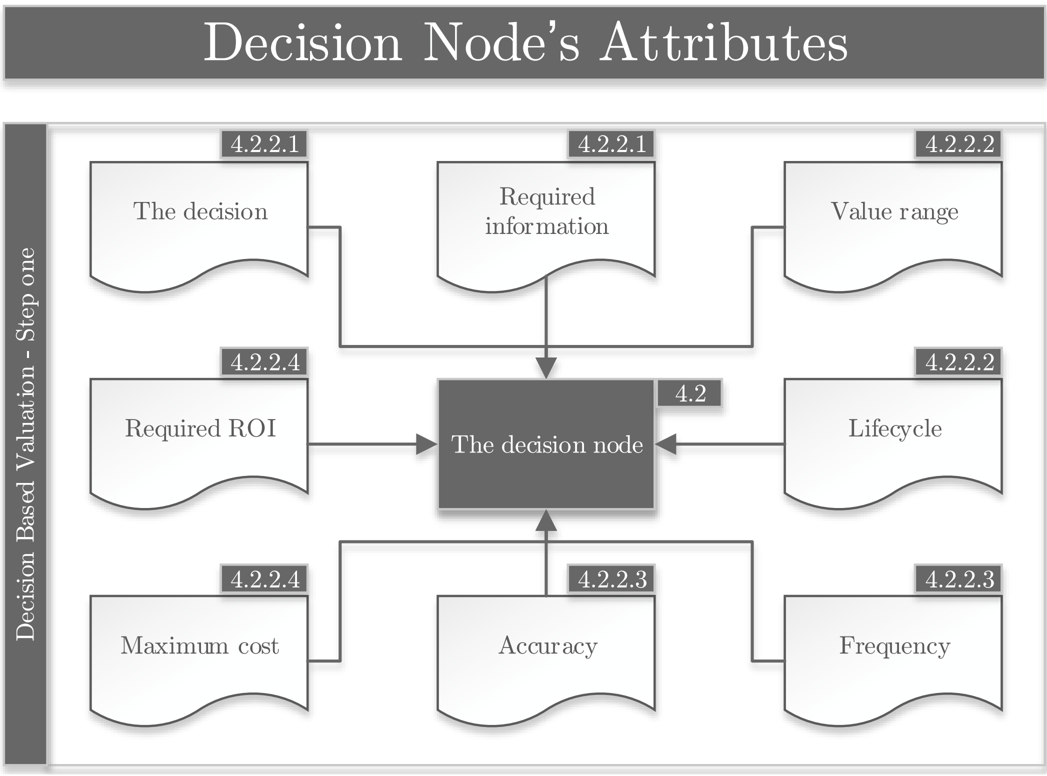
The majority of data within an organization should be Type *A-1/2-H* as this data's value is quickly realized and would offset most, if not all, the data and information system's costs. The next preferred data type is *B-1-L*, this type can be seen as the big value contributors which churn a net profit for the organization. The least favourable type is *C-2-L* and all data of this type should be avoided and kept at a minimum.

For the purpose of this study, the scope of DBV is limited to just Type A and B data. Type C and D data requires probabilistic and prediction models to determine if and when events may occur as well as determining their magnitude, which directly relates to the perceived value of those types of data. Type C and D data is left for later development and would be added to DBV when completed.

4.2 Decision Nodes

The Decision Node is the core component, and first step, of DBV. Its purpose and its attributes will be discussed as well as how it incorporates established valuation methods. A summary of the attributes of a Decision Node are shown in Figure 4.1 – the entire DBV process is available in Appendix A.

Figure 4.1: A Decision Node's Attributes



4.2.1 Function and Description

A Decision Node is analogous to a physical asset. A physical asset is made up of parts with a certain function in mind. These parts, and subsequently the physical asset itself, have a certain cost and performance attributed to them. Lastly, this physical asset has the potential to generate value for the organization who owns it. Similar to physical assets, Decision Nodes are made up of parts with a certain function in mind. Furthermore, a Decision Node has a cost associated with it (depending on its parts) as well as the potential to generate value for the organization who owns it.

A Decision Node's parts are such things as information and its quality, in fact there are eight core parts or attributes that make up a Decision Node. Lastly, a Decision Node is only able to generate value for its organization if it can perform the function it was created for.

The concept of a Decision Node, as previously mentioned, is the central mechanism of the DBV method. Moreover, it is the principle building block towards making information accountable. It adds the necessary structure to data and information that these resources are lacking. Decision Nodes can therefore be seen as intermediaries that help bridge the gap between mere resource and a valuable asset for an organization.

4.2.2 Attributes

A Decision Node comprises of eight core attributes as illustrated in Figure 4.1. These attributes help define the value of the node as well as how that value can be achieved. Furthermore, these attributes can be grouped into four distinct categories:

1. Structure: Information Requirements and The Decision;
2. Value: Value Range and Lifecycle;
3. Quality: Information Frequency and Accuracy; and lastly
4. Thresholds: Maximum Cost and Required ROI.

These four distinct categories cover what is essential to not only discovering the value of data but also generating new value.

A more in-depth look into these attributes is covered in the following sections. Furthermore, an example of a documented Decision Node (known as a Decision Node Record), Figure 4.2, is provided of this section. A Decision Node record is the statement of the node and its attributes. As is seen in the example, it is important to make sure that the Decision Node record is clear and easily understandable. The descriptions of how values are arrived at should also be provided for the various attributes. This assists in the auditing process of Decision Nodes by third parties. It is also important to ensure that accurate calculations and deductions are made when determining the value of the attributes. Though not included in Figure 4.2, the Decision Node record also contains the cost of the node and its valuation - after the attributes have been detailed. A Decision Node record is dynamic and is updated regularly to accurately portray the state of the node.

4.2.2.1 Structure

The structure attributes detail the specifics and uniqueness of the decision and directly relates to the information it requires. The differentiation between using the same information for a different decisions is also clarified by these attributes. These attributes also guide the user in determining what the potential value range of the Decision Node will be. Therefore, these are the first attributes that should be determined for a Decision Node, this aligns with the valuation framework and the top-down approach detailed in chapter 3.

The Decision

The first attribute of a Decision Node is the decision it was created for, this identifies the uniqueness of the information through its use as described in section 2.3.3.1. Similar to assigning a physical asset a function, such as an electric motor for a conveyor system, a node requires to be assigned a function too. The decision also gives context to the rest of the parameters supplied by the node, which may remove ambiguity in some cases. Furthermore, if information is unclear, it can be interpreted by looking at the decision and determining what would be needed. For this reason, it is important to be clear and concise when describing the decision the node was created for. It is also important to avoid any form of ambiguity or need for interpretation.

Information Requirements

The most important parameter of a Decision Node is what information is required. Here the details of the information need to be described in full, as well as preference on how it should be presented; for example format and medium. It is critical to include the output variables for the information which will help determine what data needs to be collected. Output variables includes things such as tons/hour or scheduled work hours. Unclear output variables can lead to incorrect or inaccurate information that then results in a loss of value for the Decision Node. Furthermore, a single node can require multiple data streams to be able to make an effective, valuable, decision. For instance, a decision could require both information about spares as well as current operating schedule of a vehicle. These are two distinct data streams (section 2.1.5) which could be processed and handled on completely separate hardware and need to come together to meet the nodes requirements and subsequently earn value. Therefore, the information requirements should be derived from the decision where its aim is to allow the user to make that decision as best they can.

4.2.2.2 Value

The value attributes describe the range and potential value that can be achieved by the information provided to the Decision Node. The information provided to the node will then realise some value in this range according to its attributes, such as accuracy. The value also includes the time or lifecycle of the decision this value applies to.

Value Range

The value range is the potential value that is achievable when using the required information. This range will be used as the basis of the financial calculation of all information in the DBV method. This attribute incorporates the income approach in section 2.3.2.3, describing the possible future benefits of an intangible asset. The main aspect to be captured with this metric is the default value produced without the information and the maximum possible value that can be produced with it. This range is then assessed and the informations value will be placed somewhere in it according to other metrics, such as the accuracy achieved.

Lifecycle

The next attribute of a Decision Node is its life cycle, the use of which occurs often; such as for data amortization in section 2.5 and net present value in section 2.2.1. Lifecycle also appears in the Relief from Royalty Approach in section 2.3.2.4 and is critical to the result of this valuation approach. The lifecycle is the length in time that it is expected that the Decision Node will remain active – as information is perishable as detailed in section 2.1.4.3. This active time can be seen as the useful life of the Decision Node, similar to that of a physical asset. After the useful life of the node is finished, then it is assumed that the node will no longer be able to generate the value as described in its value range. This can often be a difficult question to answer as the usefulness of information can change overnight. Therefore, as with the value range, this information is likely to change during the duration of the Decision Node's life.

Therefore, knowing the lifecycle of the node is important for two cases: (1) projecting the future income derived from the node, and (2) determining the amortization schedule for the node. Without knowing the lifecycle of the node, the aforementioned information becomes significantly more difficult to calculate as well as less accurate.

4.2.2.3 Quality

The quality attributes are the main deciding factors on what value the provided information will be able to realise from the potential range. Therefore the accuracy and correctness of these attributes is crucial towards having an accurate valuation of the information provided to the Decision Node.

Frequency

The frequency parameter of a Decision Node is the frequency at which the decision requires new or updated information. Furthermore, there is a tolerance to changes to the aforementioned frequency to account for if information is supplied early or late. Supplying information early would have little to no effect on the decision and the value obtained by the node; it merely gives the organization a larger buffer to make use of said information. However, if the information is supplied late or not refreshed at the required frequency, a significant loss in value will occur. This becomes especially true when the information is provided after the event for which the decision was for, in such a case, almost all value would be lost for that information. Subsequently, it is important to determine the tolerance for frequency changes for each node.

It should be noted that if information is provided early then one could use that information to forecast future events thus improving strategic decisions. However, with the construction of Decision Nodes, the information provided could be destined for forecasting decisions or day to day decisions. Thus allowing high frequency (early) information to obtain greater value would overlap in such cases as it would already be early with respect to the event the decision is linked to. Consequently, for the scope of this study, information gaining value due to a higher frequency is excluded. Furthermore, having a range of possible values while still within the frequency tolerance is also beyond the scope of this study. In such cases, the information maintains the same value up until the point where it is no longer within the frequency tolerance.

Accuracy

The accuracy of the information is perhaps the most important metric when determining the value information can obtain. There is a proportional relationship between the increase in accuracy of information and its value as shown in sections 2.1.4.4 and 2.4. Perfect information can obtain the maximum value of a decision while imperfect or inaccurate information can obtain almost no value, or in some cases lose value. Thus it is important to specify the level of accuracy required by Decision Nodes as well as the loss or gain in value due to the changes in accuracy. The loss in accuracy results in a loss in value for the information up until a point where the information cannot be used and no longer has any value, this is called the accuracy floor. Conversely, the increase in accuracy can result in an increase in the value of the information, once again to a point, called the accuracy ceiling. Both the accuracy floor and ceiling have an accompanying rate at which the information gains or loses its value according to its accuracy.

The complexity of the relationship between accuracy and the value gained depends on the type of information and the organization. Moreover, unless specified, the basic calculation of value obtained by meeting the required accuracy is equal to 100% - (Ceiling % - Required %)(Value Gain Rate).

4.2.2.4 Thresholds

The threshold attributes are largely used for evaluating the performance of the Decision Node. They are also used as a deciding mechanism on whether an organization should develop or obtain the information themselves or if they should get a third party to do so.

Maximum Cost

The maximum cost is the limit the organization sets on node for the collection of data and processing thereof to obtain the required information. This parameter is an optional one and can be interchanged with the required ROI. Depending on the organization's preferences, one would be used over the other. Furthermore, cost is used for the Cost Approach as shown in section 2.3.2.1 and for replacement cost in section 2.2.6. Thus, by setting the maximum costs, organizations will be able to judge the performance of the node but also receive quotes for obtaining the information when done by third parties.

Required Return on Investment

The required Return on Investment (ROI) is the minimum ratio of cost to value earned that is deemed acceptable by the organization. Therefore, if the collection and processing of data cannot be achieved while meeting the ROI then it should not be pursued. Return on Investment is also strongly linked to Real Options as seen in sections 2.2.3 and 2.3.2.6 where capital can be reassigned according to current performance. ROI is not a required attribute of a decision node and can be interchanged with maximum cost. Even though it is not required, it is still recommended that either a required ROI or maximum cost is derived for each node. Figure 4.2 illustrates the implantation of a Decision Node.

Figure 4.2: Decision Node Example Part 1 (Created with Apple Numbers)

Decision Node

ID: 9090999

The Decision

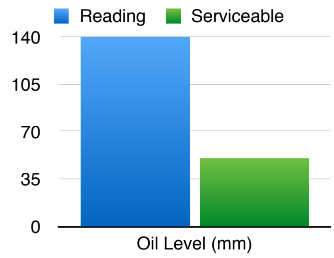
The decision is when to take a hauler truck off rotation and take it in for a service.

Information Required

The following information is required:

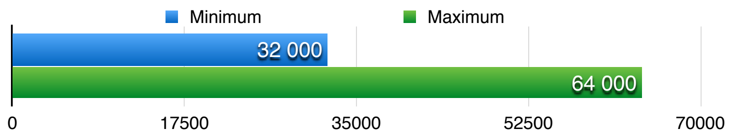
- Oil Level (mm)
- Oil Quality (ppm contaminants)
- Tread Depth (mm)
- Kilometres Driven (km)

The above information needs to be shown in a histogram beside the serviceable levels.



Value Range

The value of the decision is the avoiding the cost of a blown engine as well as losing operational losses associated with that blown engine. The operational losses for a blown engine are R50000 and the replacement of the engine is R14000. It is noticed that if a hauler truck exceeds one of the above thresholds there is a 50% chance of failure within the next few days. Thus it can be considered that half the total failure cost would be incurred if a fixed schedule was adhered to,



Life Cycle

The life cycle of a hauler truck is 10 years thus the life cycle of the decision is equally long.

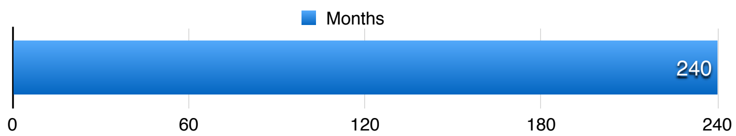
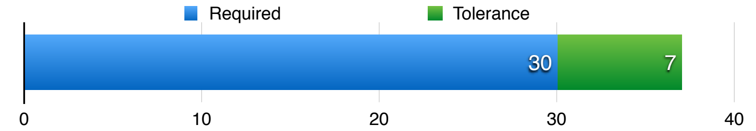


Figure 4.3: Decision Node Example Part 2 (Created with Apple Numbers)

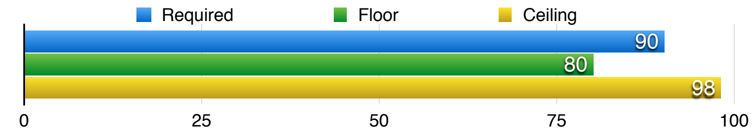
Information Frequency

The information is required once a month (30 days) due to the activity of the truck. Seeing as a motor can fail within a few days of exceeding a threshold, the tolerance is only 1 week (7 days).

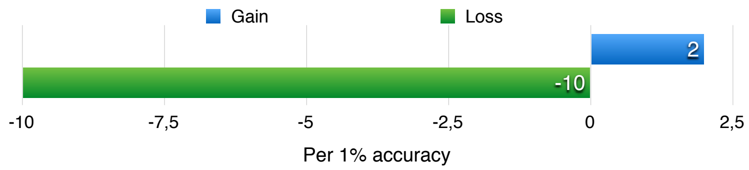


Information Accuracy

The information accuracy needs to be 90% but no less than 80%. Due to measuring inaccuracies, accuracy can only go up to 98%.

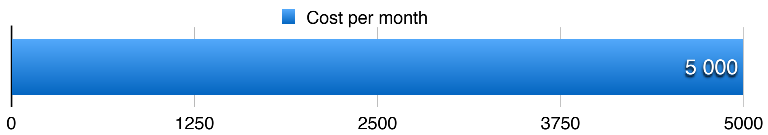


The information gains 2% value per accuracy above the required and losses 10% per accuracy below the required.



Maximum Cost

The maximum cost allowed for this decision node is R5000/month.



Required ROI

Maximum cost was selected thus ROI will not be used.

4.2.3 Criteria of Intangible Assets

Part of the reason for creating Decision Nodes is so that they can help information meet the criteria for intangible assets (see section 2.3). The criteria and how the Decision Node achieves them will now be discussed.

4.2.3.1 Separable

By the inherent nature of Decision Nodes - created for a single decision - information and data can be separated from each other. The node is therefore the mechanism in which information is separated into a form which can be sold or valued. Thus, the separable criteria for intangible assets is met.

4.2.3.2 Future Economic Benefits

Through inclusion of the value range of the decision for which the information is for, the expected economic benefits can be calculated. Where the decision's value range is the range of economic benefits and the value the node achieves is the expected future economic benefit. Consequently, there is proof of future economic benefits and the second criteria for intangible assets is met.

4.2.3.3 Measurable Costs

The costs of a node are well defined for its information inputs; with a reliable and standardized approach to calculating them. These include both indirect and direct costs in the acquisition, storage, processing, and consumption of the information and data. Therefore, the costs are measurable and meet the criteria for intangible assets.

4.2.4 Handling

This section details how an organization can implement and control use of Decision Nodes. There are two main mechanisms to achieve this control: (1) Intangible Asset Registers (IAR), and (2) Decision Node cost centres, both of which will be discussed below.

4.2.4.1 Intangible Asset Register

One of the main methods of tracking and keeping record of physical assets within an organization is the use of Fixed Asset Registers (FAR). These registers hold, to name only three; the description, location, and value of all the assets an organization owns. The concept of a FAR can easily be adapted for use with Decision Nodes, and thus the creation of the Intangible Asset Register (IAR). The IAR functions exactly the same as a FAR but instead of being a record of assets, it records intangible assets or in this case, Decision Nodes.

