

**A DATA PROTECTION METHODOLOGY TO PRESERVE
CRITICAL INFORMATION FROM THE POSSIBLE THREAT OF
INFORMATION LOSS**

BY

TARYN SCHWARTZEL

**Submitted in fulfilment of part of the requirements for the
degree of Master of Technology in Information Technology**



In the Faculty of Management

University of Johannesburg

SUPERVISOR: Dr. E. Mnkandla

ABSTRACT

Information is a company's greatest asset that is continually under threat from human error, technological failure, natural disasters and other external factors. These threats need to be identified and quantified and their relevant protection techniques need to be deployed. This research will allow businesses to ascertain which of these data protection strategies to embrace and deploy, thereby highlighting the balance between cost and value for their business needs.

Every commercial enterprise should understand the business value of their data and realise that protecting this data is of utmost importance. However, company data often resides on different mediums, in different locations and implementing a data protection strategy is not always cost effective in terms of the cost of storage mediums and protection methods. The challenges that businesses face is trying to distinguish between mission-critical data from other business data, excluding any non-business or invaluable data that resides on their systems. Thus a cost-effective data protection strategy can be implemented according to the different values of business data.

This research provides a model to enable an organisation to:

- Utilise the model as a framework or guideline in determining a strategy for protection, storage, retrieval and preservation of business critical data.
- Define the data protection strategy to meet the organisation's business requirements.
- Define a cost effective data protection solution that encompasses protection, storage, retrieval and preservation of business critical data.
- Make strategic decisions based on an array of best practices to ensure mission-critical data is protected accordingly.

- Draw a conclusion between the costs of implementing these solutions against the real business value of the data that it protects.



DECLARATION

I, Taryn Schwartzel, declare that this research report is my own work. It is submitted in fulfilment of the requirements for the degree of Master of Technology in Information Technology at the University of Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Name: Taryn Schwartzel

Date:



DEDICATION

I dedicate this research to Thuraisha Moodley who has motivated me through the years to achieve and complete my Master's studies during challenging times of trauma and ill health. She has taught me to effectively live my life and contribute to my passions, in this context, for the field of Technology.



ACKNOWLEDGEMENTS

I would like to acknowledge and thank my first supervisor Dr. Dave Augustyn for supervising me in my initial research.

Secondly, I would like to acknowledgement and thank Dr. Ernest Mnkandla for taking over as my new supervisor and directing me very effectively in my studies. I would like to thank and make an acknowledgement to Jonathan de Magalhaus for spending significant time and effort on in-depth interviews with me in order to gain knowledge on the topics covered.

In addition my thanks extend to those consultants who were willing to be interviewed and share their experiences in this field of study.

I would like to thank the language editor, Dr. Andrew Graham, for editing my dissertation.

I further acknowledge Thuraisha Moodley for her dedicated support to me, my life and in this context my studies.

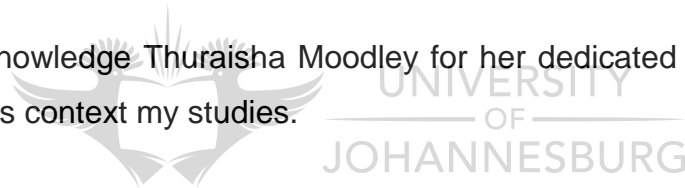


TABLE OF CONTENTS

ABSTRACT	II
DECLARATION.....	IV
DEDICATION	V
ACKNOWLEDGEMENTS.....	VI
LIST OF TABLES.....	X
LIST OF FIGURES	X
CHAPTER 1: INTRODUCTION	1
7.1 PURPOSE OF STUDY	1
7.2 PROBLEM STATEMENT	1
7.3 SIGNIFICANCE OF STUDY.....	2
7.4 BACKGROUND DISCUSSION	2
7.5 CONCEPT CLARIFICATION.....	4
7.6 ASSUMPTIONS	5
CHAPTER 2: LITERATURE REVIEW	6
2.1. INTRODUCTION	6
2.1.1. SCOPE OF LITERATURE	6
2.1.2. RELATED RESEARCH	8
2.2. THE BUSINESS VALUE OF DATA	10
2.2.1. DATA AS A CRITICAL RESOURCE.....	11
2.2.2. DATA VALUE CHANGES CONSTANTLY.....	12
2.2.3. THE CORRELATION BETWEEN DATA VALUE AND DATA STORAGE	13
2.2.4. BUSINESS VALUE OF IMPLEMENTING DATA PROTECTION TECHNIQUES	13
2.3. THREATS TO DIGITAL DATA AND THE IMPACT OF DATA UNAVAILABILITY OR DATA LOSS ON BUSINESS OPERATIONS	15
2.3.1. DATA DISCOVERY PLANNING.....	15
2.3.2. TYPES OF THREATS.....	16
2.3.3. THREAT ASSESSMENT	19
2.3.4. DATA UNAVAILABILITY ON BUSINESS OPERATIONS.....	20
2.3.5. IMPACTS ON BUSINESS WHEN IS DATA UNAVAILABLE	21
2.3.6. BUSINESS IMPACT ANALYSIS	22
2.3.7. DOWNTIME COST CALCULATIONS	23
2.3.8. MANAGING CHANGING VALUE OF DATA	24
2.4. MEASURES TO PROTECT CRITICAL BUSINESS INFORMATION	27
2.4.1. BEST PRACTICE CONCEPTS	27
2.4.2. BUSINESS CONTINUITY MANAGEMENT	28

2.4.3.	IT SERVICE CONTINUITY MANAGEMENT	28
2.4.4.	DISASTER RECOVERY	29
2.4.5.	BACKUP AND RESTORATION	29
2.4.6.	KNOWLEDGE MANAGEMENT	29
2.5.	TO EVALUATE THE LIFECYCLE OF DATA PRESERVATION	31
2.5.1.	DATA PRESERVATION STRATEGIES	31
2.5.2.	DOCUMENT RETENTION REGULATIONS OR POLICIES	32
2.5.3.	DATA STORAGE OR PRESERVATION MEDIUMS	32
2.5.4.	THREATS TO DIGITAL PRESERVATION	32
2.5.5.	THE INFORMATION AND PRESERVATION LIFECYCLES	34
2.6.	CONCLUSION FOR LITERATURE REVIEW	35

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY 37

3.1.	RESEARCH DESIGN AND DATA COLLECTION	43
3.2.	DATA ANALYSIS AND INTERPRETATION	45
3.3.	VALIDITY AND RELIABILITY	46
3.3.1.	EXTERNAL VALIDITY	46
3.3.2.	INTERNAL VALIDITY	47
3.3.3.	RELIABILITY	48
3.4.	LIMITATIONS	48
3.5.	CONCLUSION FOR RESEARCH METHODOLOGY	48

CHAPTER 4: RESULTS 50

4.1.	ANALYSIS	50
4.2.	INTERVIEWS	55
4.3.	CASE STUDIES	83
4.4.	CONCLUSION FOR RESULTS AND ANALYSIS	91

CHAPTER 5: FORMULATION OF GENERIC DATA PROTECTION MODEL 93

5.1.	GENERIC DATA PROTECTION MODEL	95
5.2.	EXAMPLE: DATA CATEGORY A	97
	SEARCH LIMITATIONS	101

CHAPTER 6: DISCUSSIONS 101

CHAPTER 7: CONCLUSION 103

7.1	KEY FINDINGS SUMMARY	103
7.2	PROPOSED MODEL SUMMARY	105
7.2.1	PRACTICALITY OF PROPOSED MODEL	105
7.3	FUTURE RESEARCH	105

REFERENCES 106
INTERVIEW QUESTIONS 110



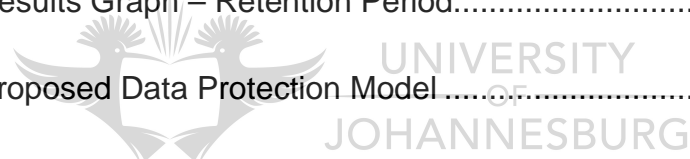
LIST OF TABLES

Table 1: Threat Categories (Bonnette, 2003).....	17
Table 2: Components of ILM (ILM, 2004).....	25
Table 3: Instrument, subject and questions per objective	39
Table 4: Analysis of Results.....	51
Table 5: Details of interview with Lechabile	56
Table 6: Background and findings of International Case Study.....	83
Table 7: Background and Findings of South Africa Case Study.....	87

LIST OF FIGURES

Figure 1: What can go wrong to assets (Backup, 2004).....	7
Figure 2: Hierarchy of Loss (Backup, 2004).....	7
Figure 3: The Viral Effect of Data (Lewis, J: 2006).....	16
Figure 4: Information Timeline.....	22
Figure 5: Recovery Timeline (Interview Supporting Documentation)	57
Figure 6: Information Availability Matrix (Interview Supporting Documentation).....	58
Figure 7: Cost of Data Availability (Interview Supporting Documentation)	59
Figure 8: Results Graph – Business Value of Data – Client Perspective	64
Figure 9: Results Graph – Business Value of Data – Consultant Perspective .	65
Figure 10: Results Graph – Threats – Consultant Perspective	67
Figure 11: Results Graph – Threats – Client Landscape	68

Figure 12: Results Graph – Impacts – Consultant Perspective	70
Figure 13: Results Graph – Impacts.....	71
Figure 14: Results Graph – Alignment – Consultant Perspective.....	73
Figure 15: Results Graph – Data Storage/Preservation and Protection – Client Perspective	75
Figure 16: Results Graph – Data Protection	76
Figure 17: Results Graph – Data Storage	77
Figure 18: Results Graph – Data Protection Methods.....	78
Figure 19: Results Graph – Data Protection – Consultant Perspective.....	79
Figure 20: Results Graph – Data Retention – Client Perspective.....	81
Figure 21: Results Graph – Retention Period.....	82
Figure 22: Proposed Data Protection Model.....	95
Figure 23: Data Category.....	97
Figure 24: Data Protection and Preservation Strategy Formulation	98



CHAPTER 1: INTRODUCTION

7.1 Purpose of study

The purpose of the research is to propose a Data Protection Model to store, protect, preserve and recover critical business data. The aim of the model is to apply different protection concepts, best practices and strategies to data according to the value the data holds within its organisation. Utilising the model will ensure that data is protected according to its value. An outcome of the model is a cost-effective strategy created in a bottom-up approach, to protect mission-critical data in its entirety. Commercial enterprises can use the model as a compliance mechanism or guideline when creating data protection strategies, policies, procedures, and processes.