The only requirement needed to list Decision Nodes in this manner is for each Decision Node to have its own unique serial number issued by the organization (as seen in the provided example in Figure 4.2). This gives an organization's finance department a recognizable way of tracking and reporting on Decision Nodes. The IAR also acts as another building block for the financial accountability of information. This IAR can be a specialized database or, for smaller businesses, it could simply be a Microsoft Excel Workbook with each Decision Node representing a line item.

4.2.4.2 Cost Centres

Another financial necessity is the use of cost centres throughout an organization; to allow for associating costs as well as budgeting. For the implementation of Decision Nodes, they should be assigned their own cost centre as well as employees which handle them. These employees would handle the collection of data, processing, and managing of all the decision nodes. Therefore, any cost incurred while maintaining and creating the Decision Nodes will be added to the cost centre. This allows organizations to easily report on the cost of maintaining all of their Decision Nodes.

4.3 Determining the Properties of Information

This section covers how an organization can determine the properties of the information it is using for a Decision Node. These properties will almost always be organization specific, therefore this section provides the inputs and possibilities to the user for determining them. These properties include determining the value range, the quality attributes of information as well as the cost of information.

4.3.1 Value Range

It should be apparent to most organizations, from Small Medium Micro Enterprises (SMME's) to large corporations, that the value of a decision does not only reside in the monetary gain received from it. The value of a decision is in fact a more complex question which relies heavily on what the decision is for. Furthermore, what one organization could deem valuable might be significantly different from another. Chief among these is safety; preserving human life and avoiding injuries. Decisions whose outcome could affect or endanger someone's life have an inherent value that could far exceed that of its monetary gain. There are of course methods to calculate the value of work place injuries and also, nearing an almost taboo topic, the monetary value of a human life. However, knowing these values is still not enough. One also needs to understand and be able to calculate the probability of these events occurring according to the outcome of certain decisions. Some would argue, and justly, that a human life is immeasurable. In such a case, the ends will always justify the means. That is to say that no matter the cost associated with obtaining the information, it will always be justified and financially viable for the organization. This illustrates how information's value is unique to organizations. Consequently, these choices on how to determine the value of a decision is left to the organizations themselves to determine. The value of information, as noted in the literature review, is unique to the organization that owns it and therefore the final say on what its value is lies with them.

When determining the value of a decision, there are a few key aspects to take note of, namely:

1. The potential monitory value of the decision with and without any information;

2. The potential for loss of life or injury with the associated outcomes of the decision;

3. The potential environmental impacts which are associated with the outcomes of the decision; and

4. The potential reputation/market value/public opinion gains or losses associated with the outcomes of the decision.

Each of the above would have different values and meanings to each organization. For instances, some organizations would not even consider the environmental impact of the decision having any inherent value. As previously mentioned, since the value of information in unique to the organization, it needs to determine the various value calculations. Even though the calculation of value is an internal matter, it is still important to have valid reasoning and justification for any calculation for the purpose of audits and external review. Typically, range is the difference between the maximum potential value with perfect information and that of making the decision without any information at all. This range forms a vital roll in determining the volatility of the decision as well the actual value that is achievable by the provided information.

It should be noted that in instances where the benefit of the information is saving the organization money, the total costs of the options can be used where the smallest cost is subtracted from the cost of the default decision to provide the potential range of savings. These types of scenarios happen often and normally occur where an organization is not sure if the cheaper options will satisfy its needs.

4.3.2 Quality

The quality of information is a combination of its accuracy and frequency as described below.

4.3.2.1 Accuracy

Accuracy is the single most important metric when determining the potential value of data. Inaccurate information is worthless, and in fact can actually cost an organization money. On the other hand, perfect information is highly valuable and guarantees certain outcomes that an organization can plan and prepare for, and take full advantage of. Thus, the accuracy of the information – together with other quality factors – describes what portion of the potential value of the information can be obtained.

Accuracy is lost at two distinct points in the value chain; at data acquisition and at data processing. The final output information stands to lose all of its accuracy if these two stages are not properly handled. The first stage – data acquisition – is perhaps the most important of the two. If inaccurate, low quality data is collected with the acquisition method, the accuracy is permanently lost and the whole set of data can be scrapped. However, unlike with data acquisition, if the data processing is deemed inaccurate, it can always be redone or modified with certain time or cost penalties. Although, these penalties would be significantly less than what an organization would suffer if they needed to reacquire all the data.

It should be noted that even though accuracy can be lost during data processing, these losses are easily avoided with adequate training and proper tools. Therefore, if the accuracy of the information is low, the organization should first look at the data acquisition and the accuracy of the data being collected. Only if the lack of accuracy still cannot be accounted for should they look at the data processing. Processing mistakes do occur and thus should never be excluded.

It is often the case that an organization is unable to provide an accuracy valuation of the quality of their data and/or data processing. In such cases, an estimated accuracy should be provided. If this is the case, then the final valuation of the information will reflect this and could possibly deviate significantly from the actual value depending on the other metrics. The accuracy of the data being acquired, and the processing methods, should each be presented by a decimal value from 0 to 1 (0% to 100%). As previously mentioned, the processing methods can be assumed to be 1 unless determined otherwise. For estimated accuracy, the organization should select one of five possible accuracies:

1. 0.2 for cases of poor unusable data,
2. 0.4 for substandard data quality,
3. 0.6 for usable and functional data quality,
4. 0.8 for excellent data quality, and
5. 1 for perfect data.

For any functional and effective data and information system, the data accuracy should be at least 0.6 or higher.

Example One

The first example is that of exact accuracy coefficients. It is possible to determine the exact accuracy of data when acquisition machines have specified tolerances and resolutions. For example, a thermometer that logs temperatures can have an accuracy of ±0.1 degrees Celsius. If the data requirement is to within 1 degree Celsius then the collected data gets an accuracy rating of exactly 0.9 or 90%.

Example Two

The next example is that of estimated accuracy coefficients. When data is collected from surveys for instances, it is difficult to attribute an exact accuracy to the data. In such cases, one of the five selected coefficients should be selected. If it is assumed that most people would answer the survey honestly while others may answer honestly but still be unsure it would be reasonable to select a 0.6 or 60% accuracy coefficient.

4.3.2.2 Frequency

Frequency is an important metric when determining the value that information can obtain from a Decision Node. If the information is not supplied at the minimum frequency, it will not be available in time for use when needed. This would lead to a significant loss, if not all, of its value. Thus it is important to understand the value gained or lost due to the frequency of information.

Each Decision Node contains a required frequency as well as a tolerance to changes in that frequency (section 4.2.2.3). Thus the frequency component that information obtains according to its supplied frequency is shown in Equations (4.3.1) and (4.3.2). In the following equations; V_{Fr} is the raw frequency component, V_F is the true frequency component, F_T is the frequency tolerance of the node, F_N is the frequency requirement of the node, and F_I is the frequency at which the information is provided.

$$I_{Fr} = \frac{F_T - \sqrt{(F_N - F_I)^2}}{F_T} \tag{4.3.1}$$

$$I_F = 1 \quad \text{IF} \quad I_{Fr} \geq 0 \quad \text{else} \quad V_F = 0 \tag{4.3.2}$$

Furthermore, if the frequencies of the Decision Node and information do not match for dynamic nodes, then the previous cycles difference is added to the frequency number for the following cycle. For example, if the required frequency is 10 and the information's frequency is 12, then for the purpose of these calculations F_I equals:

- Cycle 1: $F_I = 12$
- Cycle 2: $F_I = 14$
- Cycle 3: $F_I = 16$
- Cycle n: $F_I = 12 + 2(n - 1)$

This is due to the nature of Decision Nodes in that if the frequency of the decision and the information differ each cycle, then a compounded error occurs. For instance, if the information needs to be in by Tuesday on a weekly basis, but the information is only available every 8 days then for the first cycle the information will arrive on Wednesday (a day late). However, for the next cycle, the information will only arrive on Thursday (two days late). As the cycles continue, the information will arrive another day later until it is no longer within the frequency tolerance and is now worthless.

This shows that only non-recurring frequency differences are able to maintain value for dynamic Decision Nodes that are not used for reoccurring decisions. This type of recurring error can occur when the processing time of the data takes a standard time which does not allow it to meet the frequency of the node. This type of problem can only be counteracted by either improving the processing time or running processing cycles in parallel (starting the data processing for the next cycle before the current one is complete).

Example

Using the same example of this chapter, Decision Node 1, the frequency component of the information will be calculated for it being provided every seven days and every eight days.

For seven day information:

$$I_{Fr} = \frac{1 - \sqrt{(7 - 7)^2}}{1} = 1 \tag{4.3.3}$$

Thus $I_{Fr} \geq 0$ therefore the frequency component is $I_F = 1$. Now the calculation is done for information provided every eight days:

$$I_{Fr} = \frac{1 - \sqrt{(7-8)^2}}{1} = 0 \tag{4.3.4}$$

Thus $I_{Fr} \geq 0$ therefore the frequency component is $I_F = 1$. However, if the node is dynamic, and requires the information again the following cycle, the information will no longer have a frequency of eight but rather nine. This would yield:

$$I_{Fr} = \frac{1 - \sqrt{(7-9)^2}}{1} = -1 \tag{4.3.5}$$

Thus $I_{Fr} < 0$ therefore the frequency component is $I_F = 0$. This shows that frequency differences are compounded each cycle and it is important to match the frequency of the information to its Decision Node. this example also shows that cycle to cycle, there can be non-recurring frequency differences while still maintaining a positive frequency component.

4.3.3 Cost

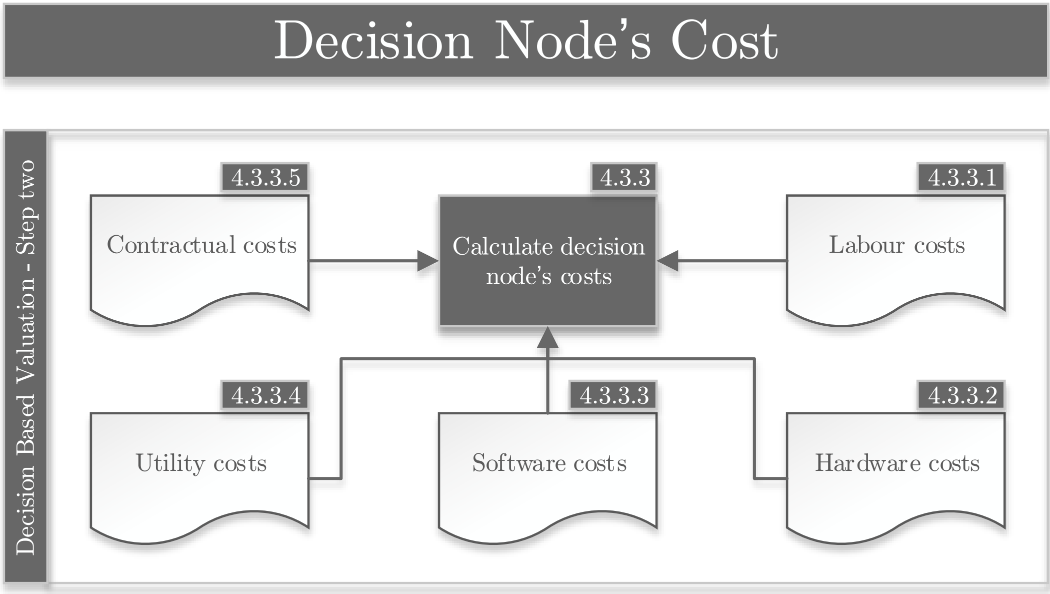
The cost of data can be broken down into five distinct categories: (1) Labour, (2) Hardware, (3) Software, (4) Utilities, and (5) Contractual. It is important to note that different strategies and optimizations will arise depending on the type of costs that are dominant in the cost of the node. For instance, if the cost of the information is mainly coming from contractual costs, then there is very little cost optimization that can be done. Instead the organization needs to ensure that the value range is sufficient to cover the costs, subsequently it becomes a value optimization problem instead of a cost optimization problem. This type of scenario occurs often in industry when business hire contractors and consultants to perform work for them. In such a case, a fixed fee is often negotiated before commencing work. Therefore, an organization would have to ensure that the fixed fee will be sufficiently offset by the value of the information received by the contractor or consultant.

The determination and calculation of costs is the second step of the Decision Based Valuation method as shown in Figure 4.4.

4.3.3.1 Labour Costs

There are both direct and indirect labour costs associated with Big Data and information processing in general. To ensure that an accurate performance ratio is determined and the net profit for various Big Data sources are accurately calculated, all costs both direct and indirect need to be accounted for.

Figure 4.4: Decision Node's Cost



The various labour costs will now be discussed.

<u>Direct Labour Costs</u>

- Hardware installation costs,
- Salaries/wages of those collecting data,
- Salaries/wages of those processing the data, and
- Any consulting fees relating to the Big Data system including Business Intelligence.

<u>Indirect Labour Costs</u>

- Employee training costs,
- Salaries/wages for those archiving and maintaining the data, and
- Salaries/wages associated with system troubleshooting,

4.3.3.2 Hardware Costs

The capturing and processing of Big Data requires various equipment. The cost and scale of this equipment is highly dependent on the scale of the Big Data system, the type of data, and the required output information. The following forms the basis of a Big Data and indeed any data and information system:

- Computer hardware for the capturing of data,
- Acquisition systems such as sensors and DAQ's to capture measured values,
- Information collection mediums such as PDA's and tablets,
- Computer hardware for data storage, and
- Computer hardware for data processing and analysis.

In some cases, computer hardware may be re-purposed for use in an organization's data and information systems. In such cases, the cost to the system should be equal to the depreciation of the hardware while in use. If the hardware has already been depreciated to zero, then that hardware's cost can be excluded from the Decision Node's cost.

4.3.3.3 Software Costs

A significant aspect of Big Data is the processing, storing, and viewing of it. This requires specialized software to accomplish unlike small data sets. This software can cost thousands of Rand and needs to be incorporated into the total costs of the Big Data system. Although the cost of this software can be significant, other software such as Microsoft Office are often used for data analysis. These types of software are typically pre-installed on the computers bought by organizations and used by employees. Therefore, they normally run at zero cost for the node and can excluded from its costs.

4.3.3.4 Utilities Costs

Utility costs includes all the overheads required to run the various equipment. These include electrical requirements as well as maintenance requirements, such the cleaning fans or sensors. Furthermore, any other miscellaneous costs should be included here as well, for instance:

- Training material,
- Safety gear,
- Transportation costs, and
- Server room and office costs (when their sole purpose is to operate the Big Data system).

4.3.3.5 Contractual Costs

Contractual costs normally appear as fixed costs that organizations pay to contractors. However, there are often cost plus items or time related clauses. When applicable, these make up the bulk of the cost of a node. Even though it may seem as if contractual costs are the only costs present for a node, in some instances there are still overhead and employee costs. This typically occurs when the contractor works on-site at the organization's premises.

4.3.3.6 Determining the Cost of A Single Node

The key to accurate net profit values for a single data stream is being able to isolate its costs and value from other data streams. This can pose great difficulty when multiple data streams are used for a single node. Perhaps the most challenging aspect of determining a Decision Node's cost are the contributions to hardware costs that are used for multiple data streams. Moreover, these hardware costs should be depreciated rapidly due to their nature, thus they will not feature as part of the costs of the node after a certain defined period of time.

Therefore the cost of a node can be broken up into two sections: (1) initial costs (usually shared), and (2) operational costs. Initial costs include both hardware and software costs which can be depreciated as physical and intangible assets respectively according to GAAP. Operational costs include a portion of the overheads required for operating the hardware as well as the labour costs for processing and handling of the data. Each node will have a maximum allowed operational cost associated with it as well as a maximum initial cost. These maximum values are linked to the required performance ratio or net profit earned that have been decided for each Decision Node by the organization. Thus, these maximums are subject to change; if for instance the information proves to be more valuable than originally calculated.

To distribute the equipment and software costs between nodes, it is essential that a list of all equipment and software specific to the data and information system be created. This list should then be branched over to the various nodes that require the use of one or more items on the list. In practice, each piece of hardware will likely have multiple nodes relying on them, distributing the costs to each of these nodes is often less than obvious, especially when operational times are not recorded or well known. Due to this fact, there are two basic methods which organizations can use to split the costs of each piece of hardware (and software) between nodes.

1. Divide the costs equally among all nodes i.e. dividing the total cost of the item by the total number of nodes using the item then allocating

each node an equal portion of the costs as well as depreciation.

2. Divide the cost of the item proportionally by utilization time. This requires knowing how many of the total operational hours of the item was used by each node.

Example: Equal Division of Cost

For the basic equal division of cost, one divides the total remaining cost of the asset being used by the number of nodes and each node then receives an equal portion of the cost. For example, a vibration sensor is used to read the vibrations of machines throughout a plant. It was purchased for R80,000 and is being depreciated linearly over five years – that sensor is currently two years old. If seven nodes require data from the sensor, then each node receives the following monthly expense for utilizing the sensor.

$$\begin{aligned}
\text{Depreciation (Monthly):} \quad & D = \frac{80000}{12 \times 5} \\
\text{Node Share:} \quad & N_S = \frac{1}{7} \\
\text{Node Expense (Monthly):} \quad & E_m = (N_S \times D) \\
& E_m = \text{R}190.48
\end{aligned} \tag{4.3.6}$$

Therefore, as seen in Equation (4.3.6), the node will have an expense of R190.48 for each month that it uses the vibration sensor. This expense will then be added to the other expenses such as labour or other hardware costs.

Example: Proportional Division of Cost

If more information is available about the time a node's data uses an asset, the proportional cost can be calculated. This requires one to know the operational hours of the asset as well as the hours the asset is being used for any specific node. Using the same example as for the equal division of cost, we have an R80,000 sensor, which is depreciated linearly over five years. However, this node is known to use the vibration sensor for ten hours per week – the sensor also only has 30 operational hours each week. Using the above information, the proportional cost for the node can be calculated.

$$\begin{aligned}
\text{Depreciation (Hourly):} \quad & D = \frac{80000}{52 \times 5 \times 30} \\
\text{Node Share:} \quad & N_S = 10 \\
\text{Node Expense (Weekly):} \quad & E_m = (N_S \times D) \\
& E_m = \text{R}102.56
\end{aligned} \tag{4.3.7}$$

The cost for the node to use the sensor is now significantly higher after incorporating the actual hours used as shown in Equation (4.3.7). This emphasises the problem with using equal division of costs, even though the method is simple and does not require any extra information, its results can be significantly different from the actual costs. Therefore, proportional division of costs should always be attempted first and only if the required information is not available should an organization use equal division.

It is obvious that option two is the most accurate however, due to the likelihood of organizations not having the required information on hardware use, option one will often be used. It should be noted that it is also possible to use a combination of option one and two. This would be useful when one node uses the bulk of the operational time of the item leading it to burden most of the costs while the other nodes can share the remaining costs equally as according to option one.

4.3.3.7 Depreciation and Hardware Cost Sharing

Hardware and software are paid off overtime; once an item is fully depreciated it no longer needs to be added to a node and only its operating costs need to still be added. Furthermore, since items are typically depreciated every month, any cost calculation should use the current (remaining) value of the item. If a new node is added to the pool of nodes currently using a item, all those node's cost values should be updated as well to reflect their new operational share.

4.4 Value Calculations

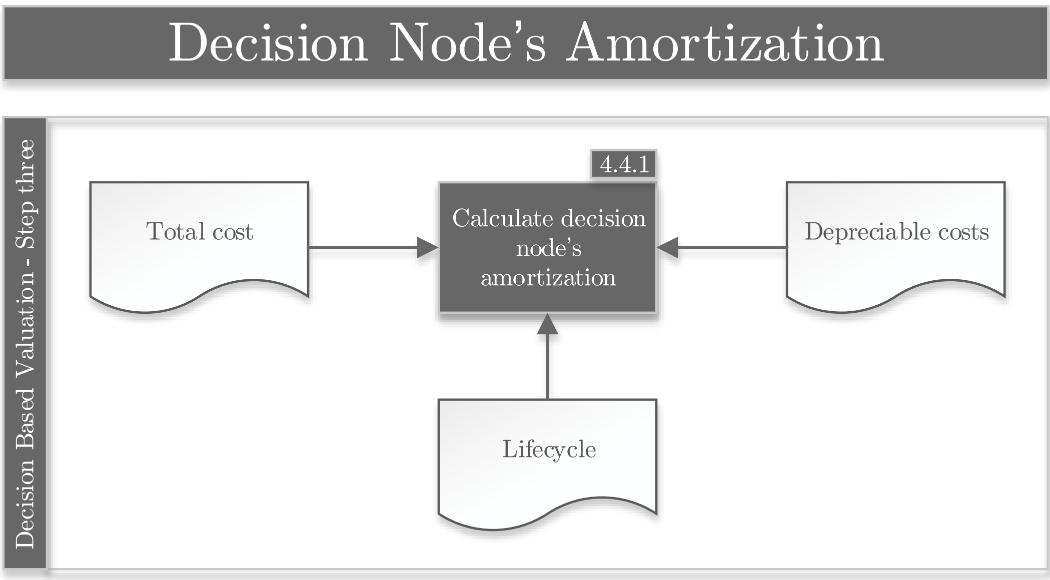
This section details the various calculations that should be followed to determine: (1) the Decision Node's amortization, (2) the Decision Node's value, (3) the performance of the Decision Node, and (4) the distribution of the Decision Node's value to its data.

4.4.1 Amortization of Decision Nodes

Depreciating or amortizing assets is a useful tool for organizations when tracking the loss of value of these assets overtime. Furthermore, it allows organizations to spread the cost of an asset over a certain length of time while also more accurately showing the yearly ROI for that asset. When it comes to data and information systems there are two distinct objects that can be depreciated; hardware and the data itself. The hardware can be depreciated and represented on financial statements according to Generally Accepted Accounting Practices (GAAP) or International Financial Reporting Standards(IFRS). However, information, while losing value overtime, still cannot be listed as an asset on financial statements and subsequently cannot be amortized either. With that said, it remains a useful tool for organizations to understand the true value of their information assets as well as understand the flow of value of their data and information system investments. Furthermore, having a reliable and standard means to amortize information will help complete the picture of information as an intangible asset and bring it one step closer to being financially accountable.

The amortization of a Decision Node is the third step in the Decision Based Valuation method as shown in Figure 4.5.

Figure 4.5: Decision Node Amortization



The practise of depreciating physical assets is to spread the cost of acquiring that asset over a certain length of time. The IAS 16.55 standard (Board, 2015) states that assets should be depreciated as soon as they are available for use by the organization up until it is de-recognized. Therefore, a Decision Node should be depreciated as soon as it is functional, that is, as soon as it has been completed and has been supplied with information. IAS 16 also states that an asset should be depreciated over its useful life, which in terms of the Decision Node is its lifecycle. Therefore, with the life cycle of the node and its costs, the Decision Node can be amortized. Due to the way Decision Nodes are constructed, a linear amortization is best suited to it. This is due to the fact that its value – that of the information that is supplied to it – is refreshed each time new information is provided. Thus the total cost of the Decision Node can be distributed across its entire lifecycle.

The next consideration is the depreciation of hardware used for a Decision Node. There are two possible scenarios for the hardware used to collect, process and store the data; the hardware is listed as an asset and is being depreciated, or the hardware acts as an operational expense and is not depreciated. Only the hardware that is not currently being depreciated by the organization will be depreciated as part of the Decision Node's amortization schedule. Therefore, the breakdown of costs of a Decision node looks as follows.

1. Depreciable Costs:
   a) Labour (all-kinds),
   b) Software (recorded as an expense),
   c) Hardware (recorded as an expense), and
   d) Utilities (recorded as an expense).
2. Non-Depreciable Cost:
   a) Software (recorded as an asset),
   b) Hardware (recorded as an asset), and
   c) Utilities (recorded as an asset).

Thus the depreciable cost (C_D) is the sum of all non-asset related costs for the node. It can also be simply written as the total cost (C_T) minus the non-depreciable cost (C_ND) as shown in Equation (4.4.1). Note that the aforementioned costs are the projected costs for the entire life cycle of the node.

$$C_D = C_T - C_{ND} \tag{4.4.1}$$

Taking the depreciable cost of the node and its life cycle (N_L), the amortization schedule for the Decision Node can be calculated as seen in Equation (4.4.2). The default period for a Decision Node's life cycle is months, if the period differs, then one should first convert the period to months before calculation.

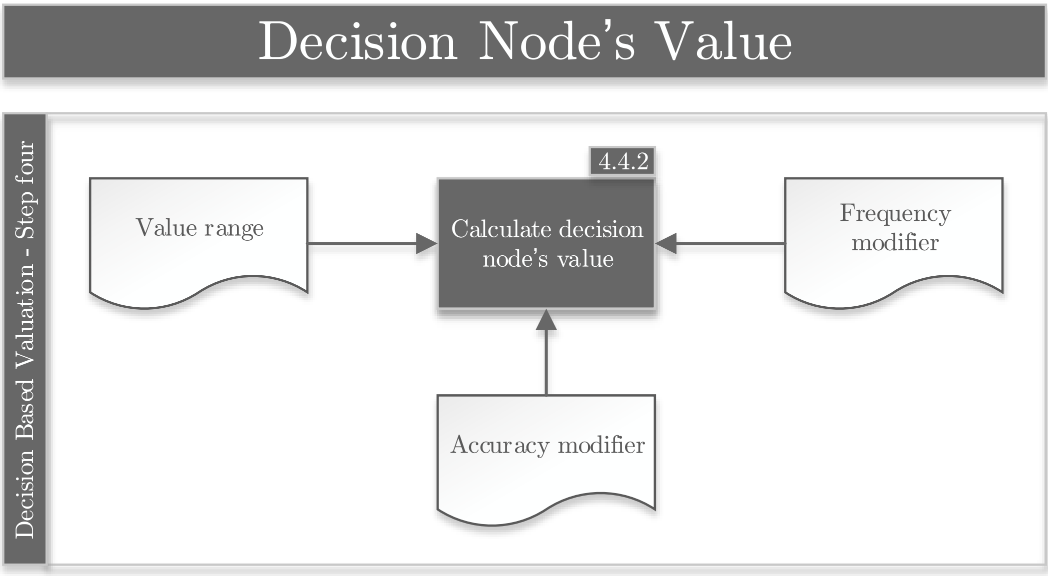
$$A_m = \frac{C_D}{N_L} \tag{4.4.2}$$

The monthly amortization amount (A_m) can then be used to amortize the Decision Node until it reaches zero value and is de-recognized.

4.4.2 Valuing of Decision Nodes

Once a Decision Node has all of its parameters, the value that node is currently achieving can be calculated. This is the fourth step in the Decision Based Valuation Method as seen in Figure 4.6.

Figure 4.6: Decision Node's Value



The first step in calculating the value of a Decision Node is to determine the expected value of the decision it is representing. The expected value V_δ is equal to the decision's average financial outcome range as shown in Equation (4.4.3).

$$V_\delta = \frac{V_{max} - V_{min}}{2} \tag{4.4.3}$$

The quality modifier Q_f is the summation of the accuracy I_A and frequency I_F modifiers as shown in Equation (4.4.4) with a range between 0 and 2.

$$Q_f = I_A + I_F \tag{4.4.4}$$

Thus, the equation for the value of a Decision Node is shown in Equation (4.4.5).

$$V_N = V_\delta \times Q_f \tag{4.4.5}$$

Therefore, the value of the Decision Node is some value between the default value of the decision and the maximum obtainable value with perfect information. The next step is to determine the frequency and accuracy modifiers, these are determined uniquely according to how they were declared in the Decision Node. Starting with the frequency modifier which can be determined as follows.

$$I_F = \begin{cases} 0 & \frac{F_T - \sqrt{(F_N - F_I)^2}}{F_T} \leq 0 \\ 1 & \frac{F_T - \sqrt{(F_N - F_I)^2}}{F_T} \geq 0 \end{cases}$$

Where F_T is the frequency tolerance of the Decision Node, F_N is the required frequency of the Decision Node, and F_I is the frequency of the supplied information. The process of determining the accuracy modifier is slightly more complex than that of the frequency modifier. If the value gain is supplied by the decision node, then the information modifier can be determined with Equation (4.4.2).

$$I_A = \begin{cases} A_{RO} + (A_{RO} - A_I) \times A_{FO} & A_I \leq A_R \\ A_{RO} + (A_I - A_{RO}) \times A_{CO} & A_I \geq A_R \\ 0 & A_I \leq A_F \end{cases}$$

Where A_R is the required accuracy for the Decision Node, A_{RO} is the value percentage gained if the required accuracy is met, A_{FO} and A_{CO} are the percentage value gained or lost respectively, A_F is the accuracy floor of the Decision Node, and A_I is the accuracy of the supplied information.

As stated previously, both I_F and I_A can obtain a value between 0 and 1 however, due to the scope of this study, the frequency modifier has been restricted to be either 0 or 1. Thus the addition of both the modifiers will be some value between 0 and 2. The range of V_N is shown in Equation (4.4.6).

$$\begin{aligned} f(x) \qquad & V_N = V_\delta \times Q_f \\ Q_f = 0 \qquad & V_N = V_\delta \times 0 = 0 \\ Q_f = 2 \qquad & V_N = V_\delta \times 2 = V_{max} - V_{min} \end{aligned} \tag{4.4.6}$$

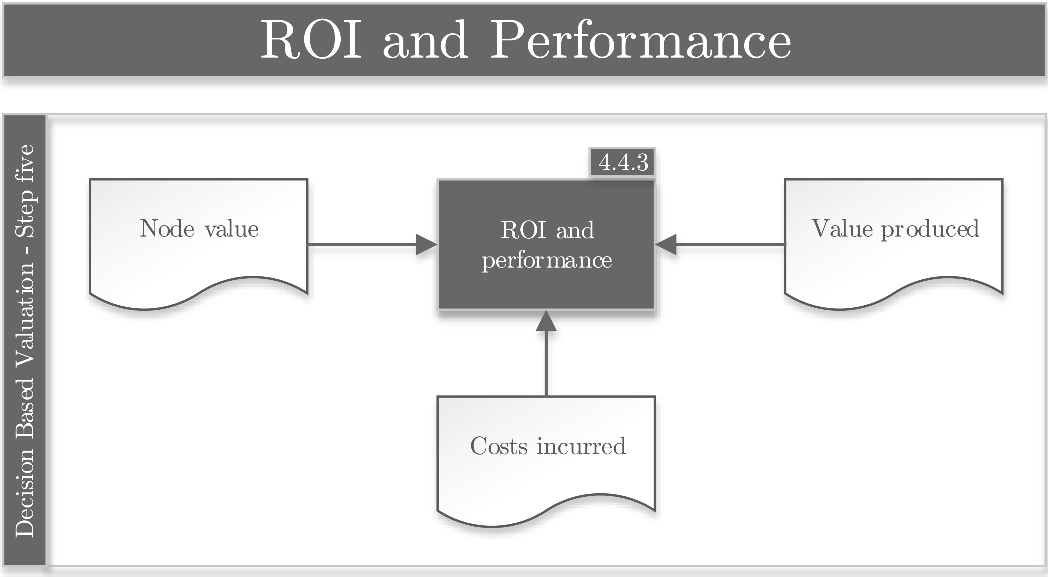
4.4.3 ROI and Performance

Determining the performance and ROI of a Decision Node is the fifth step in the Decision Based Valuation method as shown in Figure 4.7.

The basic description of Return on Investment (ROI) is the generated value over the incurred costs as shown in Equation (4.4.7.)

$$\text{Return on Investment} = \frac{\text{Value Produced} - \text{Cost Incurred}}{\text{Cost Incurred}} \tag{4.4.7}$$

Figure 4.7: Decision Node's Performance



It is important to note that there is a distinct difference between potential value and realized value. This difference is shown in Equation (4.4.8), which highlights the performance with regards to achieving the full value of the Decision Node (V_N) and subsequently the information.

$$\text{Value Performance} = \frac{\text{Value Produced}}{V_N} \times 100\% \tag{4.4.8}$$

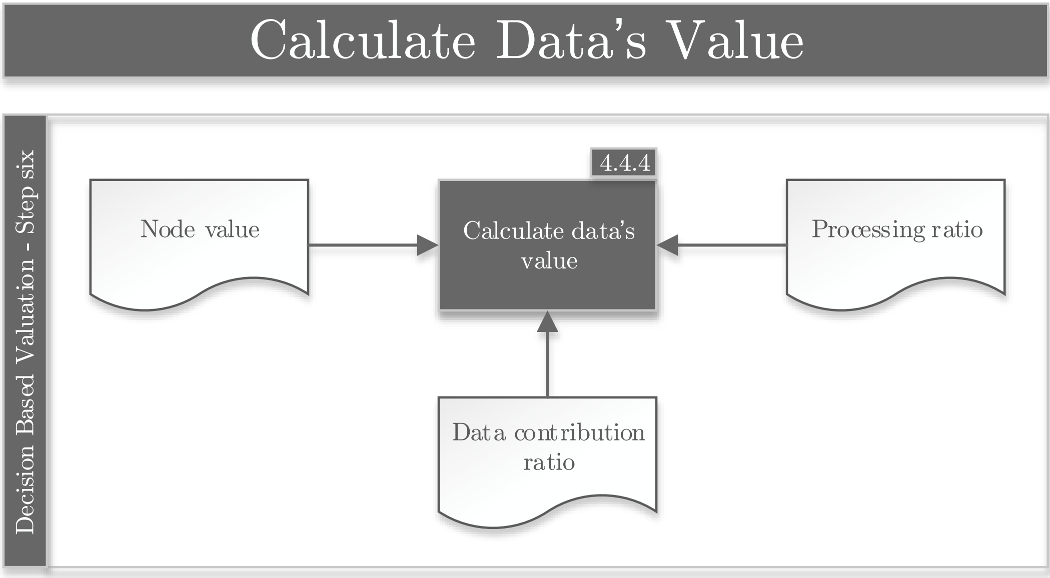
The ROI calculated in (4.4.7) is then used for the evaluation of the Decision Node – whether or not it achieved its required ROI. There are other performance metrics that can be calculated which are excluded from this study; such as value produced per cycle, return on accuracy, and tolerance to frequency variation. These metrics can be useful to organizations when the Decision Node is very costly and needs to be optimized to reduce costs. For instance, if the Decision Node has a high tolerance to frequency variation, the frequency of information can be reduced while retaining near equal value. Such forms of optimizations are beyond the scope of this study however, they will be briefly discussed in Chapter 5.

4.4.4 Distribution of Value

The calculation of the data's value through distribution of the nodes value is sixth and final step of DBV – as shown in Figure 4.8.

Once the value of the node has been determined, the value can be distributed to the information that was provided to it and then to the data.

Figure 4.8: Decision Node's Data's Value



There are two options available for the distribution of value, similar to that used for cost distribution in section 4.3.3.6. The first option is to distribute the value evenly among inputs throughout the value chain until the final value for the data has been determined, taking into account the value gained/lost during processing. The second option is to distribute the value through a weighted system where more significant inputs are given a greater share of the value. This method however requires a greater understanding of the contribution of each input which is often difficult to have. A more likely scenario is a combination of both approaches; in most instances there will be auxiliary information in the node that does not have as much weight as the primary information. In such cases, the primary information would get the majority of the value while the auxiliary information would share the rest. Furthermore, the organization needs to understand the type of information being used according to the classifications detailed in section 4.1.3.

The basic formula for calculating the value contribution of a data source is shown in Equation (4.4.9). The value contribution (V_D) is equal to the Decisions Nodes value (V_N) multiplied by the percentage contribution (V_R) and divided by the processing value (V_P).

$$V_D = \frac{V_N(V_R)}{V_P} \tag{4.4.9}$$

Therefore, there are three parameters required to calculate the data's value contribution: (1) the Decision Node's obtained value, (2) the percentage of the Decision Node's value attributed to the data's value chain, and (3) the value gained through processing according to the data classification as detailed in section 4.1.3. These formulas are summarized below for Type H and Type L data respectively.

$$V_P = 1 + 0.05 \quad \text{to} \quad 1 + 0.1$$

$$V_P = 1 + \frac{t_P}{t_A}$$

The value V_R will be some number between 0 and 1. An example of distributed data value for a value chain which contributed 60% of the Decision Nodes value, and has an equal time processing as it does acquisition is shown in Equation (4.4.10).

$$\begin{aligned} V_P &= 1 + \frac{0.5}{0.5} \\ V_D &= \frac{V_N(0.6)}{2} \\ V_D &= 0.3V_N \end{aligned} \tag{4.4.10}$$

4.5 Chapter Summary

Chapter 4 presented the Decision Based Valuation (DBV) method, the solution proposed by this study. DBV aims to meet the objectives; 2c; 2d; and 4b as listed in section 1.5. These objectives were based on the following aims; *To show that information and data can be valued through the development of a new valuation method* and *To show that there are grounds for information to be financially accountable as intangible assets*. DBV is able to fulfil these aims and objects as discussed below.