7.2 Problem statement

Every commercial enterprise should understand the business value of their data and realise that protecting it is of utmost importance. However, company data often resides in different media and locations, so implementing a data protection strategy is not always cost-effective. The challenges that businesses face are: trying to distinguish between mission-critical data from other business data, and excluding any non-business or invaluable data from their systems. From this a cost-effective data protection strategy can be implemented according to the different values of business data.

Against this background, the objectives of this research are to:

- define the business value of data.
- identify threats to digital data and the impact of data unavailability/data loss on business operations.
- identify best practices and current trends deployed to protect critical business information.
- evaluate the lifecycle of data preservation.

- propose a Data Protection Model for preserving critical information.

7.3 Significance of study

The study provides a model as a framework or guideline for data protection by enabling an organisation to make strategic decisions that would ensure their mission-critical data is protected by using the model to shape the strategy in a bottom-up approach. In this way companies that do not have a data protection strategy in place can define these strategies to meet their business requirements in a cost-effective manner.

7.4 Background discussion

Fogleman, (2003) states that information is a tool and must therefore be protected, and that in an Information Technology (IT) revolution there would, for instance, be a fundamental change in the conduct of warfare, indicating the requirement to build a defensive capability to protect the information and data exchange on which powers are becoming dependant on. An observation can be made in ancient history when, according to *Banned Books* (2003), libraries of Egypt, Greece and Rome which consisted of papyrus scrolls, disappeared and what remained were second-hand reports. There are no remains of the libraries and no one is certain of what became of them. The cause of this event was not certain but the loss was said to have been intentional.

Meanwhile, according to Tsamaidis (2003), the Alexandrian Library, which was the great library of Egypt with more than 400,000 scrolls, used for academic purposes, lost an estimated 40,000 papyrus scrolls in a fire. Kahle (2003) compares the Library of Congress's 26 million books, 40 times more than the Library of Alexandria, and argues for the digitising of information. In terms of corporations and company data, Cane (2002) states that there are two types of companies, those that have experienced a serious data loss and those that will lose their data at some stage. Looking back at the Library of Alexandria, the moral of the story may not to be to

digitise data only but potentially to look at different options of protection against different threats, as a loss of information is a loss, no matter what the cause.

Most companies think that their existing data storage and backup plans will protect their information from a massive data loss. This may be false security as an estimated 60 percent of vital data is stored on individual computers with little or no protection: “Corporations have moved critical applications and data from the mainframe to servers and now to desktop and mobile computers. Employees use these computers as a primary tool for productivity in their jobs. Information assets are created and only exist on these computers” (Cane, 2002).

Moore (2003) states that “Storing data is one thing; retrieving data is everything.” This implies that critical information, once stored and protected, would have to be retrievable in the event of a disaster occurring. Background literature to the topic suggests that since business data is critical to decision-making and in setting companies apart, it must be protected from threats that may cause major or minor data loss. Information loss is not a new topic, as seen in the Library of Ancient Egypt example. Reasons for data loss may differ and the format and location of the data may differ, but if data is targeted by any threat there is a potential for data loss, with varying impact.

It could be argued that as a company's greatest asset, information is continually under threat from human error, technological failure, natural disasters and other internal and external factors. These threats need to be identified and quantified, and relevant protection techniques deployed. This will allow businesses to ascertain which data protection strategies to embrace and deploy, thereby highlighting the balance between cost versus value for their business needs.

The literature reviewed in Chapter 2 will focus on different areas of storing, protecting, preserving and recovering mission-critical data, helpful to understanding the different areas so as to propose a generic Data Protection Model encompassing all angles.

7.5 Concept Clarification

Backup and Restore: A backup is a copy of data from one medium to another (Massiglia, 2003). The reason for having a backup is for the data to be preserved in the event that an equipment failure or a catastrophe should occur. Restoring the data is when the data files are retrieved from the backup medium to be used once again (SearchStorage.Com, 2009).

Business Continuity – Business Continuity Management is made up of processes and procedures to ensure that business functions can continue after a disaster or incident occurs. Business Continuity planning is the activity of preventing interruption of mission-critical business services and functions (SearchStorage.Com, 2009).

Data Preservation – “Data preservation does not compel either collection or retention of data; it is essentially a "do-not-delete" order pertaining to existing data. For instance a data preservation scheme provides that particular data that has already been collected can be preserved to prevent its deletion.” (Council of Europe, 2009).

In terms of this study, data preservation means the protection of data against destruction as well as being able to conserve it for long periods of time.

Data Storage – Data storage is storing or holding data in an electromagnetic format to be processed by a computer process when required (SearchStorage.Com, 2009). “Primary storage is data contained in the random access memory (RAM) and other "built-in" devices. Secondary storage is data on a hard disk, tapes, and other external devices.” (SearchStorage.Com, 2009).

Disaster A disaster is an unplanned event which can be derived from many causes both human (attack) and non-human (floods) in turn interrupting an organisations ability to function (Massiglia, 2003).

Disaster Recovery The impact of a disaster on the organisation and the mitigation of the impact of the disaster drives disaster recovery to ensure that business functions could be recovered in an optimal time for the organisation to function. Disaster

Recovery is the science of mitigating the impact of disasters, no matter what causes them (Massiglia, 2003).

Information Lifecycle Management (ILM) – ILM is an approach to manage data during its lifecycle from the time of creation to when the data is deleted or becomes obsolete (SearchStorage.Com, 2009).

In terms of this study, Information Lifecycle Management means that information as an organisational asset changes value over time and this lifecycle would need to be managed.

7.6 Assumptions

In the initial stages there was limited literature in this field, with little more found by the end of the study. The limited resources and literature will therefore influence the outcome of the research. The research project spanned a four year period with literature dating back to 2003. This literature did not limit the researcher's specific objectives.



CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

In this chapter existing research on data protection will be examined. This work proposes a generic Data Protection Model around common trends and findings discussed in this chapter. Furthermore, the results of the literature review will assist in drafting interview questions for this study whereas the aim of the proposed model is to store, protect, preserve and recover critical business data, the review will focus on these four actions and their relevance to a model designed to protect information from possible theft or loss.

2.1.1. Scope of literature

Figure 1 illustrates a number of potential problems facing assets within an organisation, whether documents, applications, operating systems (OS), storage devices, central processing units (CPUs), networks, power and a building (Backup, 2004). According to Backup (2004):

- A “freeze” is when the application or OS does not respond to the keystrokes.
- Corruption is when a document, application, OS, storage or network is corrupted and cannot function.
- A loss is when the document, OS, storage and/or building can no longer be used due to theft or a disaster, such as fire, and therefore needs to be replaced (Backup, 2004).

	Documents	Applications	OS	Storage	CPUs	Network	Power	Building
Freeze		✓	✓					
Corruption	✓	✓	✓	✓		✓		
Loss	✓	✓	✓	✓	✓	✓	✓	✓

Figure 1: What can go wrong to assets (Backup, 2004)

Backup (2004) also highlights the following types of losses that can occur:

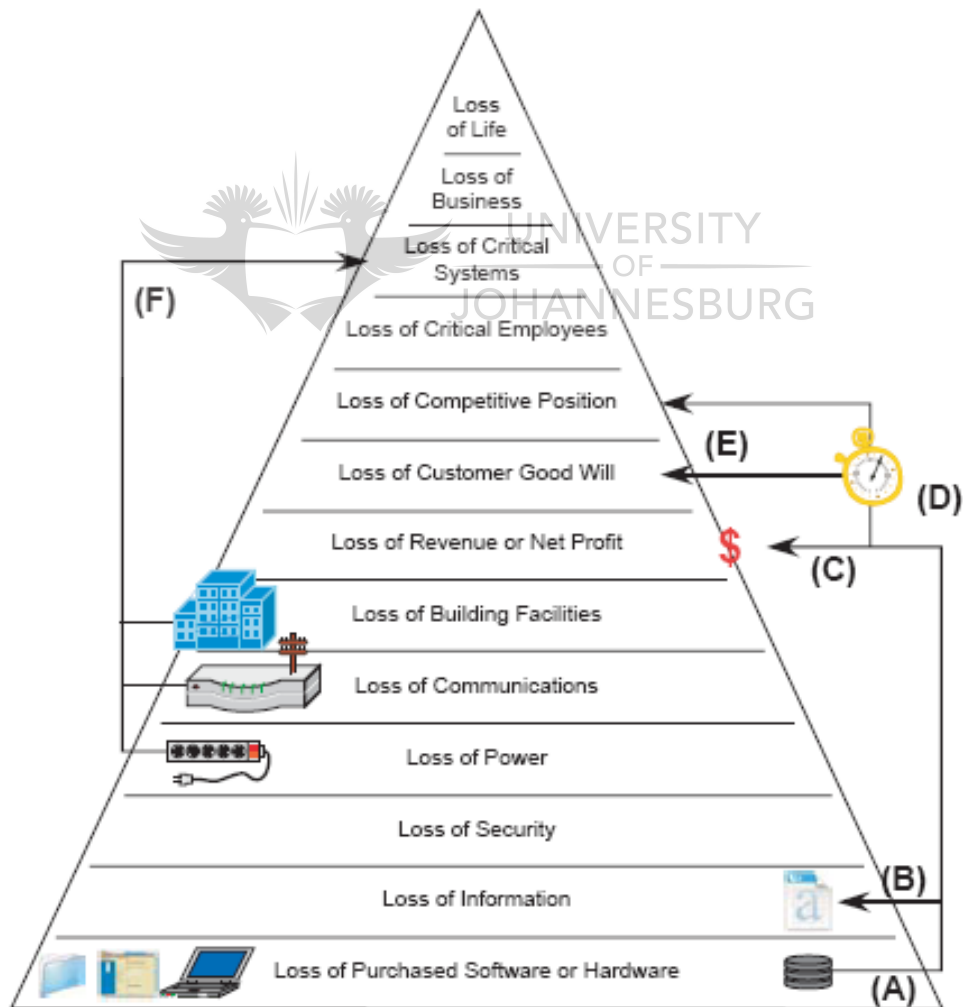


Figure 2: Hierarchy of Loss (Backup, 2004)

This study focusses on the loss and protection of information (B), which in turn will lead to a loss of revenue (C), and over time a loss of customer good will and loss of competitive position (E).