DBV borrows the most from the Income Approach and Cost Approach covered sections 2.3.2.3 and 2.3.2.1 respectively. These two approaches gave two distinctly different valuation results for intangible assets. Subsequently, DBV produces both an internal value for information as well as an internal cost. Another important aspect included in DBV is the lifecycle of the information which is similar to physical assets; information, for most practical purposes, has a finite useful life. These elements therefore encompass the entirety of what makes information valuable; the cost to produce it but also the reward for using it over its life time. By including these valuations, organizations are able to easily determine Return on Investment for their data and information systems while having a firm understanding of their value.

DBV is able to value information in part due to the inclusion of Decision Nodes; a core concept of the method. These nodes attempt to mimic certain aspects of physical assets while bringing together established valuation methods. Moreover, it complies with the fundamental principle of information's value; information can only produce value if used. Therefore, Decision Nodes provide a mechanism for organizations to evaluate if certain information is able to produce value. Decision Nodes also act as the foundation for data and information to be considered intangible assets by incorporating the asset criteria as stated by IFRS (Board, 2015).

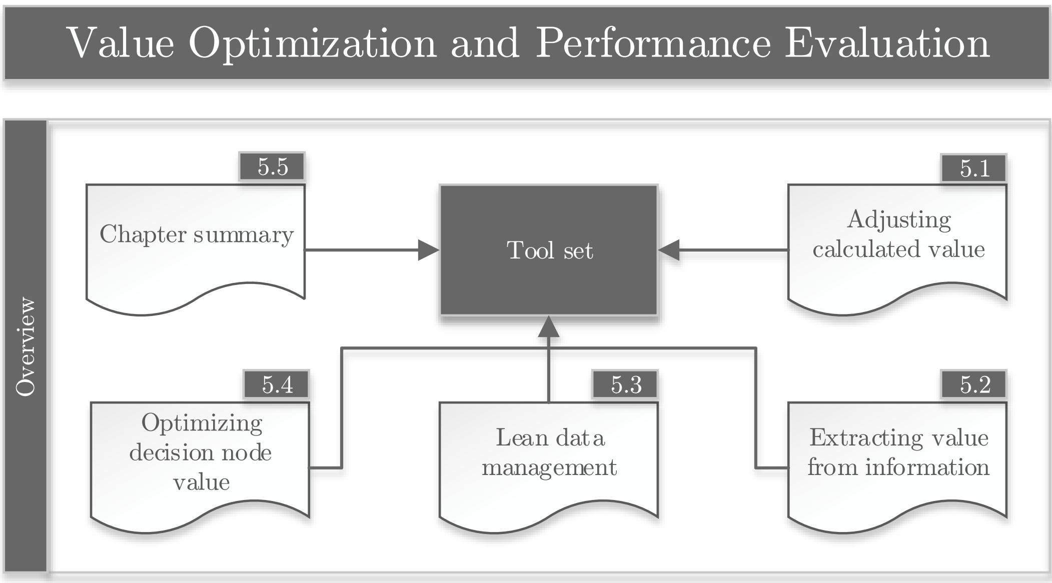
Finally, the value of information is calculated through a set of equations (section 4.4) which aim to capture the core attributes of information that influence its value; decision range value, accuracy, and frequency. These core attributes were discussed and elaborated to provide organizations with a means to determine them. DBV also provides a method to distribute this value for information to its data by using data classification discussed at the beginning of the chapter, thus completing the data value chain.

The following chapter presents the value optimization and performance assessment tools. Even though these tool can be used independently from the rest of the solution, they have been catered towards use with DBV – the next chapter concludes the solution provided by the study.

Chapter 5

Value Optimization and Performance Assessment

Chapter 5 presents basic methods that can be used by an organization to improve their Big Data and information systems; through reducing costs and maximizing value. The presented methods simplify the process of optimization and follow the principles of the valuation framework and Decision Based Valuation (DBV). Furthermore, methods will be presented to help adjust and improve the accuracy of an organization's data and information valuations. Finally, It should be noted that even though the methods presented in this chapter apply to DBV, they are still applicable to most organization's data and information systems and can be applied independently to DBV and the valuation framework.



5.1 Adjusting Calculated Value

Due to the nature of the valuation method, there will often be inaccuracies, and at times assumptions. These issues may occur when:

- It is difficult to separate the value gained from one source of information due to multiple sources being used at once;
- Income earned is difficult to attribute to a single function within the organization;
- The quality of the data does not meet the required level of accuracy or consistency;
- Changing markets and buy/sell prices;
- Industry changes and the implementation of new and untested processes, business models, and equipment; and
- Human error.

Thus it is important to review the valuation results of critical data and compare them to the realized value. Correcting such valuation irregularities may not provide much benefit to once-off decisions however, for Decision Nodes with long lifecycles it can provide better valuation accuracy. This requires looking at historic valuation results, anywhere from one month to one year back, and comparing them to actual realized value. The method presented below is based on the principle of exponential smoothing (Gardner, 1985). This principle is implemented to allow for historic results to affect the value adjustment exponentially less the older they are.

Subsequently, discrepancy in the predicted and realized value should then be accounted for through an uncertainty variable β and Equation (5.1.5). The adjusted valuation V_I^* is calculated using previous month's predicted values V_{I-i}, their realized values V_{I-1}^R, and the calculated β value.

$$V_I^* = \frac{\sum_{i=1}^{n} V_I \times (1 + \beta^i (V_{I-i}^R - V_{I-i}))}{n} \tag{5.1.1}$$

The calculation for β , (5.1.2), requires at least four previous predicted valuations and realized values.

$$\sum_{i=1}^{n} \beta^i (V_{I-i}^R - V_{I-i}) = n(V_I^R - V_I) \tag{5.1.2}$$

It should be apparent that the calculation for β becomes significantly more difficult as more months are used. Thus, it is advisable to use just four months and solve the cubic equation as shown in Equation (5.1.3).

$$\beta^3(V_{I-3}^R - V_{I-3}) + \beta^2(V_{I-2}^R - V_{I-2}) + \beta(V_{I-1}^R - V_{I-1}) - 3(V_I^R - V_I) = 0 \quad (5.1.3)$$

If four previous months are not available, then use Equation (5.1.4) as a approximate solution for Equation (5.1.2).

$$\beta = \frac{\sum_{i-1}^{n}(1 - \frac{V_{I-i}^R}{V_{I-i}})}{n}, \quad n \leq 3 \quad (5.1.4)$$

With the above, a three month value adjustment can be calculated using Equation (5.1.1), which becomes Equation (5.1.5) as seen below.

$$V_I^* = \frac{V_I \times (1 + \beta(V_{I-1}^R - V_{I-1}))}{3} + \frac{V_I \times (1 + \beta^2(V_{I-2}^R - V_{I-2}))}{3} + \frac{V_I \times (1 + \beta^3(V_{I-3}^R - V_{I-3}))}{3} \quad (5.1.5)$$

5.2 Extracting Value from Information

A problem many businesses are facing is how to create value form their information. This falls back to the question of how does information become valuable? Concluding from this study it should be apparent that information gains value when it is used, specifically it gains value depending on what it is used for. Thus, by using the top-down approach alone, an organization will be able to extract value from their information and identify the information that has no value. However, there is still room for masses of data to produce value even though it is not destined for a decision; through data mining and other data analytics. But data mining activities and the like can often yield no usable results, therefore how does an organization approach the situation where they have lots of data but they are not generating value from it?

5.2.1 Applying Top-Down Approach and DBV

The first step should always be to apply a top-down approach and generate a value chain. After this value chain has been created, it can be analysed using DBV. The combination of these processes should then yield a more definite value for certain information. Although, it might not always be possible to perform a top-down approach and DBV, especially when the data was collected without intention from the start and was just collected because the organization was able to. If that is the case, then an organization needs to structure their approach to the data as detailed in 5.2.2.

5.2.2 Structured Mining

If an organization is facing masses of data that was never destined for any decision or activity in the organization, then it is difficult to determine its value. This scenario often appears within businesses where protocols and procedures generate data that is never used and only ever archived. With the improvement in technology and data mining techniques, organizations are beginning to mine this data to extract value from it. However, they often don't know what they are looking for and generally rely on correlation and patterns. Even though this study does not provided a structured method for these situations, the same principles can be applied to help steer data mining activities. This can be done by creating outcomes for the data mining to help improve the chances of actually finding useful data that can be processed and used. Thus, an organization can ask itself the following questions to help with the process of generating outcomes.

1. Are there processes within the organization that are sub-optimal and need improvement?
2. Are employees wasting a significant amount of time on administration?
3. Are employees wasting a significant amount of time trying to obtain various sources of information?
4. Are there production capabilities that need to be optimized to improve their output?

Once an organization has answers to these questions, they can then start determining specific attributes that they would need to solve and optimize the above situations. By doing so, they are essentially creating the decision and information requirements of a Decision Node. This can then be used to structure the approach and determine what data they should look for during their data mining procedures. This will allow organizations to provide specific data that they require that will then increase the quality of the data mining substantially; data analysts will now be able to program the search specifically for what the organization is looking for, including variants of it. If this data is successfully identified, then the organization will have the tools it needs to generate value from it.

5.3 Lean Data Management

Part of the performance model is removing unnecessary and low value data and systems from the overall Big Data system. To do this, the simple concept of lean management can be applied to an organization's Big Data system. Lean management, according to lea (2014), is an idea aimed at maximizing value while minimizing waste. They go on to say that it is not a process but a way of thinking and can be applied to any industry or process. Subsequently, this way of thought can be applied to Big Data.

A key aspect of lean thinking is removing waste, in this case, data that does not generate value for its organization. By removing wasteful data and its capturing and processing systems, the overall performance of an organization's Big Data system will improve. Furthermore, by simplifying the data and its sources it allows the organization to focus on the important data and improving its quality and consistency.

The lean principle can easily be applied to data systems as follows.

5.3.1 Data Accountability

The first step in lean system construction is ensuring that all data that is collected has a purpose. This requires that, attached to a request for data collection, is a form that details the data's value chain and desired decision output. This alone will help eliminate superfluous data collection as well as speed up the more in-depth analysis of the data at a later stage. Furthermore, this value chain report will help different business functions and personnel to easily identify the reason and destination of data acquisition systems. This "transparency" provided by the value chain report also aids in technical reporting of data systems and would prove invaluable when contracting outside businesses.

The components of a basic value chain report are:

- Data being collected (source, location, frequency etc.);
- Destined decision;
- Information output format;
- Acquisition hardware;
- Software requirements; and
- Storage location.

The above act as the core of a value chain report. An organization can further augment the report with supervisor details, the person who requested the data and so forth. It should be noted that this report does not act as DBV report, the latter being far more in-depth and requiring a significant increase in details on the data, information, and its decision.

Once a template has been created to suit the organization, this process can be completed efficiently and quickly, saving lots of time and money in the long run.

5.3.2 Auditing

To help eliminate data that is no longer needed (or never was), an organization should conduct regular audits on their data and information system. This should also be the first step when going through DBV for the first time. Furthermore, during the audit, data accountability should be practised and all data should be given a value chain. If the audit is not the start of the first DBV, value chains should be reassessed to determine if the output decision is still being made and then whether it is still using the same data. As time goes on, better and more accurate sources of data can become available rendering older sources redundant.

To help facilitate this process of auditing, an organization should make use of an IAR, as detailed section 4.2.4.1. This register would speed up the auditing process significantly and should be implemented even if the DBV method and its Decision Nodes are not used. The IAR also forces organizations into making decisions on what information and intangibles are of significant value.

5.4 Optimizing Decision Node Value

At a certain stage of an organization's Big Data system, there will be key value chains in place delivering data to the organization for important decisions. However, there is still room to optimize the existing systems to reduce cost and improve value (through accuracy and consistency).

5.4.1 Value Chain Branching

Value chain branching is the practice of branching at a certain stage of the value chain to generate information for another decision. For example, looking at Figure 5.1 (a simple value chain) one could branch off at the information stage and use that information for another decision. This would then significantly reduce the costs for the second decision. Thus, the value has been increased while maintaining the costs. Another example would be to branch off at the data stage, then go through a separate processing stage for another decision. This would still reduce the costs for the new decision but it wont completely remove them as was the case with the first branching option. However, this still generates more value for the initial value chain. These two options are the core branching methods available.

A primary value chain with the two aforementioned branches is illustrated in Figure 5.2. This method of value chain branching is an effective means of increasing the value of an organization's Big Data systems while maintaining costs. Thus, the more branching of value chains within an organization's Big Data systems, the greater its performance will be.

Figure 5.1: Simple Value Chain

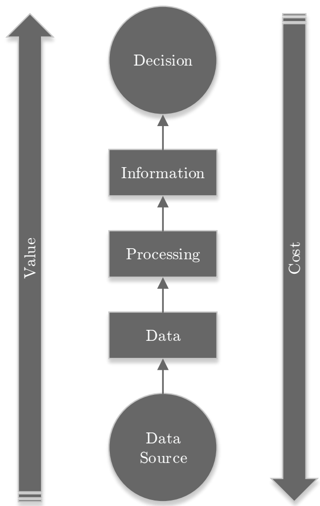
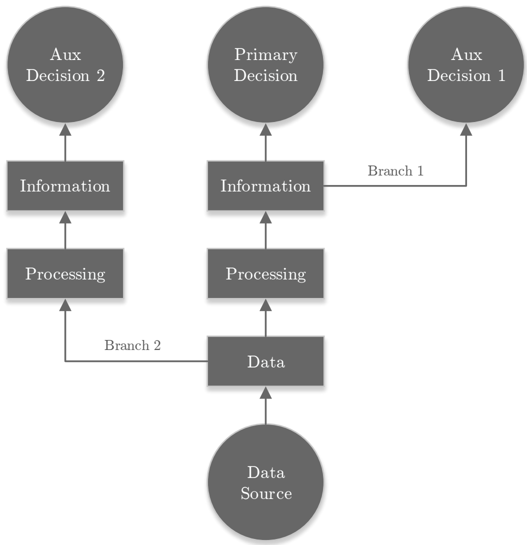


Figure 5.2: Branched Value Chain: *A simple value chain with two auxiliary branches added to it to increase the overall value*



5.4.2 Performance Metrics

Olsen et al. (2007) mentions that even though 49% of senior executive said they relied on intangible assets, only 5% track their performance. Furthermore, Marr et al. (2004) and Lev and Daum (2004) state the importance of measuring knowledge assets and intellectual capital, identifying the link of these intangible assets with global competitiveness. It is apparent that assets should be measured and tracked, subsequently a few key performance metrics have been created for use on Decision Nodes.

Once a Decision Node is operational, these performance metrics can be calculated to help optimize and track aspects of the Decision Node to improve its performance. The basic performance metric, Return on Investment, was covered in section 4.4.3 and acts as the fundamental reporting mechanism when seeing how well a Decision Node is operating. However, there are many more metrics that can be calculated that can help provide insight on various aspects of the Decision Node. Unfortunately, these metrics are outside of the scope of this study and will only be briefly discussed to highlight the possibilities available to organizations. It is important to note that these metrics are situational and are dependant on the type of data and information being collected. Furthermore, it would only be worth calculating these metrics on Decision Nodes that have a significant potential value gain. Otherwise, the potential gains from optimizations will be so minimal that it would not be worth the effort to calculate and analyse. There is also that fact that optimization has a greater effects Decision Nodes with longer lifecycles than those for once-off decisions.

5.4.2.1 Value versus Frequency

One metric that an organization can investigate is the change in obtained value versus the change in the frequency of the supplied information. This can highlight the relationship between frequency and value to the organization. This has a few advantages, for instance; if an organization identifies that the value of a Decision Node is only weakly linked to the frequency, they can save costs and reduce the frequency of the supplied information. This can lead to various alterations to the frequency at which data is collected and supplied, leading to various cost savings or value improvements.

5.4.2.2 Value versus Accuracy

When constructing the Decision Node, organizations need to supply the relationship between accuracy and value for the information being collected. This relationship will, in most cases, be an approximate. Being able to refine the true relationship can lead to better, more accurate valuations as well as improved strategic decision making.

5.4.2.3 Value Decay

Another useful metric is the value decay of the Decision Node over its lifecycle. Information has a tendency to devalue over time, especially when other sources become available. Thus, being able to see the trend in value can help organizations determine when to terminate a Decision Node. Knowing how the value changes overtime also helps the organization to determine the amortization schedule and the amortization rate; linear versus non-linear. This will assist organizations to accurately report on the profits earned versus expenses incurred as is the case with physical assets.

5.4.2.4 Time to Completion

Although not incorporated in the current Decision Node structure, the time from data acquisition to supplying information for the Decision Node can provide useful insights. This is especially true when determining the cost to supply the information when labour contributes a significant portion of the overall costs of the Decision Node. This can also provide management with lead times and help with scheduling of employees for projects.

5.5 Chapter Summary

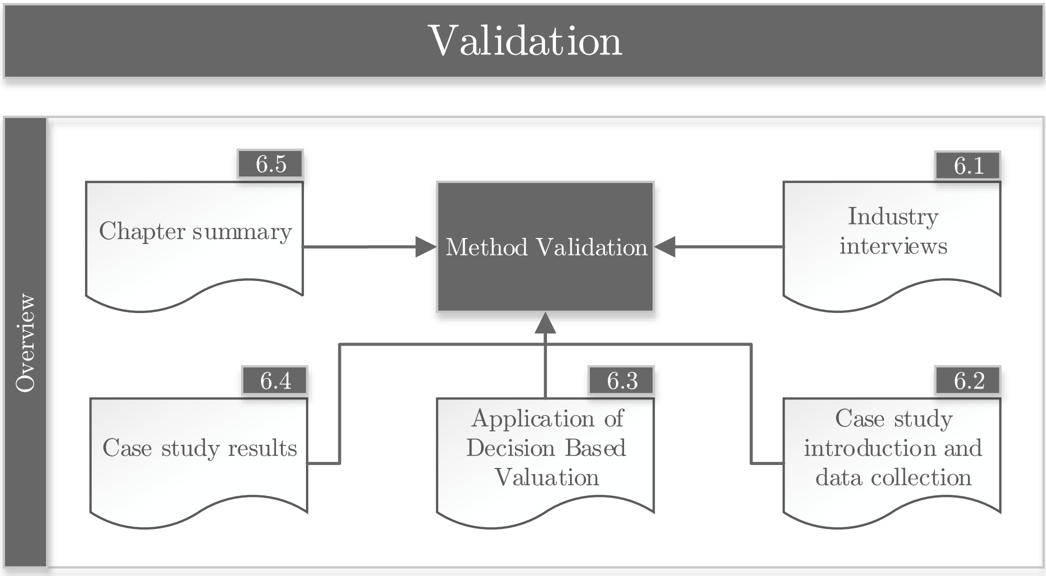
Chapter 5 presented tools and methods that aim to: (1) provide organizations with tools in which they can assess the performance of their data and information systems, and (2) detail a set of methods that can be used to optimize the value of these same systems. Furthermore, these tools and methods have a strong connection to DBV while still being able to be implemented independently to it. These tools directly address the objectives of the study, namely 5a and 5b. This chapter does not provide a complete set of tools for organizations to take full control over their data and information systems. However, the chapter still meets the aim of objective 5; *Improve the management and control of data and information's costs and value* as discussed below.

The first tool provided was for adjusting the value of information according to predicted and realised value. This tool is used when an organization implemented DBV but were unsure of the accuracy of the information provided. This tool is also only beneficial if the Decision Node being assessed has a long lifecycle. Following value adjustment, methods for extracting value from data were provided, these aim to help organizations with existing data, but with low utilization, to extract additional value. Next, the concept of lean data management was introduced: it provided organizations with a method to remove wasteful data, but importantly this tool highlights that not all data is useful and worth collecting. Ending off this chapter were value optimization and performance assessment tools specifically catered towards DBV. As previously mentioned, this chapter was successful in meeting the objective of improved information management and concludes the solution presented in this study. The next chapter will present the validation of the method and its need through interviews and case studies.

Chapter 6

Validation

Chapter 6 presents several interviews conducted with industry professionals as well as two case studies at a Middle East utility company. These are used to validate the need for Decision Based Valuation (DBV) and its success as a proof-of-concept. The results and conclusions of the interviews and the case studies are provided, followed by a discussion of whether DBV achieved its aims.



6.1 Industry Interviews

The following statements are provided by industry professionals who deal with assets (physical and intangible) and information on a daily basis. Each of the interviewees were asked to provide their interpretation and understanding of the value of information. These statements serve two purposes: (1) to help validate the need for an information valuation method such as the one developed in this study, and (2) provide valuable insight as to what makes information valuable. The latter helps ensure that the method captures the true value of information as seen by industry. These statements come directly and unmodified from the interviewees and are presented in no particular order.

6.1.1 Interviews

The statements are as follows.

Manager in Strategy and Operations
PwC South Africa

> "*Selling solutions to clients have become increasingly difficult without practical examples and evidence of the problems they face. In the day to day activities, clients often do not recognize the problems they face in making informed business decisions. Data analytics have become an increasingly important tool in the consulting space, often used to "clinch the deal". Problems in business processes specifically are effectively highlighted by showing clients real-time inaccurate data produced by their own processes. As clients are dependent on accurate and up-to-date data to make various business decisions (budgeting, capital investments, business process engineering, etc.), effective and accurate data analytics have become an integral part in making informed business decisions.*"

Manager Asset Management - Expert Services
Tetrapak Sweden

> "***How do you determine if data is worth collecting?***"

> "*In the TP supplier area we are collecting information that is needed for risk management I am not aware of a specific process to determine what type of data to collect. Anyhow we collect financial information and also TP share of our supplier business. When it comes to our customer information we collect information as a part of business intelligence to be used for finding business opportunities also information that can support our product development*

and product maintenance. We also collect financial and strategic information as a part of our risk management and account planning. As of today I am not aware of any formal process determine which data to collect furthermore we need to understand better how we transform data in to knowledge and use this knowledge better. In addition to this data we also collect data regarding demographic information, global economics and consumption patterns etc, as an input to TP corporate strategy"

"***What makes information valuable to you as an organisation?***"

"*All information we can turn to knowledge is of value*"

"***How would a better understanding of this value affect how you handle said data and information?***"

"*We probably would invest more in the area of data gathering and analysis if we could articulate the business value in a sharper way*"

Director in Asset Management
Gaussian Engineering

"*Since the late 80's and early 90's organisations have amassed data. How much? Well probably enough to proverbially sink a small country. So much data that we will probably never be able to quantify it. The question we should be asking is how much data is used and what for. Working in a number of industries a common theme is that operational data is used for operational purposes. Flow, temperature, speed etc measures are used at daily and weekly meetings. Tonnages to measure performance and throughput. Financial information is collected to be reported on monthly and yearly. The cost of collecting and storing is never really considered as they are seen as 'funny' money costs. The operator or engineer are on site anyway and we have the servers and cloud storage to operate the business. Reviewing old data is usually done in reports showing trends or when a failure or problem occurs. The value of the data is transactional as it helps solve the problem before being sent back to the server.*

Very few organisations use the data collected strategically. There are some, but the majority don't understand what is available and its value. Taking data and transforming it in to informed decisions *isn't a common practice. So if we do not value the data it will never have a value. To be strategic and sustainable organisations must realise that the value in the data lies in how it is used. The cost of the data management process should be much lower than the benefits derived from having it and using it effectively. Until we realise what we have we will never fully use it.*

Internet based companies understand the value of data, search for a holiday to Dammam and every non travel related page has an advert of accommodation in Dammam, or flights to Dammam subtly placed on the web pages. How does the BBC news page know to show adverts specifically focussed on my current interests? They collect and use my browsing history, preferences and likes. The value to me is that I don't need to search for stuff and the conversion rate for me to see and select something is high."

Director in Asset Management
Gaussian Engineering

"*Data from an asset management point of view is fundamental to informed decision making. Without it you are simply making best judgement decisions, which experience has shown can be contrary to the purpose one is trying to achieve. With data that you can depend on through simulation, modelling, comparing and a variety of other methods make better informed decisions. In addition data is the basis for developing any basis for knowledge.*"

Director in Power and Utilities
PwC Middle East

"*In summary, the cost of data (on a volume basis) has become virtually free. The value of data analysis - especially fast data analysis - has risen.*

Over the past three decades we have become accustomed to and take for granted more and more ubiquitous machine data, and have become better and better at turning data into enterprise value.

Going back four decades production rates and tighter product quality tolerances demanded faster, more accurate and more consistent machine control. Ever faster and more powerful processors, compact and reliable solid state sensors and standardized networking communication protocols enabled this shift from human-operated electro-mechanically controlled machines to automated networked computer controlled machines. Productivity, yield and prod-

uct quality gains in the seventies and eighties were enormous. Compared to the previous generation of machines, process and performance data was now much cheaper to gather automatically and the application of statistical process control became widespread.

As these automated networked computer controlled machines became more and more ubiquitous, the cost of storing their interval data for later analysis became cheaper and cheaper. Corporations could mine their process and performance data sets to find trends and make statistically strong correlations from previously uncorrelated data to support management decision making, leading to incremental improvements in productivity, yield and quality. By the late nineties most corporations could exploit their on-premise process and performance data, and many could do so corporate-wide. Data mining and its value to support management decision making became greater, driving up the volume of data locked in local and corporate servers.

With the advent of large global fibre data networks, starting in the late nineties, data communication costs plummeted. Corporations could now collect virtually unlimited process and performance data from their distant plants or sites cheaply.

Going into the 2000s the volume of process and performance data flowing into corporate data centers drove new IT innovations for on-line analytical processing (OLAP). More corporate data could now be analyzed cheaper than ever before. The value of analysis was by now widely understood. To remain competitive the next improvements in data analysis needed to come from speed.

In this decade two technology developments continue the trend of allowing more data to be processed, and to do so faster than ever before. More efficient data storage architectures that work across large networks (such as in the Cloud) and in-memory data processing have shrunk big process and performance data analysis from hours to seconds, achieving near real-time performance. In most industries (renewable power generation and distribution or energy futures trading for example) the speed of big data analysis mission critical."

Director in Global Strategy Group
KPMG London

"*Big data itself is not new and businesses have been gathering a wealth of digital information for many years. Business lead-*

ers all understand the value of hard data and seek to apply it in improving their decision making and business performance. However, the application has often been constrained by siloed, backward looking data sets, that typically present "snapshots" of particular operational, commercial or financial performance on a rolled up basis. These views, while useful in generating insight, often mask the underlying drivers of cost or value. This is rapidly changing. With increasing digitisation and the development of more powerful data interrogation tools, businesses in all sectors are now exploiting the abundance of information to gain advantage through more granular insight and, now, real time decision support. Amazon's customer analytics are well known. McLaren analyses terabytes of data in real time during a single 90 minute formula one race, finding advantage in races where milliseconds can mean a podium finish. Utilities, often with millions of customers, are now able to identify and help individual customers reduce energy usage, especially during peak grid hours. Perhaps more importantly, they are now much better placed to identify customers that may be vulnerable and in need of support. Understanding population sentiment, improving healthcare provision, better meeting customer needs. The list is almost endless. It is clear that the application of big data, when used to support opportunity identification and decision making, has the potential to improve many aspects of business, science and life. The full potential is yet to be reached."

Asset Management Consultant
Gaussian Engineering

"*The decision makers issue often stems on the haunting statement "you don't know what you don't know". This lurking fear drives you to maintain the mantra of more is better when it comes to collecting data and information. Unfortunately, this often results in a delay to decisions as we await more clarity on a situation through the lens of information, with issues compounded by not being fully aware of what lens that information is highlighting at the detriment to other information. Having the ability to differentiate data and information at hand based on experience or systems in place allows pertinent weight to be applied to available information for analysis when making a decision.*"

"*Thinking on this, information to me is the foundation from which a decision can be argued and justified based on the situation and environment the decision was made. In this way you are able to revisit decisions and assess your decision methodology and critique* *for improvement in the future. The greatest advantage to having a well structured and properly weighted information framework at hand is that regardless of the outcome of a decision you are able to pull new lessons from them, and ensure that every decision made is a learning curve that gives an advantage to yourself or your business in the future.*"

6.1.2 Summary

In light of the interview responses, there is apparent consensus among the industry professionals that it is now easier than ever to collect large quantities of data. This has resulted in many companies possessing terabytes of data that they are often not fully utilizing. The interviews highlight another important factor; even though companies have access to a large variety of data and processing methods, they are struggling to identify and harness data's value. The interviewees also agree on the importance of data and information analytics, where the benefits of utilizing Big Data were considered numerous.

Seven conclusions were ascertained from the statements provided, namely:

1. There is a relationship between data's volume and variety and its value;
2. Organizations are currently ill-equipped to value the data they collect;
3. Organizations are not harnessing the full potential of Big Data and information;
4. Few organizations make strategic decisions when it comes to their data;
5. Information is everywhere and easily accessible to all;
6. Information is critical to strategic decision making; and
7. There is a disconnect between the cost of data and information systems and the value they provide.

The aforementioned conclusions validated the need for the DBV method, which directly addresses at least three of the issues raised in the interviews. Furthermore, there is partial validation of the classifications of data, highlighting that certain types of data scale with volume. The interviews further validate the need for a more strategic view of data and information which chapters 3, 4 and 5 have directly addressed. The toolset provided by these chapters help organizations understand how value is created from data and how to optimize their data and information system(s). In addition, the toolset facilitates a greater awareness of the cost of data for organizations, thus allowing them to provide superior motivation for their decisions on data collection.

In closing, the industry interviews illustrated that the method and toolset developed in this study are indeed needed by organizations. Furthermore, these methods and toolsets directly address some of the major issues organizations are currently facing with Big Data. The insight gathered from the interviews meet the objectives as stated in section 1.5, namely: 3a; 3b; 3c; and 4c. The following sections will present the case studies, beginning with an introduction and application of the valuation framework from Chapter 3.

6.2 Case Studies Introduction and Data Collection

The aim of these case studies is to identify whether DBV, as a proof-of-concept, was successful. Furthermore, these case studies help identify any flaws in the method and, subsequently, which areas require further development. In choosing companies for the case studies, the following criteria had to met:

- Collects large volumes of data equivalent to that of Big Data;
- Financially benefits from this data to some degree;
- Has specialized hardware and software to handle Big Data;
- Conducts business intelligence and/or data analysis of the stored data;
- Maintains data archives; and
- Collects and stores superfluous data.

The organizations selected, and how they met these criteria, is discussed below.

6.2.1 Selection of Case Studies

Important: Due to the nature of the information disclosed in the case studies, the names of the parties involved have been changed.

Two case studies were conducted in the Middle East at one of its major utility companies, one of which was provided by a consulting company in the region. Due to the sensitivity of the information disclosed for this study (financial information and otherwise) both organizations have chosen to remain anonymous.

These companies frequently make decisions that require information from various sources however, this information comes at a cost – often a substantial one at that. The high level employees at these companies note that these decisions are often made with complete disregard to their cost implications, as well as the value of the information being collected. By applying DBV at these companies, their high level managers were able to receive insights to some of the information they collect and use. Furthermore, these companies deal with Big Data on a regular basis; using data analytics and processing tools on this data. Both companies also use data centres and servers to handle the large volume of data they collect. Therefore, the criteria for the case studies are met by both organizations.

6.2.2 Data collection

The data used for the case studies was collected on-site from the organizations in the Middle East. Furthermore, strict confidentiality was adhered to when conducting the case studies. The type of data collected was:

1. Cost of labour;
2. Cost of equipment;
3. Cost of utilities;
4. Value of the decision;
5. Metrics to define the required information: frequency, accuracy, lifecycle.

The aforementioned data came in various formats such as:

1. Excel formatted time sheets filled in by employees,
2. Contracts,
3. Rates and billable documents,
4. Microsoft Project files,
5. Supplier quotes, and
6. SAP exports and reporting.

The above data was then analysed to extract the required information for DBV.

6.3 Application of Decision Based Valuation

The first step to applying DBV was to use the valuation framework (Chapter 3) to identify a data value chain within the organization. First, management was asked about frequent decisions that they make, preferably one that would provide a significant reward to the organization if made effectively. Once a decision was identified, the information it required was determined, as well as who would be providing this information. During this process, the second case study was identified; they were to provide the information used in the first case study. After the required information had been determined, the next steps of the valuation framework were followed; identifying processing techniques and the data required. The data was processed with a combination of human processing and the use of Microsoft Office – SAP was used to export data but was not used by to do the processing itself.

After identifying the required data needed by the information, its sources and collecting methods could be determined. In the case studies, the data was collected from SAP and shared networks while being sourced from employee time reports and others as mentioned in section 6.2.2. The identification of the data value chains provided sufficient insight and understanding to implement DBV. After applying the valuation framework, the DBV process was followed as outline in Appendix A.

As previously stated, the second case study was identified while determining the data value chain of the first. The result of which is that same decision is viewed from two different perspectives; illustrating two distinct scenarios and information needs. It further illustrates how information's value is unique and can shift as described in chapter 2. Moreover, it illustrates that organizations can incur different costs, and receive different value, from the same information. This fact is made clear through the use of the information by the organizations; in the first case study it was to fulfil an agreement and in the second it was to make a strategic decision. A more detailed look at each case study's decision is provided next.

6.3.1 Case Study One

This case study is conducted from the perspective of the consultant having to choose the right team composition to handle a deliverable, and delivering it on time to the client. The deliverable was information required for the second case study. Once again, this illustrates how information's value changes and is dependent on the organization use of it.

6.3.1.1 The Decision Node

Step 1

Step one of the Decision Based Valuation method is the construction of the Decision Node as detailed in section 4.2. The following are the attributes of the Decision Node for the first case study.

The Decision

How many resources should be assigned to determine whether the client should use smart or sequential asset tagging system and the documentation thereof (deliverable project).

Information Required

The following information is required to make this decision:

1. The cost of labour options;
2. The return on the project;
3. The time to complete the project; and
4. The penalties for completing it late.

The information should be presented on a linear graph showing the cost of the project, including resources, over time for utilizing either one or two employees.

Value Range

The default decision is to use two engineers for the duration of the project to ensure that the deadline is met. However, some combination of one or two engineers – on and off – can be used if adequate information is provided. These engineers come at a cost of \$150 per hour and work 8 hour days. It is estimated that the project will take 12 weeks. Thus the minimum potential value is \$0 and the maximum potential value is $\$150 \times 8 \times 5 \times 12 = \$72,000$

It is assumed that if more engineers are placed on the project than are needed, the project will not necessarily be done sooner. This assumption can be attributed to two reasons: (1) if the engineers aren't pressured by large quantities of work, they will just work slower, and (2) a lot of the necessary information requires inputs from third parties and therefore there will still be a waiting period for that information; consequently there will be a lot of paid down time of the engineers.

Lifecycle

This decision will repeat over a 3 month period.

Frequency

Once a month with a tolerance of 1 week.

Accuracy

The required accuracy for all cost related information is 100% and for time related information it is 80%. There is no ceiling or floor for cost information as 100% accuracy is required. For time information, the floor is 60% with a loss of 5% in value per percent accuracy lost and a gain of 1.5% for each percent accuracy gain, with a ceiling of 100%.

Maximum Cost

The ROI option was selected.

Require ROI

The ROI should be no less than 20%.

6.3.1.2 Cost

Step 2

The second step of the Decision Based Valuation method is determining the cost of the Decision Node as described in section 4.3.3.

There were no hardware or software costs as the equipment and its software were already paid off. The labour cost is for the manager who does the calculation and makes the decision. The utility costs include transport to and from the site as well as lodging and per diem costs.

This information will take an estimated 3 hours to compile, therefore the following costs apply.