2.1.2. Related Research

In the journal *Protecting Data from Disaster* (1995), different threats to data are highlighted, ranging from magnetic interference and condensation to fire and floods. It is felt that the best methods of protection need to be taken into account alongside backup methods. In addition, it must be noted that the actual media used to make the backup may be as vulnerable as the backed up data itself. *Protecting Data from Disaster* (1995) adds that these threats to data security need to be anticipated and considered. It is important to note that anything that affects the actual premises of a business (such as the building and systems) is a threat to computer-based information. Fire, floods, sabotage and terrorist attack can all destroy a business, and whilst equipment can quickly be replaced, if the data has been lost it will take longer to recover, if it is even possible. A modern business needs to treat its data as if it were money, as well as understanding patterns of 'wear and tear' and expected media lifetimes. Consequently, it can be argued that a company's business data can be seen as an asset to the organisation therefore having a monetary value attached to it.

According to Burnie (2002), conventional data protection methods are not sufficient to succumb to the challenges that businesses face regarding the growth of data needs and the availability requirements of working twenty-four hours a day and seven-days-a-week for business operations. It is becoming increasingly difficult to protect data with tape alone and additional data protection methods must also be employed. Poker (1996) argues that when a Disaster Recovery Plan (DRP) is created it needs to ensure that the plan will allow work to resume with the least amount of effort and to the same standards as before the disaster. A DRP should outline how a business operation would be affected if there had been a power-cut over certain time periods, e.g. 24 or 48 hours. In addition, it defines the scope of

restoration and, establishes responsibilities for actions to be taken once the disaster has occurred.

Furthermore, Poker (1996) notes that there should also be an area describing how a particular disaster could be prevented, as it will be easier to prevent it than it will be to repair the damage it will have caused. Poker (1996), highlights that a risk analysis should be completed to determine when the loss of each application could become critical. Together with a risk analysis, vulnerabilities should be identified and the appropriate protection controls assigned to each vulnerability so as to develop a proactive protection method. For Poker (1996), there are four basic types of threats to computers, namely natural and physical threats, and intentional and non-intentional threats. Disaster recovery plans are not foolproof, but they should increase chances of surviving a catastrophe.

Internationally, there are a number of data protection acts, for instance in the United Kingdom the Data Protection Act of 1984 applies to any corporation, whether or not operating for profit, that uses computers to process data related to people. Each organisation must register with the Data Protection Registry and assure them that they are following eight data processing principles set down in the law, to ensure accuracy of information, confidentiality, and data security. Failure to comply with the law is punishable by a fine (Peers, 1985).

Data Protection and Privacy acts are typically for the privacy of personal information to ensure that it is secure and protected. As mentioned, data protection for the purpose of this research is viewed as the protection of data against information loss and destruction.

According to Stevens (1988), there are three methods of protecting electronic data, namely hardware security, password protection and encryption. For instance, there is a product that puts a band of metal around a *Mac SE* computer and prevents people from accessing data on the computer or easily removing it. Further to this protection method, there are other products that perform a range of functions, including *NightWatch*, which is a comprehensive and secure hard disk password system that provides fast and convenient protection; *Sentinel*, a comprehensive

encryption program; *P-C Privacy*, which allows file purging and easy data transmission; and *Packer*, which compresses and encrypts files simultaneously.

The results of the literature survey show that data protection concepts and methods do exist and are in use. However, having data protection methods alone are not sufficient due to challenges that businesses face in terms of data growth and availability requirements for business operations.

Section 2.2 through 2.5 aims to gain understanding of certain concepts that will lead to the development of a Data Protection Model.

The literature in Section 2.2 through 2.5 linked to the following:

- define the business value of data.
- identify threats to digital data and the impact of data unavailability/data loss on business operations.
- identify best practices and current trends deployed to protect critical business information.
- evaluate the lifecycle of data preservation.

Once the literature has been reviewed, the research methodology and design will be discussed, as well as the analysis and results of the interviews and case studies. Consequently, the Data Protection Model for preserving critical information will be deduced and formulated from the findings and analysis thereof.

2.2. The Business Value of Data

The objective of this section is to define the value of data to an organisation. Each organisation would have different types of data, creating business value within that particular company and contributing to its competitive position. Whether the business data is customer information, accounting records or billing data, it would support the company's core business and provide it with a competitive edge.

For the purpose of this research it is important to understand how different organisations perceive their business data and whether or not they classify it in

different categories. In this way, once there is an understanding of how an organisation values their data and knows where different types are stored, it is easier to ascertain whether storage is appropriate and in the right location to protect and retrieve it accordingly. This objective will be further pursued during interviews to understand the current landscape of the business value of data within corporations. The focus is on data classification and data valuation.

2.2.1. Data as a critical resource

According to Moore (2003), “data is the DNA of the organisation in the Information Age”, and in terms of the uniqueness of company data, each company’s data is unique regardless of where it resides. A comparison between two users’ computers can be made; both having the same type of computer with the same software. However, the data on each computer will be unique, therefore its base of value is created by information technology and not computation alone (Moore, 2003).

Massiglia (2003) states that it is important to know what data is recorded for business operations, as this is critical to enterprise resiliency. An example of this is an airline, unable to function without passenger reservations, maintenance records and supply inventory records. In the same way a software company cannot function without a source code.

This can be argued against the student journal *Protecting Data from Disaster* (1995) that mentioned that a modern business needs to treat its data as if it were money, as well as understanding patterns of ‘wear and tear’ and expected media lifetimes.

In order to understand how this applies in industry, the researcher will use this point in the interview questions to understand whether organisations actually view their business data as a critical resource or asset that has a monetary value attached to it.

2.2.2. Data value changes constantly

Regarding the rate of change of data value, businesses are changing in pace with the rate of change of the value and importance of their data (Croy, 2006). Certain business data is critical to support business processes, whether for decision-making or to inform their employees. This makes the alignment between data value and its storage more difficult to manage. For Croy (2006), data has a lifespan and therefore decreases its value and importance to the business over time. A company should therefore develop a framework to identify and track its changing value of data, so as to ensure that the data is stored according to its value. The framework approach would assist in lowering the total cost of storage and enabling the company to retrieve its most important data first in the event of a disaster. Croy (2006) also points out that the responsibility of identifying the value of data lies with the IT department:

which is required to match the data value with the appropriate storage to support the business functions. It is difficult for companies to determine this value due to the changing value of data in a competitive and regulated business environment.

Croy (2006) suggests that IT departments within the organisation are required to work with the CEO and other business departments to address the business value of data within each department. Further to this there are a limited number of tools to identify the value of a company's data, to monitor its value and store it according to its current value.

This can be linked to Burnie's (2002), statement that conventional data protection methods are not sufficient to succumb to the challenges that businesses face regarding the growth of data needs and the availability requirements of working twenty-four hours a day and seven-days-a-week for business operations.

Since the value of data is changing constantly, the state in which the data is in would therefore need to be proactively managed to close data protection gaps and challenges.

2.2.3. The correlation between data value and data storage

According to Croy (2006), there is a correlation between data value and storage meaning that business data can be classified in a Data Valuation Framework and then stored according to its value, importance and recovery needs to keep the business functioning. This is in line with business continuity and disaster recovery plans to ensure that an organisation's most critical and important data is accessible, accurate and secure at all times.

Furthermore, data storage should enable this continuity, integrity and security of valuable business data but the data management and allocation of data to data storage should be managed more effectively Croy (2006).

Since the objective of the Data Protection Model is for companies to use the model to create a cost-effective strategy to protect its data, the correlation between data value and data storage will be addressed in the interviews.

It could be argued that a cost effective strategy would imply that a company would need to know the value of its business data, categorise the data according to its value and then store it using the appropriate storage medium. In this way, non critical data is not stored on expensive storage medium.

2.2.4. Business value of implementing data protection techniques

Even if the IT department had to work closely with each business department to understand its data value, according to Croy (2006), the problem would be that whilst most business managers say that their data needs to be highly available, few identify which should be accessible and which can be stored in locations that are less accessible. Secondly, organisations have large amounts of data which changes in value frequently. Thirdly, fewer companies can afford the cost of storing all data in a highly accessible and immediately recoverable manner. It is therefore a challenge for business continuity and disaster recovery professionals to address restoring valuable data in a way that is not only timely but also cost-effective for the business (Croy, 2006).

This can be argued against the correlation between data value and data storage in Section 2.2.3 indicating that there is a need to store data on the appropriate storage medium to bring down the complexity and cost of the solution. In this way recoverability of the data will be addressed with the most important data being recovered first for critical business operations.

According to Croy (2006), when looking at the return of investment (ROI), storing data in the most highly accessible and recoverable way is not always the best option in terms of cost effectiveness. For this reason, only a select set of data that is mission-critical needs to be protected in a highly available format and using replication technologies that reduce recovery times to minutes or even seconds.

Furthermore it can be argued that understanding the business value of data and storing it according to its value addresses the challenges of cost effectiveness. Understanding that the value of data changes over time indicates that in order to have a cost effective solution over time, the data that is constantly changing is required to be re-assessed at intervals to change its data storage and protection methods according to its value.

In summary, the literature suggests that commercial enterprises define and classify their business information, using a valuation framework and process to categorise their information and store the information accordingly. This is to ensure that the most important data is highly available and recoverable, and the total storage investment is more cost-effective.

Section 2.3 highlights different threats to an organisation's data and the impacts if business data is not available when required.

2.3. Threats to digital data and the impact of data unavailability or data loss on business operations

The objective of this section is to identify the different threats to information, as well as to highlight the potential impacts of these threats and disasters on the commercial enterprise. In addition, the correlation between data availability and business operations will be determined. For the purpose of this work this section will also look at where data resides, make threat assessments and categorise disasters. In this way an understanding of the threats for corporations will help in ascertaining how to protect the critical data from the threats in an appropriate way. In terms of data availability, the literature provides some insight into the impact on business when data is not available, and the acceptable availability rate for certain data.

2.3.1. Data Discovery Planning

In terms of where data resides, Lewis (2006) states that the location is different from company to company, as each has a unique storage environment. It is thus important that each company understand where their data resides and in doing so determine how much it will cost the company to keep its data in each of the different locations and on different media.

For instance, Lewis (2006) notes that business data can reside in e-mails, among other locations (Figure 3). E-mails are typically hosted on a server, which may contain multiple e-mail accounts. E-mails are also stored on the employee's laptop hard drive. Business information can also reside on an employee's laptops or desktop machine, or on company servers. This is one example of where an e-mail data resides.

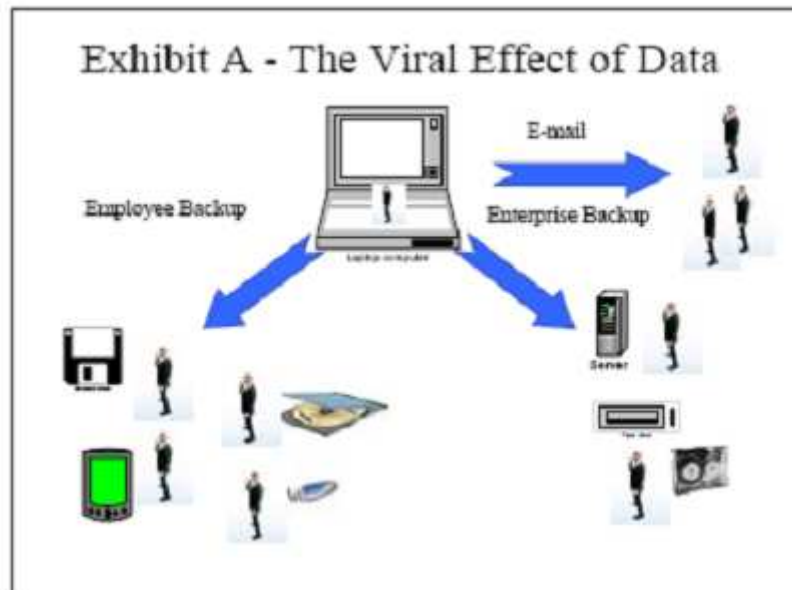


Figure 3: The Viral Effect of Data (Lewis, J: 2006)

2.3.2. Types of Threats

Literature suggests that there are a different sub-set of threats which can impact the availability of data and its business operations. Table 1 outlines the different types of threat in human, non-human and mixed categories (Bonnette, 2003).

This can be argued against the threats mentioned in Section 2.1.2, where different threats to data were highlighted, ranging from magnetic interference and condensation to fire and floods. In addition, it must be noted that the actual media used to make the backup may be as vulnerable as the backed up data itself. The threats to data security need to be anticipated and considered. Fire, floods, sabotage and terrorist attack can all destroy a business, and whilst equipment can quickly be replaced, if the data has been lost it will take longer to recover, if it is even possible.

As Poker (1996) mentions, together with performing the Risk Analysis, vulnerabilities should be identified and the appropriate protection controls assigned to each so as to develop a proactive protection method. There are four basic types of threats to computers, namely natural and physical threats, and intentional and non-intentional

threats. Disaster recovery plans are not foolproof, but they should increase chances of surviving a catastrophe.

Table 1 lists a number of threats in human, non-human and mixed categories (Bonnette, 2003):

- Human Category: Individuals from inside and outside the organisation.
- Non-human Category: This consist of natural disasters
- Mixed Category: This is a mixture of human and non-human threats

Table 1: Threat Categories (Bonnette, 2003)

Threat Categories	
Category	Sub-groups/examples
Human Category	Hackers: Hackers play with computer systems to test its capabilities
	Crackers: Crackers are seen as criminals that steal, destroy and implement denial of systems with malicious intentions in mind against data or systems <i>Example: Cybercrime</i>

	<p>Insiders: Insiders have a certain level of access to the company's systems. They may have malicious intentions and take advantage of the access or permissions they have to sabotage the company.</p> <p><i>Example: Employee Sabotage</i></p> <hr/> <p>Partners: Partners could include third part vendors, business partners, service providers or employees with access to the company's systems. The damage could also be intentional or unintentional</p> <hr/> <p>Competitors: These could be international or local competitors with intentions of gaining competitive advantage by exploiting the company's information systems. Crackers could be hired to perform this or any other method to get to the company's sensitive information</p> <hr/> <p>Terrorists: Terrorists (political or social organisation) would also have malicious intentions in mind with the aim of being disruptive.</p> <p><i>Examples: Information and Software Warfare</i></p>
<p>Non-human Category</p>	<p>Fires, floods, earthquakes, tornadoes, hurricanes and severe storms.</p>
<p>Mixed Category</p>	<p>Malicious code (Trojan horse, virus, worms) created by a person to with malicious intentions in mind.</p>

2.3.3. Threat Assessment

According to Bonnette (2003) that in the information security practice, it is a challenge to identify and assess threats. Performing a threat assessment is a component of an Information Security Risk Evaluation in order to understand threat sources and prioritise vulnerabilities for remediation purposes. In addition, existing security controls need to be evaluated to determine their effectiveness within the organisation.

Bonnette (2003) suggests that the security professionals in a company must understand the source of the attacks, with the likelihood of their occurrence and related impact. "This is important as one in five organisations have experienced a security breach of some nature."

There are two types of companies, those that have experienced a serious data loss and those that will lose their data at some stage (Cane, 2002).

Therefore it can be argued that the criticality of performing a Threat Assessment within the organisation is important as the likelihood that a company will be a target to a threat is extremely high.

Furthermore, with reference to the fire at the Library of Alexandria from Section 1.4, different threats need to be protected against, as a loss of information is a loss, no matter what the cause.

Kaomea (2003) noted that data threats include: fraud and theft, employee sabotage, loss of physical and infrastructure support, malicious hackers, industrial espionage, malicious code, and foreign government espionage. He adds that the detection of many attacks is difficult until the damage has occurred in the information network. For Bonnette (2003), each organisation has its own overall risk assessment programme, to be used to develop and implement methods to evaluate threats based on its unique circumstances.

The following formula shows the relationship between Risk, Threats and Vulnerabilities.

$$\text{Risk} = \text{Threats} * \text{Vulnerabilities} \text{ (Equation 1)}$$

The formula for Risk states that risk (the possibility that “bad things might happen”) is a function of a threat (a source of harm or attack) acting on a vulnerability (a weakness or deficiency in controls). The severity of the risk will also be influenced by the value of information assets that might be damaged or destroyed due to an exploit. (Bonnette, 2003).

2.3.4. Data unavailability on business operations

The time that data is accessible by applications when it is expected to be available is known as data availability (Massiglia, 2003).

Data availability is often measured as a percentage of a year. Example, 99.95% availability equals to 4.38 hours of unavailability in a year for a set of data that is expected to be available all the time (Massiglia, 2003).

$$0.0005 * 365 * 24 = 4.38 \text{ (Equation 2)}$$

High Availability is the ability of a system to perform its function without interruption for a longer period of time (Massiglia, 2003).

In contrast Burnie (2002), mentions that conventional data protection methods are not sufficient to tackle the challenges that businesses face regarding the growth of data needs and the availability requirements of working twenty-four hours a day and seven-days-a-week for business operations.

Thus it can be deduced that conventional and stand-alone data protection methods are not sufficient to tackle the challenges of data growth and the changing value of data. Therefore a hybrid solution or an alternative approach is required to prevent information loss from occurring. Due to the nature of a business to be able to operate twenty-four seven this implies that this solution would need to be highly strategic in nature and not only technology or single-concept driven.

2.3.5. Impacts on business when is data unavailable

Downtime can cause different impacts for a business function (Massiglia, 2003).

Those impacts fall into three categories:

1. Financial impacts – this could be due to losing revenue due to the downtime, or the cost of replacing the system/function or the time taken for an employee to recreate the information therefore the result is an increased expense, or a lost opportunity. E.g. Sales, billing, collections and service functions.
2. Company's customer base – loss of a customer's trust in the product or service or the company's reputation are impacted by downtime. This is critical if the media or a competitor knows about who can strengthen its own product or service to look superior in a media stint. This often leads to losing customers while senior leadership is distracted as they attempt to repair the organisation's reputation. This type of loss has impact on finance.
3. Legal and regulatory impacts - this includes fines and investigations from violation of industry regulations, potential lawsuits from breach of contract, negligence, any other obligations, budget reductions or even closure.

Impacts can be measured:

- The timing of the event (keeping in mind the severity of the disaster)
- The scale of it happening.