Labour:

Manager: $\$213/\text{hour} \times 3 = \$639$

Utilities:

Per Diem: $\$65/\text{day} = \$\frac{65\times3}{8} = \$24.38$

Lodging: $\$114.74/\text{day} = \$\frac{114.75\times3}{8} = \$43.03$

Total

The decision will be repeated three times, therefore: $\$706.40\times3 = \$2,119.21$

6.3.1.3 Calculations

After the Decision Node is created and its cost is determined, the value for the node can be calculated as shown in section 4.4.

Amortization
Step 3

The calculation of the Decision Node's amortization is the third step (section 4.4.1) of DBV and is calculated as follows.

The depreciable cost is calculated as shown in Equation (6.3.1), noting that in this case, there are no asset costs that are already being depreciated.

$$\begin{aligned} C_D &= C_T - C_{ND} \\ C_D &= \$2,119.21 - \$0 \\ C_D &= \$2,119.21 \end{aligned} \tag{6.3.1}$$

The amortization schedule is then calculated as shown in Equation (6.3.2).

$$\begin{aligned} A_m &= \frac{C_D}{N_L} \\ A_m &= \frac{\$2,119.1}{3} \\ A_m &= \$706.40 \end{aligned} \tag{6.3.2}$$

Valuing
Step 4

The fourth step to Decision Based Valuation is determining the value of the Decision Node as detailed in section 4.4.2. First, the calculation of the value range is done as shown in Equation (6.3.3).

$$
\begin{aligned}
V_\delta &= \frac{V_{max} - V_{min}}{2} \\
V_\delta &= \frac{\$72,000 - \$0}{2} \\
V_\delta &= \$36,000
\end{aligned}
\tag{6.3.3}
$$

Then, the frequency modifier is calculated.

$$
\begin{aligned}
I_F &= \frac{F_T - \sqrt{(F_N - F_I)^2}}{F_T} \\
I_F &= \frac{7 - \sqrt{(30 - 31)^2}}{7} \\
I_F &= 0.86 \geq 0 \\
I_F &= 1
\end{aligned}
\tag{6.3.4}
$$

Next the accuracy modifier is calculated as seen in Equation (6.3.5), where $A_{RO} = 100\% - (100\% - 80\%)(1.5) = 70\%$.

$$
\begin{aligned}
I_A &= A_{RO} + (A_I - A_{RO}) \times A_{CO} \\
I_A &= 70\% + (90\% - 80\%) \times 1.5\% \\
I_A &= 85\% \ \text{ or } \ 0.85
\end{aligned}
\tag{6.3.5}
$$

Using the above modifiers, the quality factor is calculated as shown below.

$$
\begin{aligned}
Q_f &= I_A + I_F \\
Q_f &= 0.85 + 1 \\
Q_f &= 1.85
\end{aligned}
\tag{6.3.6}
$$

Finally, the value that the Decision Node should obtained is calculated in Equation (6.3.7).

$$
\begin{aligned}
V_N &= V_\delta \times Q_f \\
V_N &= \$36,000 \times 1.85 \\
V_N &= \$66,600
\end{aligned}
\tag{6.3.7}
$$

Performance
Step 5

Step five of Decision Based Valuation is the calculation of the performance of the Decision node (section 4.4.3). The performance of the Decision Node is calculated using the equation below.

$$
\begin{aligned}
\text{ROI} &= \frac{V_N - C_T}{C_T} \\
\text{ROI} &= \frac{\$66,600 - \$2,119.21}{\$2,119.21} \\
\text{ROI} &= 3,043\%
\end{aligned}
\tag{6.3.8}
$$

When comparing the estimated value to the actual value, the disparity between the actual completion time of the project versus the estimated time was assessed. It was found that it took a week longer than anticipated to complete the project, thus costing an additional week of the engineer's time. Due to the deadline still being met, no penalties were issued. Thus, the additional week of work costs $\$150 \times 8 \times 5 = \$6,000$. This resulted in the net saving of $\$72,000 - \$6,000 = \$66,000$ for the project and is used in equation below.

$$
\begin{aligned}
\text{Value Performance} &= \frac{\text{Value Produced}}{V_N} \times 100\% \\
\text{Value Performance} &= \frac{66{,}000}{66{,}600} \times 100\% \\
\text{Value Performance} &= 99\%
\end{aligned}
\tag{6.3.9}
$$

Distribution of Value
Step 6

The final step of DBV, as detailed in section 4.4.4, is the distribution of the Decision Node's value to the information's data.

The bulk of the value came from the information on how long employees take to complete a project. It should be noted that these values are somewhat subjective and often just an estimation of the true value of the data. However, if one is to analyse how the value increases and decreases, it will be obvious that the aforementioned data does in fact contribute the majority of the value of the information. The data sources for the various information requirements are as follows – the order will be maintained for the rest of the calculations and is represented by the numbers one through three (one data source is repeated).

1. The cost of labour options - *Employee Billables*;
2. The return on the project - *Project Contract*;
3. The time to complete the project- *Employee Time Sheets*; and
4. The penalties for completing it late - *Project Contract.*

Subsequently, the percentage contributions V_R are:

1. *Employee Billables*: $V_R^1 = 0.2$
2. *Project Contract*: $V_R^2 = 0.4$
3. *Employee Time Sheets*: $V_R^3 = 0.94$

The data is classified as Type L data as per the classifications in section 4.1 therefore, the equations for Type L data are used. The processing time took approximately 20 minutes and the acquisition of the data took approximately 30 minutes. These times result in a V_P calculation as shown in Equation (6.3.10).

$$\begin{aligned} V_P &= 1 + \frac{t_P}{t_A} \\ V_P &= 1 + \frac{20}{30} \\ V_P &= 1.667 \end{aligned} \tag{6.3.10}$$

Subsequently, the data's value contributions are calculated as shown below.

$$V_D = \frac{V_N(V_R)}{V_P}$$

$$\begin{aligned} V_D^1 &= \frac{\$66,600(0.02)}{1.667} \\ V_D^1 &= \$799.20 \\ \\ V_D^2 &= \frac{\$66,600(0.04)}{1.667} \\ V_D^2 &= \$1598.40 \\ \\ V_D^3 &= \frac{\$66,600(0.94)}{1.667} \\ V_D^3 &= \$37,562.40 \end{aligned} \tag{6.3.11}$$

As can be seen in the data's value distribution, the majority of the value came from the *Time Sheets* as well as processing. Since the processing was done by an employee, it emphasises the importance of choosing the right person to do the job.

6.3.2 Case Study Two

The second case study is based on the perspective of an organization using the services of a consultant to produce an information output. This information output will then used by the organization to invest in a certain asset tagging system. The process followed when applying DBV is the same as with the first case study. Therefore, for the sake of brevity, the references to the chapters used will be omitted.

6.3.2.1 The Decision Node

Step 1
Below are the attributes used to create the Decision Node for the second case study. The creation of the Decision Node is the first step of DBV.

The Decision

Whether to implement a smart or sequential asset tagging system.

Information Required

The following information is required by the decision:

1. The cost of one system versus the other;
2. The requirement of the company;
3. The practicality of one system versus the other; and
4. The time impact of the systems.

The above information should be presented on a single histogram comparing the three parameters with the recommendation.

Value Range

The value range is calculated by determining the difference in total cost of either decision; provided by quotes from suppliers as well as internal costs. The following information was accounted for:

1. Cost of labels,
2. Cost of equipment,
3. Cost of logistics,
4. Cost of administration, and
5. Cost of training.

This resulted in the cost of the options being $\$2,003,808.04$ and $\$733,000$ for smart and sequential numbering respectively.

LifeCycle

This is a once-off decision however, the costs are spread over a four month period or 16 weeks.

Frequency

This is a once off decision, the deadline is 31 August 2015 with a tolerance of two to three weeks.

Accuracy

The accuracy required is at least 90% with an accuracy floor of 80% and a ceiling of 100 %. The value loss is -5% per percent loss in accuracy and +1% per percent gain in accuracy.

Maximum Cost

The maximum cost is set by the contract and excludes internal costs. The contract cost is $\$88,607.50$, excluding any additional costs.

Require ROI

Not chosen as maximum cost is used.

6.3.2.2 Cost

Step 2

The determination of the Decision Node's cost is the second step of Decision Based Valuation. The cost of the information was as follows:

Contractual Costs

The deliverable covering the asset tagging choice incorporated other requirements as well. The section for the asset tagging choice only accounted for one quarter of the total cost of the deliverable.

$\$354,430 \times 0.25 = \$88,607.50$

Labour:

Of the total time spent on the various deliverables by the company's project manage for the contract, a total of 48 hours can be attributed to this section.

Project Manager: $213/hour $\times$ 48 = $10,224

Utilities:

However small, the organization had to provide electricity and office space for the consultants. Budgeting in the hours spent for the section of deliverable, the total cost for utilities is as follows. Noting that the laptop used electricity at a 75 Wh consumption and was used for 8 hour days.

$\$0.12/\text{kWh} \times 0.075 \times 8 \times 5 \times 16 = \$5.76$

As can be seen, the utilities costs are negligible in comparison to the other costs.

Total

Thus, the total cost for the information is: $98,837.26

6.3.2.3 Calculations

The following calculations go through the last four steps of Decision Based Valuation

.

Amortization
Step 3

The depreciable cost is calculated as shown in Equation (6.3.12), noting that in this case, there are no asset costs that are already being depreciated.

$$\begin{aligned} C_D &= C_T - C_{ND} \\ C_D &= \$98,837.26 - \$0 \\ C_D &= \$98,837.26 \end{aligned} \tag{6.3.12}$$

Subsequently, the amortization schedule is calculated as shown in Equation (6.3.13).

$$
\begin{aligned}
A_m &= \frac{C_D}{N_L} \\
A_m &= \frac{\$98,837.26}{4} \\
A_m &= \$24,709.32
\end{aligned} \tag{6.3.13}
$$

Valuing Step 4

First the average value range was calculated.

$$
\begin{aligned}
V_\delta &= \frac{V_{max} - V_{min}}{2} \\
V_\delta &= \frac{\$2,003,808.04 - \$733,000}{2} \\
V_\delta &= \$635,404.02
\end{aligned} \tag{6.3.14}
$$

Then, the frequency modifier is calculated (in weeks).

$$
\begin{aligned}
I_F &= \frac{F_T - \sqrt{(F_N - F_I)^2}}{F_T} \\
I_F &= \frac{3 - \sqrt{(16 - 18)^2}}{3} \\
I_F &= 0.33 \geq 0 \\
I_F &= 1
\end{aligned} \tag{6.3.15}
$$

Next, the accuracy modifier is calculated with Equation (6.3.16), where $A_{RO} = 100\% - (100\% - 90\%)(1) = 90\%$.

$$
\begin{aligned}
I_A &= A_{RO} + (A_I - A_{RO}) \times A_{CO} \\
I_A &= 90\% + (95\% - 90\%) \times 1\% \\
I_A &= 95\% \text{ or } 0.95
\end{aligned} \tag{6.3.16}
$$

Using the above modifiers, the quality factor is calculated as shown below.

$$
\begin{aligned}
Q_f &= I_A + I_F \\
Q_f &= 0.95 + 1 \\
Q_f &= 1.95
\end{aligned} \tag{6.3.17}
$$

Finally, the value that the Decision Node should obtain is calculated in Equation (6.3.18).

$$\begin{aligned} V_N &= V_\delta \times Q_f \\ V_N &= \$635,404.20 \times 1.95 \\ V_N &= \$1,239,037.84 \end{aligned} \tag{6.3.18}$$

Performance
Step 5

The performance of the decision node is then easily calculated using the equation below.

$$\begin{aligned} \text{ROI} &= \frac{V_N - C_T}{C_T} \\ \text{ROI} &= \frac{\$1,239,037.84 - \$98,837.26}{\$98,837.26} \\ \text{ROI} &= 1,154\% \end{aligned} \tag{6.3.19}$$

The actual (realized) value wasn't available for the value performance calculation. This is due to the Decision Node still being operational and the information inputs are still being provided. It should be noted that the financial reward for this section of the deliverable is greater than the others thus, the value gained from this section will offset possible losses in the others.

Distribution of Value
Step 6

The data is classified as Type L data as per the classifications in section 4.1, therefore the equations for Type L data were used. The time to acquire the data made up approximately 30% of the total time. Thus the value of the combined data sources is calculated as follows.

$$\begin{aligned} V_P &= 1 + \frac{t_P}{t_A} \\ V_P &= 1 + \frac{70}{30} \\ V_P &= 3.33 \end{aligned} \tag{6.3.20}$$

$$\begin{aligned} V_D &= \frac{V_N(V_R)}{V_P} \\ V_D &= \frac{1,239,037.84(1)}{3.33} \\ V_D &= \$371,711.35 \end{aligned} \tag{6.3.21}$$

6.4 Case study results

This section covers the challenges identified in the case study, the performance of the method in valuing information, and a summary of its application.

6.4.1 Challenges

There are two distinct challenges with applying the DBV method in practice: (1) the calculation of the decision's value range, and (2) the calculation of the information input's accuracy. These two challenges were overcome, but with a certain amount of subjectivity.

An example of challenge one is seen with case study one and two, if the information provided in case study one required the manager to select a combination of one or two employees at different time intervals, the provided value range would be inaccurate and would have to be altered. This altering of the value range can only be done once the information has been provided. This problem is even more apparent in case study two; where the value range could only be calculated after initial information was provided. This highlights one of the major challenges; the value range of the decision cannot be reliably determined before the Decision Node has been fulfilled. This doesn't affect the functioning of the Decision Node and has a lesser impact on repeating decisions compared to once of decisions. This issue also indicates that the value range should be a dynamic metric that can change with the decision.

The next challenge is that of accuracy, as was seen in the case studies, none of the information came from machines where accuracy can be reliably measured. Most of the information came from either estimated prices – that relied on many inputs – or user completed documents. Both of these sources' accuracies are difficult to reliably measure or estimate. It was shown that the estimations for case one were accurate in the end however, the final details on case study two are not yet know and cannot be compared. Considering that the information accuracy is more difficult to determine in case two, it is unlikely that there will be aa good a value performance of the estimated and actual value. Standardising the approach in determining the accuracy of such data would be developed during the iteration of the method.

In summary, these two challenges provide obstacles to DBV and its implementation by organizations. Consequently, these challenges should be investigated in future research as detailed in section 7.2.

6.4.2 Performance

The DBV method performed remarkably well for case study one – even though there was a lot of vagueness in the information provided. Furthermore, it was able to provide a value for the information in case two, though the actual achieved value is not yet known; therefore its true performance is not known either. For both of the case studies, the Decision Node could be completed and the value calculations done (taking note of the aforementioned challenges). That is not to say that the method, as it stands, is perfect. DBV still requires maturing and iteration with a more standard approach to some of the key areas that were challenging to it. However, as a proof-of-concept it was able to achieve the aim of the study, namely: *To show that information and data can be valued through the development of a new valuation method.* The application of the method also provides valuable insights to the organization, and the various benefits that those insights bring.

One such benefit is showing the value of employee' time sheets, a main contributor to the Decision Node value in case study one. With this information, the company can take steps to improve the accuracy of these time sheets as well as improve the format for processing (to speed up the processing of the data on it). Another benefit for the organization is being able to substantiate and defend the cost of the project. For case study two, money saved from the consultant's reports and research outweighed their cost in the end. This benefit also applies to the consultant when tendering for future projects. Therefore, being able to understand both the cost and value of information provided valuable insights for both parties.

There is still improvement required for the accuracy of the method, although even with only being partially accurate, the insights it provided are still useful. Once the method has matured, there is definitely an argument to be made to establish Decision Nodes within organizations and to make use of DBV.

6.5 Chapter Summary

The validation chapter aimed to meet the objectives listed in section 1.5, namely: 3a; 3b; 3c; 3d; 4c. These objectives are for the aims; *To validate the need and success of this method through case studies and interviews* and *To show that there are grounds for information to be financially accountable as intangible assets.* The depth interviews provided validation for the need of an information valuation method. They show that organizations are dealing with increasingly more data and information while not having full control of this resource or how to extract value from it. An information valuation method such as DBV directly addresses this issue by providing organizations with a method to value their data and information. By knowing this value, managers are able to make more informed and strategic decisions on how to harness data's potential. The interviews also illustrated the importance of data and information and how these resources are in fact assets to the organization.

Following the interviews are the case studies that were conducted at a utility company in the Middle East. These case studies were conducted to test DBV as a proof-of-concept valuation method (objective 3d) and determine if it can successfully value real data and information. It was shown that DBV could in fact value information and its data and was able to achieve a high level of accuracy for the first case. However, through the implementation of DBV there were a few subjectivity issues. These issues centre around determining the accuracy of human generated data as well as determining the value range for the decision. These issues will be solved through further development and iteration of the method as stated in the limitations of the study in section 1.6.3.

The application of DBV is gradual and well laid out when using the top-down approach provided by the valuation framework (chapter 3). By using the valuation framework and the top-down approach, the cost and inputs for the data value chain were properly identified. This resulted in capturing the true cost and value of the information and, in part, resulted in an accurate valuation. The construction of the Decision Node is straight forward for both case studies, although potential inaccuracies were identified in the value range of the first case study were the decision to have changed. Furthermore, the decision's value range in case study two required initial information that was used to meet the Decision Node's criteria. This highlights the main issue with the value range; in most instances it is a dynamic attribute. Unfortunately the current version of DBV does not capture this dynamic nature.

Once the Decision Node was created and its cost were determined, the application of the value calculations were simple. Thus, the importance of getting the Decision Node and it cost right is crucial to the success of the method. Lastly, the identification of the data as per the data classifications in section 4.1 assisted in identifying how to distribute the Decision Node's value to its data. The data of the case studies were easily identified by the provided classifications.