Together with this it can be argued that not only financial and customer base impacts, but in the literature regarding the Library of Alexandria from Section 1.4, there could be a loss of information which may not be able to be regained. Therefore there is a loss in the context of the information which may not be able to be retrieved again. Together with this there may be a time period where the information loss implies that information may not be able to be re-created to its original state, or even not being able to recreate the information at all due to loss in context. This may result in “setting a company or society back” in terms of knowing that information.

The Information Timeline in Figure 4 was created by the researcher to illustrate how a possible information loss could influence the initial data that was created by increasing the time to first create the information (Time A). Once the loss has occurred, to trace back as much information as possible (Time B) and then to start recreating the data to as close to its original state as possible (Time C). This also implies that the state of data recreation does not match up exactly to its original state. Furthermore, the context of the data or information's meaning may change over time and recreation.

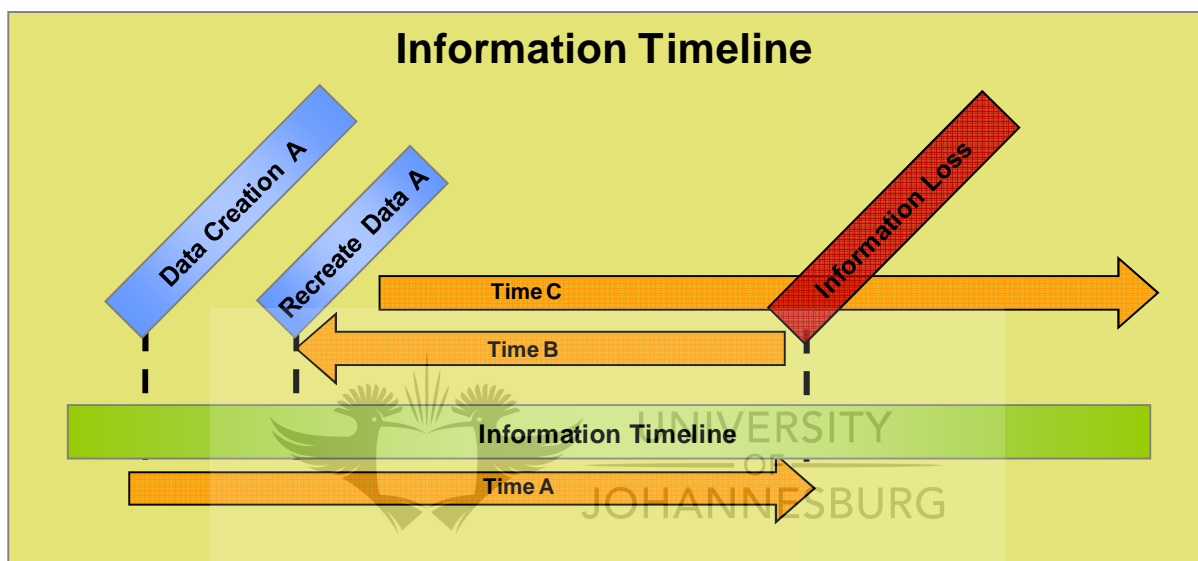


Figure 4: Information Timeline

As a Disaster Recovery Plan (DRP) is created to ensure that the plan will allow work to resume with the least amount of effort and to the same standards as before the disaster Burnie (2002), so would some type of an Information Recovery Plan be required to get information back to the standard and state it was in before the loss occurred.

2.3.6. Business Impact Analysis

According to Wrenn (2005), a Business Impact Analysis (BIA) is used to prioritise the security efforts throughout the enterprise. Alongside the Business Impact Analysis the appropriate human resources are assigned and prioritised for each

incident-response activity. The output is a report which lists the incidents that are likely to occur with its related business or operational impact associated to time and cost.

Aligned to different impacts to the organisation due to downtime or information loss, a Business Impact Analysis could be stated as a relevant concept and method to understand what incidents are likely to occur and the associated impacts if the incident were to take place.

2.3.7. Downtime cost calculations

In order to understand the financial aspects of business downtime or information loss, it would be valuable for an organisation to actually understand the associated costs.

According to (Krayton, 2003), in tolerating downtime, the biggest cost factors are:

- Time to recover
- The value of the time lost to market
- Value/worth of lost or unrecoverable data

When a data loss occurs there are data recreation costs attached to the loss. According to Krayton (2003), the following job recreation cost variables exist:

- Active jobs: Total number of active jobs the user has on her computer at the time of the system crash and data loss.
- Creation time: Time in hours it took (on average) to originally create each of the currently active jobs.
- Re-creation time: re-creating a job takes less time than it did to create it in the first place. This is approximately 75 percent (.75) of the original amount of time.

10 hours to get as far as you did get (factor X), it will take you less time (factor Y) to get to the same place. $X*Y$, or $50 \text{ Hours} * 75\%$ (Formula 1)
--

- Percent lost: in the case when the job cannot be recreated or there is not enough time to recreate.
- Rate: This is how the person who lost the documents cost the company per hour.

Massiglia (2003) states that the financial impact of losing revenue due to the downtime, or the cost of replacing the system or function or the time taken for an employee to recreate the information can result in an increased expense, or a lost opportunity. E.g. Sales, billing, collections, and service functions can cause loss of revenue from downtime. Some of the revenue loss can be permanent or deferred.

2.3.8. Managing changing value of data

In order to optimise a company's storage investment, an Information Lifecycle Management (ILM) strategy is required to keep up to date with the transformation of the value of data in an organisation (Croy, 2006). Since the value of information changes constantly, its general worth can therefore change from being critical to the organisation to being worthless within a short space of time. In this way an Information Lifecycle Management strategy is required to keep track of those changes in value. ILM optimises the storage investment as more critical business information can therefore be stored accordingly (Croy, 2006).

ILM can be attached to the statement that there is a correlation between data value and data storage thereby using this correlation and the management of the changing value of data as a combined effort to optimise a company's storage investment.

In an Information Lifecycle Management (ILM, 2004), data is managed in a dynamic way over a lifecycle. From the time that the data is created to the time it is deleted its value has changed. The business benefit of this concept is to maximise the value of information extracted while at the same time minimising the total cost of storage and management of the data. Therefore, highly valuable data will be stored more effectively. Implementing ILM enables storage as a service. ILM is not to be seen in

isolation; rather it is an enterprise-wide strategy to look at information holistically not as a product “thrown at” a problem.

The operational and tactical reasons for a business organisation to implement ILM are outlined in Table 2.

Table 2: Components of ILM (ILM, 2004)

Operational Components	Tactical Components
Tiered storage	Meet business and regulatory requirements for data retention and access
Data Classification Policy Definition	Provide storage service level
Data Movement	Improve application and file system performance
Data Movement	Lower storage hardware costs
Data Movement	Speed up data management operations (e.g. backup, restore, upgrades or replication)

It can be suggested that an Information Lifecycle Management strategy is required to track the value of data and store and preserve it for business operations and future reference (ILM, 2004).

In summary, the results of literature survey suggest that a company’s data resides in different locations and on different media, and that there are different threats to

digital information and preservation thereof, which have different impacts on the organisation and data loss, as well as cost variables attached to the loss.

The following section highlights different measures or best practices to protect critical business information.



2.4. Measures to protect critical business information

The objective of this section is to identify best practices and current trends deployed to protect critical business information.

The challenge that conventional data protection methods are not sufficient to succumb to the challenges that businesses face regarding the growth of data needs and the availability requirements of working around the clock for business operations needs to be taken into account when reviewing data best practices (Burnie, 2002).

The strength of one concept would need to compensate for a weakness in another concept during the data protection model development in order to have a turnkey data protection solution.

2.4.1. Best Practice Concepts

There is a high demand to store, retrieve and communicate information in order to make it available as needed. The amount of information in general is growing at an extremely fast pace with information growing ten times in the next five years (Moore, 2003).

Digital storage is also growing at an accelerated pace and higher availability of that information is required (Moore, 2003). Therefore the key challenge in the storage industry is the management of data and storage thereof (Moore, 2003).

Global security threats (such as September 9/11 attacks on the USA's World Trade Centre) have made companies aware of the need to look into data protection and disaster recovery technologies from a different angle (Moore, 2003).

Business continuity and disaster recovery plans are required to meet a general business requirement that information is accessible, accurate and secure at all times even when there is an interruption (Croy, 2006).

Data storage should enable continuity, integrity and security. In the past companies operated using filing systems with information which did not change frequently.

Today the complexity of data has increased and it is scattered across the organisation on many different storage devices, servers, networks, and laptops.

2.4.2. Business Continuity Management

Business Continuity Management (BCM) has to do with managing risks to ensure that an organisation can continue operating at a minimum predetermined level at all times.

Business continuity is a proactive process which is a responsibility for the entire business organisation (Synergistic Online Solution, 2009). In this way business continuity should be a business goal with the appropriate planning and tools to support it and in turn protect the organisation's data. To manage the protection of data results in minimising data loss and maximising business continuity. The most important factor to maintain continuity is to have a data backup system, ensure the system is designed to fit the business data model and to ensure that equipment is up and running.



2.4.3. IT Service Continuity Management

IT Service Continuity manages an organisation's ability to provide a pre-determined and agreed level of IT services which support the minimum business requirements following an interruption to the business (NSS, 2003).

IT Service Continuity Management (ITSCM) is part of BCM. IT Service availability is ensured by combining risk reduction measures (e.g. installing reliable systems) as well as providing recovery options (e.g. backup systems and redundant systems) (NSS, 2003).

The aim of ITSCM is to support Business Continuity Management (BCM). It ensures that the IT infrastructure, services and support can be restored within a predetermined time once a disaster has occurred. ITSCM should also be included as a business objective. The scope of ITSCM depends on the outcomes of the risk assessment. By using effective risk analysis and risk management, ITSCM reduces

the impact of a disaster or interruption to the business as well as the vulnerability and risk to the business (NSS, 2003).