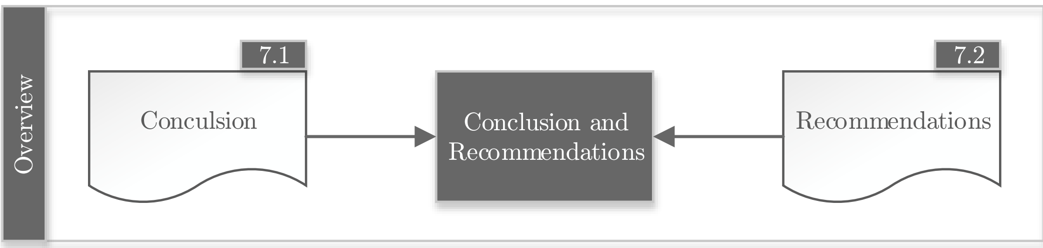
In conclusion the application of the method showed valuable insights about the value of the information being used and produced. More importantly, both parties were able to justify the expenses used to generate the information by seeing its resulting value, which is aligned with the objectives of this study. Furthermore, by being able to value data and information, the criteria for intangible assets can be met.

Chapter 7

Conclusion and Recommendations

Chapter 7 concludes the entire study and discusses the process of developing Decision Based Valuation, its aims, objectives, and which aspects of the method require further development. These aspects that require further development are described for future research.

Conclusion and Recommendations



7.1 Conclusion

This study set out to develop a proof-of-concept method that could be used to value data and information, while showing that it is possible to classify certain information as intangible assets. The aims and objectives of this study were met; consequently, the null hypothesis was rejected and hypotheses one and two were not rejected.

These hypotheses, as stated in section 1.5, were:

> H_0: *Information cannot be valued because it is not an intangible asset.*
>
> H_1: *Current methods have failed to determine the value of data and information because they were not specifically created to do so.*
>
> H_2: *Information that can be valued, can be regarded as an intangible asset.*

The first aim of the study was: *To identify established valuation methods for physical and intangible assets.* This aim and its objectives were fulfilled by Chapter 2, where a review of established valuation methods revealed that no single method was able to provide a complete process that could reliably value information. All of the reviewed methods required particular situations, such as active markets, for valuing intangible assets that do not exist for most information. Thus, the literature review provided evidence to not reject hypothesis one. The fourth aim of the study, *To show that there are grounds for information to be financially accountable as intangible assets*, was also partly fulfilled by Chapter 2. It detailed the classification of assets and the classification of intangible asset by International Financial Accounting Standards (IFRS), as well as the criteria for doing so. These criteria were then incorporated into Decision Based Valuation (DBV), developed in this study, in order for information to be recognized as a financially accountable intangible assets.

The second aim of the study was: *To show that information and data can be valued through the development of a new valuation method.* This aim was fulfilled by chapters 2, 3, and 4. The development of DBV was influenced by the Cost and Income approaches identified in Chapter 2, as well as through the incorporation of several other methods identified in the chapter. Chapter 3 presented the development of the valuation framework for DBV and the first part of the solution to the valuation problem. This framework addressed the fundamental principle of the value of information; it only has value if used. Furthermore, the framework sought to identify how the value of data and information, and their costs, occur throughout their value chains. Chapter 3 therefore enabled the effective identification and use of the inputs for DBV.

Chapter 4 presented the DBV method, starting with a description of classifications aimed at differentiating between data types. This differentiation was used to determine how certain data gain and lose value during the distribution of a Decision Node's value. These Decision Nodes are the core principle behind DBV and bring together the different valuation methods, as well as creating an analogy of a physical asset for information. Information is provided to these Decision Nodes, and if the criteria are met, that information is able to gain a certain value depending on its attributes. Decision Nodes also allow organizations to handle intangible assets in a more familiar way. After presenting Decision Nodes, Chapter 4 provides an in-depth description of how to determine the attributes and cost of data and information. In addition, calculations used by DBV were provided.

Following the presentation of DBV, tools and methods were provided for optimizing value and determining the performance of Decision Nodes, as seen in Chapter 5. This chapter directly fulfils aim five and its objectives, namely: *Improve the management and control of data and information's costs and value*. Chapter 5 initially presented a tool that organizations could use to adjust the value produced by DBV to account for errors. Following this tool, a set of methods were presented which could be used to extract value from data as well as improve the efficiency of data and information systems. This chapter then presented value optimization and performance assessment tools. These tools were created for DBV with the intention of providing organizations with better control of their data. Chapter 5 concludes the valuation solution presented in the study.

Chapter 6 provides the validation of both the need and success of DBV and directly fulfilled aim three of this study: *To validate the need and success of this method through case studies and interviews.* Several industry interviews conducted during the course of this study provided validation of the need for DBV. The interviews identified key needs within industry with regards to data and information. For example, although organizations understand the value of data and information, they currently unable to value these resources effectively. Furthermore, these organizations had limited control over their data and information system costs, a deficit that DBV would be able to address. Two case studies were conducted to test DBV on real data in a practical setting. These case studies successfully showed that information can be valued. However, it should be noted that issues of subjectivity emerged when applying the method in two primary areas, namely; the determination of the decision's value range and calculating the information's accuracy. DBV is currently unable to standardize the approach and methodology for these aforementioned areas – to do so would require further iteration of DBV. This iteration of DBV would involve improving the standardisation of the method by: (1) developing an improved approach of determining information accuracy, (2) improving the approach to determining the decision value range, and (3) refining the calculations through quantitative testing with various case studies. Noting that the issue with accuracy resides with the data and information and not DBV.

Lastly, the method set out to value both Big Data and information although, during its implementation is was shown that Big Data does not affect the valuation method. Valuing information derived from Big Data versus normal data follows the same process - distributing that value to the data is also similar. Therefore, it can be concluded that this method is suitable for Big Data and presents no issues with being implemented. In summary, this study was successful in showing that data and information can be valued and that these resources can be financially accountable and handled as intangible assets. However, further iteration of DBV and its components is still required before it can be reliably implemented by organizations. The areas of DBV that require further development will be discussed in the section below.

7.2 Recommendations

Recommendations for the future research and development of DBV centre on improving the standardization and consistency of the method. These aspects are difficult to capture during the first iteration of a method and would require additional development through quantitative analysis and further field testing. The most significant areas for future research are discussed below.

1. Data classifications as proposed in section 4.1 require refinement. The classifications discussed in this study described different data types and how they gain and lose value throughout their data value chains. The standardization and further development of these classifications will help improve the consistency of future iterations of the DBV method. The creation and testing of formulae specific to these classifications will also assist with this endeavour.

2. Accuracy is perhaps the most important aspect of information, which relates to how much of it potential value it are able to obtain. In saying that, the accuracy calculations developed and implemented in this study still require further research and field testing. Currently, these formulae require a certain amount of subjectivity from the user; if DBV is to be consistent and reliable, this subjectivity of accuracy estimation needs to be removed.

3. Another area for future development is the implementation of Decision Nodes within an organization. This includes the handling, control, and

the strategic decisions based off these Decision Nodes. This study briefly discussed two handling approaches – Intangible Asset Registers (IAR) and cost centres – however, there are many more interactions within an organization that need to be covered. By detailing these other interactions and management techniques, organizations will have greater control over their intangible assets and information.

4. One important attribute required for the valuation of data and information is the value range of the decision it is based off of. If this value range is not accurately captured, then the results of DBV could be significantly different to what the actual value of the data and information is. There is currently still a reasonable amount of subjectivity involved in determining this value range, as was evidenced in the case studies employed by this study. Greater standardization in the approach to determining a decision's value range would add even more consistency to the method, thus ensuring that the results are more reliable.

5. The final recommendation for further research is the iteration of DBV itself, namely; the calculations it uses, the Decision Node, and framework it is based on. As described and implemented in this study, it was shown that DBV can be used to value information. However, it is still the first iteration of the method. Further development would thus improve the accuracy of the method while increasing its functionality, all of which would contribute to information being financially accountable.

All of the recommendations above address the issue of the consistency and reliability of Decision Based Valuation. Further development of the aforementioned areas will establish DBV as a viable method for the valuation of information and indeed other intangible assets.

Bibliography

(2004, March). Ias 38 - intangible assets.

(2014). *Asset Management - Overview, principles and terminology.*

(2014, November). What is lean.

Aggarwal, C. C. (2007). *Data streams: models and algorithms*, Volume 31. Springer Science & Business Media.

Aggarwal, C. C., J. Han, J. Wang, and P. S. Yu (2003). A framework for clustering evolving data streams. In *Proceedings of the 29th international conference on Very large data bases-Volume 29*, pp. 81–92. VLDB Endowment.

Agrawal, D., S. Das, and A. El Abbadi (2011). Big data and cloud computing: current state and future opportunities. In *Proceedings of the 14th International Conference on Extending Database Technology*, pp. 530–533. ACM.

Albanesius, C. (2011). Every place samsung and apple are suing each other. *Message posted to http://www. pcmag. com*.

Alexander, E. R. and R. R. Hiner (2001). Accounting for asset retirement obligations. *Journal of Accountancy 192*(6), 49.

Allen, F. and D. Gale (2000). Financial contagion. *Journal of political economy 108*(1), 1–33.

Amadi-Echendu, J. E., R. Willett, K. Brown, T. Hope, J. Lee, J. Mathew, N. Vyas, and B.-S. Yang (2010). What is engineering asset management? In *Definitions, concepts and scope of engineering asset management*, pp. 3–16. Springer.

Andriessen, D. (2004). Ic valuation and measurement: classifying the state of the art. *Journal of intellectual capital 5*(2), 230–242.

Anson, W. (2007). The intangible assets handbook. *Maximizing Value from Intangible Assets*.

Antonopoulos, C. and P. Sakellaris (2011). Estimating computer depreciation using online auction data. *Economics of Innovation and New Technology 20*(2), 183–204.

Arrow, A. K. (2002). Intangible asset deployment in technology-rich companies: how does innovation affect return on assets? *International Journal of Technology Management 24*(4), 375–390.

Avison, D. and G. Fitzgerald (2003). *Information systems development: methodologies, techniques and tools.* McGraw Hill.

Baker, N. R. and W. H. Pound (1964). R-and-d project selection-where we stand. *IEEE Transactions on Engineering Management 11*(4), 124–134.

Baker, W. (1978). Life-cycle costing. *Plant Eng.(Barrington, Ill.);(United States) 32*(11).

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics 9*(1), 3–18.

Barney, J. A. and J. R. Barney (2003, April 29). Method and system for rating patents and other intangible assets. US Patent 6,556,992.

Beichelt, F. (2001). A replacement policy based on limiting the cumulative maintenance cost. *International Journal of Quality & Reliability Management 18*(1), 76–83.

Bekenstein, J. D. (1981). Energy cost of information transfer. *Physical Review Letters 46*(10), 623.

Berg, M., A. De Waegenaere, and J. L. Wielhouwer (2001). Optimal tax depreciation with uncertain future cash-flows. *European Journal of Operational Research 132*(1), 197–209.

Bleazard, D. G. and P. T. Khu (2001). Physical asset management. *Engineering and Mining Journal 202*(3), WW36.

Bloom, N. and J. Van Reenen (2002). Patents, real options and firm performance. *The Economic Journal 112*(478), C97–C116.

Board, I. A. S. (2015). *International financial reporting standards (IFRS's): including international accounting standards (IAS's) and interpretations as at.* London, International Accounting Standards Board.

Bollier, D. and C. M. Firestone (2010). *The promise and peril of big data.* Aspen Institute, Communications and Society Program Washington, DC, USA.

Borio, C. E. and P. W. Lowe (2002). Asset prices, financial and monetary stability: exploring the nexus.

Borkar, V., M. J. Carey, and C. Li (2012). Inside big data management: ogres, onions, or parfaits? In *Proceedings of the 15th International Conference on Extending Database Technology*, pp. 3–14. ACM.

Borthakur, D. (2007). The hadoop distributed file system: Architecture and design. *Hadoop Project Website 11*(2007), 21.

Bosworth, D. and M. Rogers (2001). Market value, r&d and intellectual property: an empirical analysis of large australian firms. *Economic Record 77*(239), 323–337.

Bott, M. (2000). Software as a corporate asset. *IEE Proceedings-Software 147*(2), 31–36.

Briggs, A. and A. Gray (2000). Economics notes: using cost effectiveness information. *BMJ: British Medical Journal 320*(7229), 246.

Brown, B., M. Chui, and J. Manyika (2011). Are you ready for the era of âĂŸbig dataâĂŹ. *McKinsey Quarterly 4*, 24–35.

Bughin, J., M. Chui, and J. Manyika (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly 56*(1), 75–86.

Burkhauser, R. V. and J. Cawley (2008). Beyond bmi: the value of more accurate measures of fatness and obesity in social science research. *Journal of health economics 27*(2), 519–529.

Cabena, P., P. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi (1998). *Discovering data mining: from concept to implementation*. Prentice-Hall, Inc.

Canadian Institute of Charted Accountants (2015). Illustrative example of intangible asset valuation.

Carcello, J. V. and A. L. Nagy (2004). Audit firm tenure and fraudulent financial reporting. *Auditing: A Journal of Practice & Theory 23*(2), 55–69.

Chareonsuk, C. and C. Chansa-ngavej (2008). Intangible asset management framework for long-term financial performance. *Industrial Management & Data Systems 108*(6), 812–828.

Chareonsuk, C. and C. Chansa-ngavej (2010). Intangible asset management framework: an empirical evidence. *Industrial Management & Data Systems 110*(7), 1094–1112.

Cheevaprawatdomrong, T. and R. L. Smith (2003). A paradox in equipment replacement under technological improvement. *Operations Research Letters 31*(1), 77–82.

Chen, H., R. H. Chiang, and V. C. Storey (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly 36*(4), 1165–1188.

Choi, W. W., S. S. Kwon, and G. J. Lobo (2000). Market valuation of intangible assets. *Journal of Business Research 49*(1), 35–45.

Clayton, J., D. Geltner, and S. W. Hamilton (2001). Smoothing in commercial property valuations: Evidence from individual appraisals. *Real Estate Economics 29*(3), 337–360.

Coffman, K. G. and A. M. Odlyzko (2002). Internet growth: Is there a âĂIJmooreâĂŹs lawâĂİ for data traffic? In *Handbook of massive data sets*, pp. 47–93. Springer.

Cole, R. J. and E. Sterner (2000). Reconciling theory and practice of life-cycle costing. *Building Research & Information 28*(5-6), 368–375.

Cukier, K. and V. Mayer-Schoenberger (2013). Rise of big data: How it's changing the way we think about the world, the. *Foreign Aff. 92*, 28.

Cummins, J. and D. Bawden (2010). Accounting for information: Information and knowledge in the annual reports of ftse 100 companies. *Journal of Information Science*.

Cusumano, M. A. (2013). The apple-samsung lawsuits. *Communications of the ACM 56*(1), 28–31.

Cuzzocrea, A., I.-Y. Song, and K. C. Davis (2011). Analytics over large-scale multidimensional data: the big data revolution! In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP*, pp. 101–104. ACM.

Dando, N. and T. Swift (2003). Transparency and assurance minding the credibility gap. *Journal of Business Ethics 44*(2-3), 195–200.

Davenport, T. H. and J. E. Short (2003). Information technology and business process redesign. *Operations management: critical perspectives on business and management 1*, 97.

Deelman, E., G. Singh, M. Livny, B. Berriman, and J. Good (2008). The cost of doing science on the cloud: the montage example. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, pp. 50. IEEE Press.

Demchenko, Y., P. Grosso, C. De Laat, and P. Membrey (2013). Addressing big data issues in scientific data infrastructure. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, pp. 48–55. IEEE.

Dittenhofer, M. (2001). Performance auditing in governments. *Managerial Auditing Journal 16*(8), 438–442.

Dittrich, J. and J.-A. Quiané-Ruiz (2012). Efficient big data processing in hadoop mapreduce. *Proceedings of the VLDB Endowment 5*(12), 2014–2015.

Dretske, F. (1981). Knowledge and the flow of information.

Eaton, J. J. and D. Bawden (1991). What kind of resource is information? *International journal of information management 11*(2), 156–165.

El-Tawy, N. and M. Abdel-Kader (2013). Accounting recognition of information as an asset. *Journal of Information Science 39*(3), 333–345.

English, L. P. (1999). *Improving data warehouse and business information quality: methods for reducing costs and increasing profits*. John Wiley & Sons, Inc.

Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (1996). Advances in knowledge discovery and data mining.

Fiala, P. (2005). Information sharing in supply chains. *Omega 33*(5), 419–423.

Fisher, D., R. DeLine, M. Czerwinski, and S. Drucker (2012). Interactions with big data analytics. *interactions 19*(3), 50–59.

Foster, B. P., R. Fletcher, and W. D. Stout (2003). Valuing intangible assets. *The CPA journal 73*(10), 50.

Franks, B. (2012). *Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics*, Volume 49. John Wiley & Sons.

Frentz, S. J. (2011). Customize appraisals for built-in gain tax purposes. *J. Passthrough Entities 14*, 39.

Gaber, M. M., A. Zaslavsky, and S. Krishnaswamy (2005). Mining data streams: a review. *ACM Sigmod Record 34*(2), 18–26.

Galai, D., A. Raviv, and Z. Wiener (2007). Liquidation triggers and the valuation of equity and debt. *Journal of Banking & Finance 31*(12), 3604–3620.

Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting 4*(1), 1–28.

Gereffi, G., J. Humphrey, R. Kaplinsky, et al. (2001). Introduction: Globalisation, value chains and development. *IDS bulletin 32*(3), 1–8.

Ghazal, A., T. Rabl, M. Hu, F. Raab, M. Poess, A. Crolotte, and H.-A. Jacobsen (2013). Bigbench: towards an industry standard benchmark for big data analytics. In *Proceedings of the 2013 ACM SIGMOD international conference on Management of data*, pp. 1197–1208. ACM.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 3–8.

Gluch, P. and H. Baumann (2004). The life cycle costing (lcc) approach: a conceptual discussion of its usefulness for environmental decision-making. *Building and environment 39*(5), 571–580.

Gordon, L. A., M. P. Loeb, and W. Lucyshyn (2003). Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy 22*(6), 461–485.

Gorry, G. A. and M. S. S. Morton (1971). *A framework for management information systems*, Volume 13. Massachusetts Institute of Technology.

Gould, J. P. (1974). Risk, stochastic preference, and the value of information. *Journal of Economic Theory 8*(1), 64–84.

Greenberg, A., J. Hamilton, D. A. Maltz, and P. Patel (2008). The cost of a cloud: research problems in data center networks. *ACM SIGCOMM computer communication review 39*(1), 68–73.