2.4.4. Disaster Recovery

A disaster recovery plan covers both the hardware and software required to run critical business applications and the associated processes to transition smoothly in the event of a natural or human-caused disaster. To plan effectively, you need to first assess your mission-critical business processes and associated applications before creating the full disaster recovery plan (Disaster Recovery Best Practices, 2007).

2.4.5. Backup and Restoration

Planning a backup and restoration of files is the most important step to protect data from accidental loss in the event of data deletion or a hard disk failure. The backup copy can be used to restore lost or damaged data (uCertify, 2009).

2.4.6. Knowledge Management

Knowledge Management (KM) involves spreading knowledge of individuals and groups across the organisation in ways that directly impact performance (Seiner, 2001).

The challenge with knowledge is that this knowledge is either kept only in hard copy format (un-searchable), on individual's local computers (un-reachable), or in employees heads (un-recorded) (Seiner, 2001). Furthermore, there is no guarantee that the information is accurate and up-to-date. A knowledge steward is therefore required to manage knowledge.

In summary, the literature suggests that there are best practices and current trends that can be deployed to either safeguard information or ensure that information is available and retained by the company. The safeguarding and availability of information can be done by looking at preventative measures and ensuring that

applications needed by employees and information is available, and recovered in the event that a disaster occurs.

The following section evaluates the lifecycle of data preservation.



2.5. To evaluate the lifecycle of data preservation

The objective of this section is to understand data preservation and the lifecycle thereof. In the previous sections threats to information were discussed, as well as Information Lifecycle Management used to manage the changing value of information. In the same way, the purpose of this section is to understand threats to digital data preservation and to evaluate whether a digital data preservation lifecycle exists.

2.5.1. Data Preservation Strategies

Preservation strategies in terms of research libraries are not unknown concepts, but now organisations face the challenges and issues of data preservation with an increasing amount of digital content, retention periods and storage medium threats (Digital Preservation Strategies, 2006).

As digital preservation is in its “infant” phase it still presents a number of technological issues. For instance since the creation of digital media, over 200 newer and different storage mediums have been invented thereby making previous media obsolete. The computers or systems that use the media also change to be able to use the latest mediums and older ones are no longer manufactured. In addition each storage format needs its own software to interpret the meaning of the data (Digital Preservation Strategies, 2006).

According to Chin (2007), there are four approaches to long-term digital content preservation: technology preservation, technology emulation, content migration and analogue conversion. Content migration is a default option, but with a potential loss of information due to the conversion process. Converting digital content onto analogue media, such as microfilm or microfiche, is a viable strategy, but this is not a good option for frequently accessed records, or those that need to be accessible over wide geographical areas by many users. Outsourcing is an emerging alternative to the traditional four approaches, but technological innovation often lags behind what the enterprise might do on its own.

2.5.2. Document retention regulations or policies

According to Croy, M (2006), there are a number of acts and regulations such as the The Sarbanes-Oxley Act, HIPAA, Graham-Leach-Bliley, and other regulatory changes which create challenges for data storage for different industries and organisations. Document retention regulations require organisations to establish, document, monitor and maintain the availability, authenticity, accessibility, security, and recoverability of their data, and sometimes retain data for some time.

2.5.3. Data Storage or Preservation Mediums

According to Bogossian, M (1998), due to the technology constantly changing, data storage issues arise from both a practical and financial point of view. Furthermore, as data file sizes grow larger, so does the need for more efficient and sizeable forms of storage media. Organisations and users are therefore required to constantly update their hardware in line with the rapid advancement of storage technologies.

Some examples of different storage or preservation media over the past 11 years include: Magnetic media, Diskettes, Fixed hard drives, Tape back-up systems, CD ROM drives, Removable cartridges to name a few.

2.5.4. Threats to Digital Preservation

It must be noted that the actual media used to make the backup may be as vulnerable as the backed up data itself (*Protecting Data from Disaster*, 1995).

The following list highlights threats to digital preservation (Baker and Keeton, 2005) include:

- Massive storage failures: Hardware or software failure or an act of war
- Mistaken erasure: Accidental deletions due to human error or deletion of information before its value has been used.
- Bit rot: No affordable digital storage is completely reliable over a long period of time (CD's have recently been shown to have a life span of only two years).

- Outdated media: Digital media becomes outdated over time. Technology is driven by innovation which leads to very short periods of relevancy before redundancy. Data stored on redundant media becomes useless if the appropriate hardware is not available to read it.
- Outdated formats, applications and systems: As hardware becomes redundant, so do file formats and the software which interprets them.
- Loss of context: Some data can be related, and this relationship can be vital to data interpretation.

An example of this might be the Rosetta Stone, discovered in Rashid, Egypt. The stone is engraved with hieroglyphics in three different languages and without the "key" of what these symbols meant no one was able to read the inscription. It took a French scholar Jean François Champollion fourteen years to decipher the inscription.

- Intentional attacks: These are executed by people who intentionally destroy or damage digital assets for a variety of reasons. Information that is currently located in open access repositories accessible via the internet is also vulnerable to attack. This is a threat to both long and short term storage.
- Lack of resources: Many institutions do not have the resources, usually financial, to consider digital preservation. These strategies are often overlooked as low priority and are likely to remain so until a major data loss scares people into action.
- Organisational failure: This is a threat to long term digital storage of any kind. Technology is so dynamic not only in innovations but also movement with vendors and competition killing off what seemed to be at one point very strong technology players. For this reason it would be short sighted to rely too heavily on any one vendor or system or sponsoring organisation because they change and often change quickly. Digital assets which need to be preserved in the long term must be protected from the failure of any one organisation, which is a challenge in a dynamic environment.

2.5.5. The Information and Preservation Lifecycles

Information can be seen as the lifeblood of a business, and the differentiating factor between companies competing in the market place. Furthermore, information is required to keep business operations running optimally.

Information has a lifecycle and over this period the value of information changes. During the lifecycle, information is required to be stored and protected using relevant technologies, processes and procedures and data protection strategies. Together with the lifecycles the organisation needs to ensure that this information is available to be used for business operations. In the case of information not being immediately retrievable for business operations or for employees to continue with their duties, the information needs to be recoverable to be used for its business purposes.

During the lifecycle of this information, the data is to be stored in different formats (hard copy, soft copy, and in e-mails, to name a few). A soft copy of the information may be on a particular storage device or medium. As technologies change and get outdated over time the data would not be accessible from these storage technologies.

During both lifecycles, information and the technology mediums on which it is stored are exposed to a number of threats which could alter the information, and ultimately its meaning, purpose and value and integrity or destroy the medium on which the information exists. These threats could be intentional or unintentional (with reference to the Library of Alexandria in Section 1.4). Either way this poses a number of challenges to an organisation in terms of getting that information back altogether, or to the level of meaning that it was before.

The challenges that an organisation faces if information is either unavailable for business operations or if there is information loss varies from having financial costs such as re-creation costs to legal and customer base impacts. Re-creation costs are associated with the loss of information in order to retrieve or re-create the information. Legal impacts may be to have a certain type of data for a certain retention period. Additionally, a competitor may have access to an organisation's key and confidential information and use that information to its advantage

For this reason, information must be protected, stored, recovered and available at all times in order for the organisation to operate optimally and cost-effectively. Failing to do so highlights a potential situation whereby an organisation may need to allocate a vast amount of money to a problem rather than implementing a proactive, cost-effective solution to begin with. For the long term data protection and data preservation is required to effectively and proactively manage data. To ensure that information is accessible regardless of the technology it resides on.

The dual challenge is: the changing lifecycle of information on a changing lifecycle of storage mediums and technologies. This poses a number of areas for an organisation to address in order to protect, store, recover and preserve information that is critical to the business.

In summary, data retention policies exist and are based on regulations and company policies. There are different digital storage or preservation media which are prone to threats such as bit rot. Information Lifecycle Management is a way to manage the lifecycle of changing information but there is currently no method to evaluate the lifecycle of data preservation. A possible digital data preservation strategy could be created using a “Digital Data Preservation Lifecycle Management” to ensure that the organisation is aware of changing storage and preservation mediums and keeping up to date with the technology changes of mediums.

2.6. Conclusion for Literature Review

With the challenges of data and information threats as well as preservation threats, the data protection model proposed in this dissertation aims to address these issues in a generic model.

The results of the literature suggest that a company’s data resides in different locations and on different media, and that there are different threats to digital information and preservation which have different impacts on the organisation. Furthermore, there are best practices and current trends that can be deployed to either safeguard information and ensure it is available and retained by the company. These include looking at preventative measures and ensuring that applications

needed by employees and information are available, so as to recover data in the event of disaster. There is no lifecycle for data preservation or retention, and this is based on regulations and company policies. The literature suggests that Information Lifecycle Management utilises the changing storage and preservation media and technologies.

Data needs to be categorised according to its value. There are different threats to be identified and assessed, using a threat assessment (activity of a risk assessment) as well as a business impact assessment to determine the impacts to business if information or systems are not available. Different best practices can be used, such as Business Continuity, to ensure proactive measures are in place. The literature further concludes that data has a lifecycle and its value changes over time. This data is required to be stored and preserved, therefore preservation strategies are required since there are threats to data preservation such as bit rot for the medium on which the data is stored.

In conclusion, the results of literature survey suggest that data needs to be categorised according to its value. There are different threats to data and these are to be identified and assessed using a threat assessment (activity of a risk assessment) as well as a business impact assessment to determine the impacts to business if information or systems are not available. Different best practices can be used such as Business Continuity to ensure proactive measures are in place to protect critical data. The literature further concludes that data has a lifecycle and its value changes over time. This data is required to be stored and preserved; therefore preservation strategies are required since there are threats to data preservation such as bit rot.

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

In the preceding chapters, the theoretical background for the business value of data was discussed as well as the threats to digital data. Best practices and current trends, such as business continuity, were identified, emphasising how the threats could be minimised before they impact data availability in business operations. In this chapter the researcher will describe the chosen research method used for the proposed Data Protection Model. Section 3.1.2 and Section 3.1.3 will cover the interview and case study for the Research Design, followed by a description of the processes used for data collection, analysis and interpretation. The data and information was collected, analysed and interpreted within the qualitative research paradigm.

The general approach conducted for this research was a synthetic or holistic one, to answer the research question and topic of a Data Protection Model to store, protect, preserve and recover critical business data, The researcher strove to understand parts of the problem by looking at the whole, firstly through a literature review, which covered concepts and topics leading up to the model to shape its outline, purpose and form the main ideas. Secondly, the researcher interviewed experts in the field of security, storage, and data protection, and thirdly, case studies were undertaken.

The research method that employed is based on qualitative research, typically used to answer the complex nature of phenomena in natural settings. Leedy and Ormrod (2001) conclude that qualitative research aims to describe and understand the multiple facets of unexplored or under-explored phenomena., the latter being more applicable to this study. The researcher sought to furnish an initial understanding of the background of critical business data, threats to data, data protection and associated benefits. Interviews were conducted with Computer Security Expert and Consultants who had broad experience across industries in security projects and assessments. Due to the sensitive nature of the questions being asked Consultants were chosen for an interview process to illicit trust instead of sending out questionnaires.

According to Henrichsen (1997), qualitative research tends to be synthetic rather than analytic and captures the 'big picture'. The researcher is then able to see how a multitude of variables work together in the real world. The research aim does however start with some preconceived notions or hypothesis, and attempts to discover, understand and interpret what is happening in the research context (Henrichsen, 1997). This researcher's preconceived idea was a Data Protection Model to cover storage, protection, preservation and recovery of critical business data from the possible threat of information loss.

A deductive approach was taken in the research, as there were already established theories and literature that had been used (Yin, 2003). The researcher created a model to address the preservation and protection of critical information and data. *The Resilient Enterprise* was the basis of the literature reviewed, covering concepts such as disaster recovery, business continuity and impacts of data unavailability and downtime. According to Henrichsen (1997), the deductive approach is driven by a particular hypothesis. The researcher has a specific, focused statement in mind and his or her objective is to prove or disprove that specific hypothesis.

In deductive research a hypothesis is necessary, that is a focused statement which predicts an answer to a research question. It is based on the findings of previous research (gained from review of the literature) and perhaps previous experience with the subject. The ultimate objective is to decide whether to accept or reject the hypothesis as stated. In this way, the hypothesis gives direction and focus to the research. Henrichsen (1997) sees subjects as the sources of data, and most research in language-related fields uses people as subjects. Their characteristics, development, opinions, attitudes, knowledge and performance are used to answer the research question.

In order to choose appropriate subjects it is necessary to decide on what the population of interest is. The instruments were used to gauge some quality or ability of the subjects were a literature review, interviews (both semi-structured and structured), and case study. In addition, the degree of control over the research context is low, as qualitative research examines naturally occurring behaviour.

Table 3 outlines the instrument, subject and questions for each objective.

Objectives:

- To define the business value of data
- To identify threats to digital data and the impact of data unavailability and/or loss on business operations
- To identify best practices and current trends deployed to protect critical business information
- To evaluate the lifecycle of data preservation
- To propose a Data Protection Model for preserving critical information (to store, protect, preserve and recover critical business data)

Table 3: Instrument, subject and questions per objective

OBJECTIVE 1: To define the business value of data.		
Instrument	Subject	Questions
Literature Review	General literature available on the Internet or related publications/books.	<ul style="list-style-type: none"> • Determine business value of data and means of storing and protecting. • Is there an advantage of determining the value of business data?
Interviews	Computer Security Consultants. These consultants have consulted to top enterprises within South Africa.	<ul style="list-style-type: none"> • How do companies value their business data? • Do companies store and protect data according to its value? • Would this be a cost effective exercise? Is it practical for companies?

OBJECTIVE 2: To identify threats to digital data and the impact of data unavailability/data loss on business operations.		
Instrument	Subject	Questions
Literature Review	Veritas Disaster Recovery Survey – refer to Appendix	<ul style="list-style-type: none"> • What are different threats to information? • Key assets of company and what can happen to them • What data needs to be available?
Interviews	Computer Security Consultants were interviewed in order to understand their client’s environment regarding security threats.	<ul style="list-style-type: none"> • Deal with interruptions to their business infrastructure? • How often does interruption occur? • Impact of downtime or data unavailability?
Case Studies	World Trade Centre: NYBOT Case Study – The Resilient Enterprise Corporate Consumer Case Study	<ul style="list-style-type: none"> • Critical data and business operations protected?

OBJECTIVE 3: To identify best practices and current trends deployed to protect critical business information (Store, protect, recover business data)		
Instrument	Subject	Questions
Literature Review	The researcher studied literature to determine best practices such as business continuity, disaster recovery.	<ul style="list-style-type: none"> • What best practices are used? • What are current trends? • Does make sense to implement them? • How are organisations dealing with data protection?
Interviews	The researcher interviewed IT Storage Corporates namely Lechabile. Lechabile gave extensive information in the form of presentations, interviews.	

OBJECTIVE 4: To evaluate the lifecycle of data preservation. (Preserve business data)		
Instrument	Subject	Questions
Literature Review	The researcher studied literature such as data preservation strategies.	<ul style="list-style-type: none"> • To understand if there is a lifecycle for data preservation and its storage/preservation medium. • To understand data preservation strategies and any problems with preservation.



OBJECTIVE 5: To propose a Data Protection Model for preserving critical information. (to store, protect, preserve and recover critical business data)	
The researcher utilised the findings from the first four objectives in order to create the Data Protection Model outlined in Chapter 5.	

3.1. Research Design and Data Collection

The research design established three main methods of data collection, namely literature review, interviews and case study. Data was also collected through attending industry conferences, such as the data storage conference held in Johannesburg in 2004.

3.1.1. Literature Review

The Literature Review was divided into four sections in order to gather information pertaining to the first four objectives of this research.

- To define the business value of data
- To identify threats to digital data and the impact of data unavailability and/or data loss on business operations
- To identify best practices and current trends deployed to protect critical business information
- To evaluate the lifecycle of data preservation.

The results of literature review helped in formulating ideas for the Data Protection Model. The results of literature survey suggest that companies rely heavily on business data and the risk of its loss varies as its value changes. *The Resilient Enterprise* (Massiglia, 2003) highlighted this point, though focusing on storage and disaster recovery efforts rather than defining which data was to be protected from which threats. The researcher built on ideas obtained from *The Resilient Enterprise* (Massiglia, 2003) and other literature reviewed to channel ideas into the model to be proposed.

3.1.2. Interviews

Interviews were conducted in two streams: semi-structured interviews with a vendor and structured interviews with computer security experts and consultants. In-depth interviews were performed with two storage and security companies, *Lechabile* and *Bytes Technology*. Information obtained from in-depth interviews with *Lechabile*

employees, including surveys conducted by the company itself for internal company research was gathered.

The first stream of the interview process was an semi-structured interview of a South African enterprise that provide storage and security solutions. The purpose of this interview was to understand some of the solutions and best practices within the security, data protection and storage practices.

The second stream of the interview process was extensive interviews with four consultants in the computer security specialist field. The consultants had been trained in best practices and trends, as well as having been exposed to advisory roles and creating architectures and solutions for corporations in terms of security and business continuity. The consultants interviewed had experience within the security industry and had been consulting for between four and twelve years. The interview was split into five sections: business value of data; threats; impacts; alignment; and methods to protect, store, preserve and retain data for the purpose of “unwrapping” the information and creating a generic model.

3.1.3. Case Study



In the event of a lack of information, Leedy and Ormorod (2001) suggest that the researcher use the case study method. Often case studies are highlighted as being difficult to generalise from one case to another, therefore researchers often try to select a representative case or set of cases. Therefore the researcher generalised findings to theory (Yin, 2003), selecting two representative cases, namely an international real-life disaster, and a South African enterprise.

The data was collected for the case study through involvement in a Business Continuity Analysis project, as well as an international case study being the source of *The Resilient Enterprise* (Massiglia, 2003). Since the study was conducted using qualitative methodology, the data collection procedures or instruments were relatively "loose" or open. Subjects had more latitude in the ways they could respond, and there was more room for the personal judgements of the researcher to enter the dialogue.

According to Henrichsen (1997), the level of explicitness in data collection procedures is also low in this type of research. The data is more impressionistic and interpretive than numerical due to its qualitative nature.

3.2. Data analysis and interpretation

3.2.1. Literature Review and Interview Analysis

In terms of qualitative research, deductive reasoning was used in order to analyse the data. To assist with this, all data collected from the formal interviews and the case studies was logically stored according to specific criteria within each objective, thus assisting the researcher at a later stage of writing up the report. Analysis of the data was based on the knowledge that the researcher had gained from the literature review, in order to better interpret the data gathered from the case study and interviews.

3.2.2. Case Study Analysis

The research provides two distinct types of case study analysis, within-case and cross-case. Within-case analysis implies that the researcher compares the case site against previous theory (this analysis, was employed by the researcher), whereas cross-case analysis implies that case sites are compared against each other. Due to the lack of case sites of this nature within South Africa, cross-case analysis was not utilised.

The added advantage of focusing on one case site is that rich data can be obtained when researchers restrict themselves to single situations (Cornford and Smithson: 1996).

Semi-structured material was analysed, such as the Lechabile's Disaster Recovery Survey report.

3.3. Validity and reliability

Research should be tested for reliability and validity. Validity implies reliability: a valid measure must be reliable but reliability does not necessarily imply validity: a reliable measure need not be valid. Lincoln and Guba (2000) regard validity as determining whether the findings are authentic and in accord with reality, and how other social worlds are constructed. In order to determine whether such findings can be trusted to influence policy and practice, these authors look for rigour in the application of the method and the extent to which the researcher defends his or her reasoning in the interpretations offered.