Gröjer, J.-E. (2001). Intangibles and accounting classifications: in search of a classification strategy. *Accounting, Organizations and Society 26*(7), 695–713.

Groves, P., B. Kayyali, D. Knott, and S. Van Kuiken (2013). The âĂŸbig dataâĂŹrevolution in healthcare. *McKinsey Quarterly*.

Gu, F. and B. Lev (2001). Intangible assets: measurement, drivers, usefulness. *New York*.

Guck, R. L. (1999, January 26). Method for storing/retrieving files of various formats in an object database using a virtual multimedia file system. US Patent 5,864,870.

Guthrie, J. (2001). The management, measurement and the reporting of intellectual capital. *Journal of Intellectual capital 2*(1), 27–41.

Hall, B. H., G. Thoma, and S. Torrisi (2007). The market value of patents and r&d: Evidence from european firms. In *Academy of Management Proceedings*, Volume 2007, pp. 1–6. Academy of Management.

Hartman, J. C. and I. C. Schafrick (2004). The relevant internal rate of return. *The Engineering Economist 49*(2), 139–158.

Hastings, N. A. J. (2010). *Physical asset management*, Volume 2. Springer.

Hastings, N. A. J. (2015). Introduction to asset management. In *Physical Asset Management*, pp. 3–23. Springer.

Higson, C. and D. Waltho (2009). Valuing information as an asset.

Høegh-Krohn, N. E. J. and K. H. Knivsflå (2000). Accounting for intangible assets in scandinavia, the uk, the us, and by the iasc: Challenges and a solution. *The International Journal of Accounting 35*(2), 243–265.

Holloway, B. P. and R. F. Reilly (2012). Intangible asset valuation approaches and methods. *Insights Journal 12*, 13–33.

Israel, D. J. and J. Perry (1991). *What is information?* CSLI.

Jagadish, H., J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, and C. Shahabi (2014). Big data and its technical challenges. *Communications of the ACM 57*(7), 86–94.

Juran, J. M. (1995). *Managerial breakthrough: The classic book on improving management performance.* McGraw-Hill New York.

Jydstrup, R. A. and M. J. Gross (1966). Cost of information handling in hospitals. *Health services research 1*(3), 235.

Kaisler, S., F. Armour, J. A. Espinosa, and W. Money (2013). Big data: Issues and challenges moving forward. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pp. 995–1004. IEEE.

Kanodia, C., H. Sapra, and R. Venugopalan (2004). Should intangibles be measured: What are the economic trade-offs? *Journal of Accounting Research 42*(1), 89–120.

Kaplinsky, R. and M. Morris (2001). *A handbook for value chain research*, Volume 113. IDRC Ottawa.

Katal, A., M. Wazid, and R. Goudar (2013). Big data: Issues, challenges, tools and good practices. In *Contemporary Computing (IC3), 2013 Sixth International Conference on*, pp. 404–409. IEEE.

King, K. (2002). The value of intellectual property, intangible assets and goodwill. *Journal of Intellectual Property Rights 7*(3), 245–248.

Klein, P. G. (2008). The make-or-buy decisions: Lessons from empirical studies. In *Handbook of new institutional economics*, pp. 435–464. Springer.

Krajewski, L. J., L. P. Ritzman, and M. K. Malhotra (2007). Operations management: processes and value chains.

Larkin, J. H. and H. A. Simon (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science 11*(1), 65–100.

LaValle, S., E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz (2013). Big data, analytics and the path from insights to value. *MIT sloan management review 21*.

Lee, M. and L. F. Cunningham (2001). A cost/benefit approach to understanding service loyalty. *Journal of services Marketing 15*(2), 113–130.

Lefebvre, L., E. Lefebvre, J. Harvey, et al. (1996). Intangible assets as determinants of advanced manufacturing technology adoption in sme's: toward an evolutionary model. *Engineering Management, IEEE Transactions on 43*(3), 307–322.

Leiblein, M. J., J. J. Reuer, and F. Dalsace (2002). Do make or buy decisions matter? the influence of organizational governance on technological performance. *Strategic management journal 23*(9), 817–833.

Leitner, K.-H. (2005). Managing and reporting intangible assets in research technology organisations. *R&D Management 35*(2), 125–136.

Lev, B. and J. H. Daum (2004). The dominance of intangible assets: consequences for enterprise management and corporate reporting. *Measuring business excellence 8*(1), 6–17.

Lohr, S. (2012). The age of big data. *New York Times 11*.

Luehrman, T. A. (1998). Investment opportunities as real options: getting started on the numbers. *Harvard business review 76*, 51–66.

MacMillan, I. C., A. B. Van Putten, R. G. McGrath, and J. D. Thompson (2006). Using real options discipline for highly uncertain technology investments. *Research-Technology Management 49*(1), 29–37.

Mahrt, M. and M. Scharkow (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media 57*(1), 20–33.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers (2011). Big data: The next frontier for innovation, competition, and productivity.

Marr, B., G. Schiuma, and A. Neely (2004). Intellectual capital-defining key performance indicators for organizational knowledge assets. *Business Process Management Journal 10*(5), 551–569.

Maxwell, S. L., J. R. Rhodes, M. C. Runge, H. P. Possingham, C. F. Ng, and E. McDonald-Madden (2015). How much is new information worth? evaluating the financial benefit of resolving management uncertainty. *Journal of Applied Ecology 52*(1), 12–20.

Mayer, R. E. and J. K. Gallini (1990). When is an illustration worth ten thousand words? *Journal of educational psychology 82*(4), 715.

Mayer-Schönberger, V. and K. Cukier (2013). *Big data: A revolution that will transform how we live, work, and think.* Houghton Mifflin Harcourt.

McAfee, A., E. Brynjolfsson, et al. (2012). Big data: the management revolution. *Harvard business review 90*(10), 60–68.

Michael, K. and K. W. Miller (2013). Big data: New opportunities and new challenges [guest editors' introduction]. *Computer 46*(6), 22–24.

Mitchell, J. S., J. E. Hickman, and J. E. Amadi-Echendu (2007). *Physical asset management handbook.* Clarion technical publishers.

Moody, D. L. and P. Walsh (1999). Measuring the value of information-an asset valuation approach. In *ECIS*, pp. 496–512.

Moore, K., K. Eyestone, and D. Coddington (2013). The big deal about big data. *Healthcare financial management: journal of the Healthcare Financial Management Association 67*(8), 60–6.

Moraux, F. (2002). Valuing corporate liabilities when the default threshold is not an absorbing barrier.

Myers, S. C. (1984). Finance theory and financial strategy. *Interfaces 14*(1), 126–137.

Neely, A. and Y. Jarrar (2004). Extracting value from data-the performance planning value chain. *Business Process Management Journal 10*(5), 506–509.

Norman, G. (1990). Life cycle costing. *Property Management 8*(4), 344–356.

Olsen, M. G., M. Halliwell, and R. P. Gray (2007). Intangible value: delineating between shades of gray. *Journal of Accountancy 203*(5), 66.

PAGAON, M. and T. Jappelli (1993). Information sharing in credit markets. *The Journal of Finance 48*(5), 1693–1718.

Pappas, S. (2014). Google flu tracker overestimated cases, study argues, pointing to flaws in âĂŸbig dataâĂŹ.

Penman, S. H. (2009). Accounting for intangible assets: There is also an income statement. *Abacus 45*(3), 358–371.

Pindyck, R. S. (1991). Irreversibility, uncertainty, and investment. Technical report, National Bureau of Economic Research.

Poikolainen, K. and P. Kärkkäinen (1983). Diary gives more accurate information about alcohol consumption than questionnaire. *Drug and alcohol dependence 11*(2), 209–216.

Poon, T. S.-C. (2004). Beyond the global production networks: a case of further upgrading of taiwan's information technology industry. *International journal of technology and globalisation 1*(1), 130–144.

Porter, M. E. and V. E. Millar (1985). How information gives you competitive advantage.

Power, M. K. (2003). Auditing and the production of legitimacy. *Accounting, organizations and society 28*(4), 379–394.

Qi, G. et al. (2006). Corporate value of separating cash-flow rights from control: Empirical study on family owned public firms in china [j]. *Accounting Research 4*, 004.

Reilly, R. F. (2008). The relief from royalty method of intellectual property valuation. *Insights. Autumn*, 20–43.

Reilly, R. F. and R. P. Schweihs (1998). *Valuing intangible assets*. McGraw Hill Professional.

Reily, R. F. and R. P. Schweihs (1999). *Valuing intangible assets*. McGraw-Hill.

Rodov, I. and P. Leliaert (2002). Fimiam: financial method of intangible assets measurement. *Journal of intellectual capital 3*(3), 323–336.

Ross, S. A. (1995). Uses, abuses, and alternatives to the net-present-value rule. *Financial management*, 96–102.

Russom, P. et al. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*.

Sagiroglu, S. and D. Sinanc (2013). Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, pp. 42–47. IEEE.

Sakalaki, M. and S. Kazi (2007). How much is information worth? willingness to pay for expert and non-expert informational goods compared to material goods in lay economic thinking. *Journal of Information Science 33*(3), 315–325.

Sakalaki, M. and S. Kazi (2008). Valuing and representing information:: the paradox of undervaluing information and overvaluing information producers. *Journal of Information Science*.

Sale, K., M. Marien, S. Brand, H. Cleveland, et al. (1997). Information technology revolution: Boon or bane? *The Futurist 31*(1), 10–15.

Schenk, A. S. (1966). Depreciation of intangible assets: The uncertainty of death and taxes. *Wayne L. Rev. 13*, 501.

Schiozer, R. F. and S. B. Suslick (2003). An evaluation of a royalty relief model for mature fields. In *AAPG International Conference, Barcelona, Available online at: http://www. searchanddiscovery. com/abstracts/pdf/2003/intl/extend/ndx_ 81913. pdf*.

Shah, A. (2013). New supercomputer uses ssds as alternative to dram, hard drives.

Shannon, C. E. and W. Weaver (1949). The mathematical theory of communication (urbana, il.

Sklavos, N. (2010). On the hardware implementation cost of crypto-processors architectures. *Information Security Journal: A Global Perspective 19*(2), 53–60.

Skyrme, D. J. (1998). Valuing knowledge: is it worth it? *Managing information 5*(2), 24–26.

Srivastava, R. K., T. A. Shervani, and L. Fahey (1998). Market-based assets and shareholder value: a framework for analysis. *The Journal of Marketing*, 2–18.

Steenkamp, N. and V. Kashyap (2010). Importance and contribution of intangible assets: Sme managers' perceptions. *Journal of Intellectual Capital 11*(3), 368–390.

Stephens, D. (1989). Variance and the value of information. *American Naturalist*, 128–140.

Sudarsanam, S., G. Sorwar, and B. Marr (2006). Real options and the impact of intellectual capital on corporate value. *Journal of intellectual capital 7*(3), 291–308.

Sullivan, W. G., T. N. McDonald, and E. M. Van Aken (2002). Equipment replacement decisions and lean manufacturing. *Robotics and Computer-Integrated Manufacturing 18*(3), 255–265.

Sullivan Jr, P. H. and P. H. Sullivan Sr (2000). Valuing intangibles companies-an intellectual capital approach. *Journal of Intellectual capital 1*(4), 328–340.

Tam, K. Y. (1998). Analysis of firm-level computer investments: A comparative study of three pacific-rim economies. *Engineering Management, IEEE Transactions on 45*(3), 276–286.

Taylor, W. B. (1981). The use of life cycle costing in acquiring physical assets. *Long Range Planning 14*(6), 32–43.

Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering 22*(2), 127–151.

Trigeorgis, L. (1996). *Real options: Managerial flexibility and strategy in resource allocation*. MIT press.

Villars, R. L., C. W. Olofson, and M. Eastwood (2011). Big data: What it is and why you should care. *White Paper, IDC*.

Wakeman, L. M. (1980). Optimal tax depreciation. *Journal of Accounting and Economics 2*(3), 213–237.

Walker, G. and D. Weber (1984). A transaction cost approach to make-or-buy decisions. *Administrative science quarterly*, 373–391.

Wang, A. (2010). Comparison of real asset valuation models: A literature review. *International Journal of Business and Management 5*(5), p14.

Westphal, C. and T. Blaxton (1998). Data mining solutions: methods and tools for solving real-world problems.

Whittington, G. (2008). Fair value and the iasb/fasb conceptual framework project: an alternative view. *Abacus 44*(2), 139–168.

Whitworth, B. L. (2003, September 16). Method of creating an index of residual values for leased assets, transferring residual value risk, and creating lease securitizations. US Patent 6,622,129.

Williamson, O. E. (1981). The economics of organization: The transaction cost approach. *American journal of sociology*, 548–577.

Woodhouse, J. (1997). What is asset management? *Maintenance and Asset Management 12*, 26–28.

Xiang-bin, L. (2007). A research on asset appraisal method of forestland resource [j]. *Journal of Harbin University of Commerce (Social Science Edition) 2*, 012.

Yu, Z., H. Yan, and T. Edwin Cheng (2001). Benefits of information sharing with supply chain partnerships. *Industrial management & Data systems 101*(3), 114–121.

Zandbergen, P. (2015). Data types in programming: Numbers, strings and others. Accessed: 2015-01-26.

Zaslavsky, A., C. Perera, and D. Georgakopoulos (2013). Sensing as a service and big data. *arXiv preprint arXiv:1301.0159*.

Zhu, K. (2004). The complementarity of information technology infrastructure and e-commerce capability: A resource-based assessment of their business value. *Journal of Management Information Systems 21*(1), 167–202.

Zikopoulos, P., C. Eaton, et al. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data.* McGraw-Hill Osborne Media.

Appendices

Appendix A

Decision Based Valuation Process

Figure A.1: Decision Based Valuation Part 1

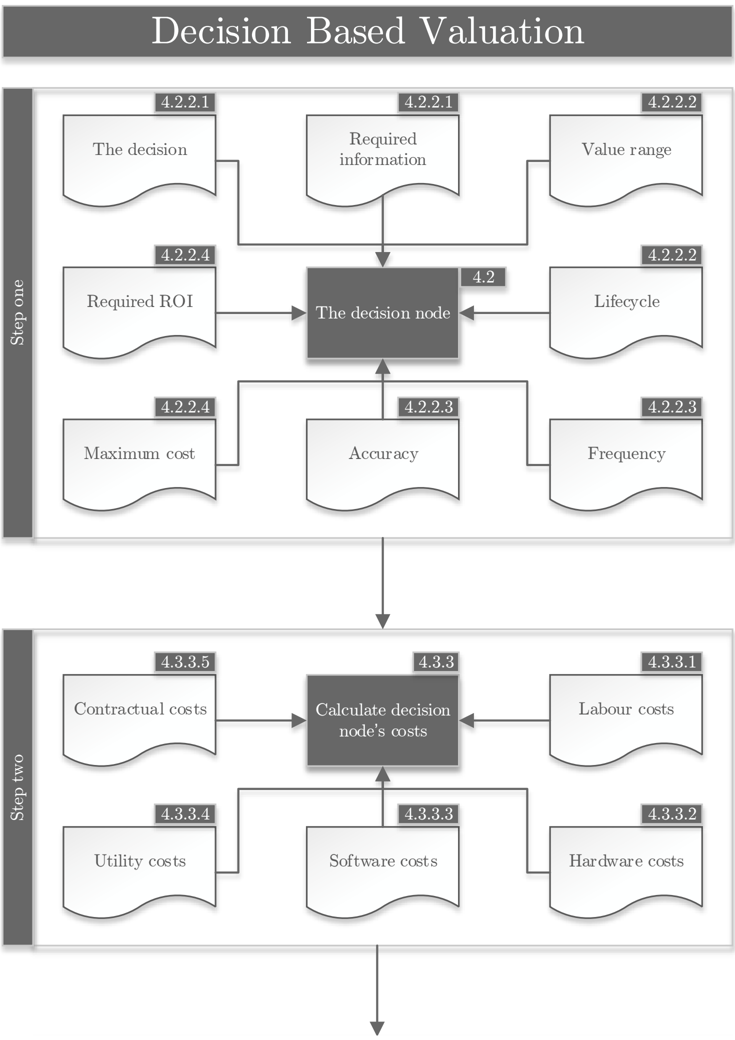


Figure A.2: Decision Based Valuation Part 2

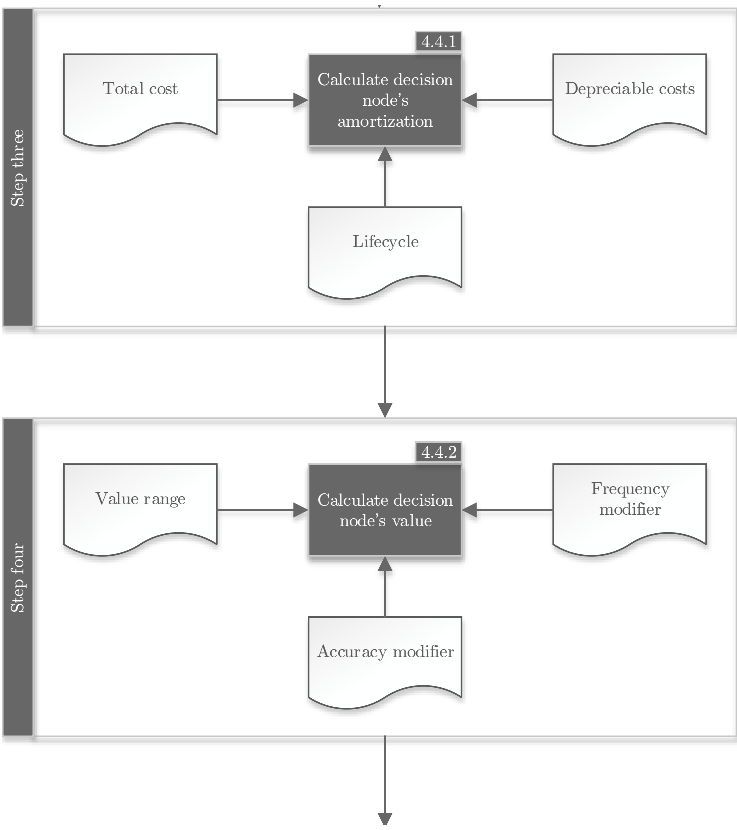


Figure A.3: Decision Based Valuation Part 3