3.3.1. External validity

External validity is concerned with the “extrapolation of particular research findings beyond the immediate form of the inquiry to the general” (Riege: 2003), that is to what extent the findings from the study can be generalised beyond the scope of the study. The theory found through the literature reviewed has been used to inform the research questions, and was later used throughout the data collection process.

According to Henrichsen, L (1997) external validity is the extent to which the researcher generalises findings to a larger group or other contexts.

Seven important factors that affect external validity are:

- Population characteristics (subjects)
- Interaction of subject selection and research
- Descriptive explicitness of the independent variable
- The effect of the research environment
- Researcher or experimenter effects
- Data collection methodology
- The effect of time.

3.3.2. Internal validity

Internal validity refers to cause-and-effect relationships. Within case study research there is intent to find generative mechanisms looking for the confidence with which inferences about real-life experiences can be made. The research tries to identify what components are significant for those examined patterns and what mechanisms produced them (Riege, 2003). These criteria were used in terms of studying the international and local case studies in order to deduce patterns and understand their causes.

According to Henrichsen, L (1997) internal validity is affected by flaws within the study itself, such as not controlling some of the major variables (a design problem), or problems with the research instrument (a data collection problem).

Factors which affect internal validity include:

- Subject variability
- Size of subject population
- Time given for the data collection or experimental treatment
- History
- Attrition
- Maturation
- Instrument/task sensitivity

The researcher therefore kept the subject size to a minimum, using the interviewees' expert experiences and best practice knowledge as the basis for the results collected. The aim was to take a holistic view of how companies operate within the topics discussed, but it did not make generalisations across these findings.

The same patterns were however found in intricate parts of the research, where formulae or frameworks were not in place for this topic, and therefore utilised by companies.

3.3.3. Reliability

Reliability refers to the extent to which operations and procedures of the research can be replicated by other researchers who then can achieve similar findings (Riege: 2003). A structured approach was taken in terms of gathering the data as well as interpretation of the data. This ensured that the researcher's own bias did not skew results and therefore the reliability of the data. In order to further increase the reliability of the data, only consultants having an objective view of their clients covering of business continuity, risk assessment and security were interviewed. This further ensured the sensitive information would not be omitted, as company employees would be more vulnerable when answering such security-related questions.

3.4. Limitations

One of the limitations of the research relates to its scope, in that it excluded threats associated with other key assets in an organisation, such as the company building and computer networks. However, attention was paid to these assets in the research, as the "houses" of data.

Limitations regarding research matter, such as data confidentiality and sensitivity, were an issue in the early stages of this research. The researcher adapted the list of interviewees to include a defined list of consultants which had a broad knowledge of the subject, as well as a vast experience across industries in order to obtain information which otherwise would be seen as sensitive for a direct company employee to answer.

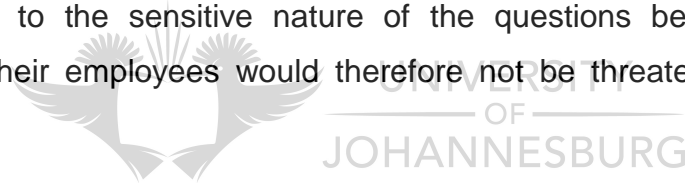
3.5. Conclusion for Research Methodology

The general approach conducted for this research was a Synthetic/Holistic one. The research question and topic of a Data Protection Model to store, protect, preserve and recover critical business data was researched in a holistic and broad approach, so as to cover the main aspects of the studies. The researcher strove to understand parts of the problem by looking at the whole. This was achieved effectively in the

Literature Review, which covered concepts and topics leading up to the model to shape its outline, purpose and main ideas. Further to this, the researcher interviewed experts in the field of security, storage and protection.

The research method that was employed is based on qualitative (interpretive, constructivist, and anti-positive) research. This is typically used to answer the complex nature of phenomena in natural settings. Leedy and Ormrod (2001) conclude that qualitative research aims to describe and understand the multiple facets of unexplored / under-explored phenomena.

Since the study is a relatively new area of research and therefore under-explored, the researcher used qualitative research in the form of the literature review and initial interviews to furnish an initial understanding of the background of critical business data, threats to data, data protection and benefits associated. Following this the researcher interviewed Computer Security Expert Consultants with broad experience across industries in security projects and assessments. Consultants were chosen as interviewees due to the sensitive nature of the questions being asked. Direct companies and their employees would therefore not be threatened by questions posed.



CHAPTER 4: RESULTS

The previous chapter outlined the research methodology that was used in order to perform the relevant research for the proposed generic Data Protection Model.

This chapter contains the analysis and results of the interviews conducted for the purpose of understanding the current landscape of corporations in terms of data protection and storage within the South African context. In addition, the results include a local and international case study, the results of which are intended to help understand the perception of business continuity and data protection within an organisation.

4.1. Analysis

4.1.1. Literature Review and Interviews

Detailed analysis of results was undertaken by gathering the data for the interviews and processing it with an online tool called, *SurveyMonkey* to get statistics on the questions where “yes” or “no” answers were required. Graphs from *SurveyMonkey* were then generated and these were analysed against the interviewees’ comments and the results of the literature review, as well as information gathered from the in-depth semi-structured interview with The Storage and Business Continuity Company.

4.1.2. Case Study

The Case Studies were analysed for any trends and commonalities in terms of behaviours, issues and ideas.

In this way the following analysis has been noted for each objective leading to a Data Protection Model to store, protect, preserve and recover critical business data.

4.1.3. Analysis of Results per Research Objective

Analysis against each objective was as follows:

Table 4: Analysis of Results

Objective	Result (s)	Research Method
To define the business value of data.	Organisation not understanding its true value of business data.	Interviews
To identify threats to digital data and the impact of data unavailability or data loss on business operations.	Organisations not knowing where all key data resides.	Interviews
To identify threats to digital data and the impact of data unavailability/data loss on business operations.	It is required to know where the data resides in the organisation.	Interviews
To identify threats to digital data and the impact of data unavailability/data loss on business operations.	Impacts of downtime are generally known but companies are not aware of full impact of information not being available.	Interviews
To identify threats to digital data and the impact of data unavailability/data loss on business operations.	Costs associated to data unavailability not been known as no formula being used in organisations to calculate this.	Interviews

To identify threats to digital data and the impact of data unavailability/data loss on business operations.	Threat Assessment needs to be performed in order to understand the threats that each organisation is susceptible to.	Interviews and Literature Review
To identify threats to digital data and the impact of data unavailability/data loss on business operations.	Business Impact Assessment is required to understand what impact a disaster or event as such has on the organisation.	Interviews and Literature Review
To identify best practices and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	Best practice concepts used but perhaps in isolation.	Interviews
To identify best practices and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	Budget constraints regarding business continuity or disaster recovery in the organisation as this is not viewed as a critical business function.	Case Study Analysis
To identify best practices and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	Ownership of end to end business continuity not defined.	Case Study Analysis
To identify best practices	Lack of buy in from	Case Study Analysis

and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	business.	
To identify best practices and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	Criticality of certain aspects of Business Continuity not taken seriously.	Case Study Analysis
To identify best practices and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	Companies have disaster recovery plans in place.	Case Study Analysis
To identify best practices and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	Business continuity plan typically not tested and therefore not fully fail proof.	Case Study Analysis
To identify best practices and current trends deployed to protect critical business information. (Storage, protection and recovery of business data).	ILM and Data Valuation not used by organisations.	Interviews

recovery of business data).		
To evaluate the lifecycle of data preservation (Preservation of critical business data).	Data preservation strategies and methods are not the focus in organisations.	Interviews
To evaluate the lifecycle of data preservation (Preservation of critical business data).	Digital preservation is not advanced; data on storage mediums is not being transferred from outdated technologies to updated or new technology mediums.	Literature Review
To evaluate the lifecycle of data preservation (Preservation of critical business data).	No Data Preservation Lifecycle Management concept.	Literature Review
To propose a Data Protection Model for preserving critical information.	It is required to categorise data and understand its value.	Interviews
To propose a Data Protection Model for preserving critical information.	It is required to create a strategy based on the data's criticality in order to have a cost effective solution.	Interviews

4.2. Interviews

Interviews were conducted in two streams: semi-structured interview with a vendor and structured interviews with computer security experts and consultants.

The first stream was a semi-structured interview with a South African enterprise that provides storage and security solutions. The purpose of this interview was to get an understanding of some of the solutions and best practices within the security, data protection and storage sector.

4.1.1 Semi-structured Interview with Vendor

The semi-structured interview allowed the interviewee flexibility of expression in responding to the questions, whilst also enabling the interviewer to return him to the topic in the event of digression. A number of themes emerged from the interview.



Table 5: Details of interview with Lechabile

Company:	<i>Lechabile</i>
Date:	12/07/2007
Interviewee:	Product Manager
Business Focus:	IT storage focus (digital), Business Continuity solutions
Customers:	FNB Namitech: Large security systems, data and access control



UNIVERSITY
OF
JOHANNESBURG

DISCUSSION OF INTERVIEW:

Current Organisational Situations and Challenges

- Backup Islands
 - Difficult to manage and administer
 - Inability to effectively track all media being used
- Slow adoption of improved technologies and methods
 - Solutions available for 24x7 operations
 - Budget and Staff constraints
 - Difficult to justify costs to business for “backup”

New Business Requirements

- Traditional Methods do not meet business requirements
 - Large data capacities
 - Growing “Backup Window” runs into production time

- Cumbersome management of backup media
- No “Real World” Service Levels for business recovery
 - The “Restore Window” is usually never determined
 - Media required for backup and restores is sometimes lost or overwritten
 - Restore tests are never conducted
 - ♦ Are the tapes recoverable?

Is the administrator familiar with the recovery procedure?

Possible Solutions

What is required from the business?

- Recovery times of the business need to be determined by the business

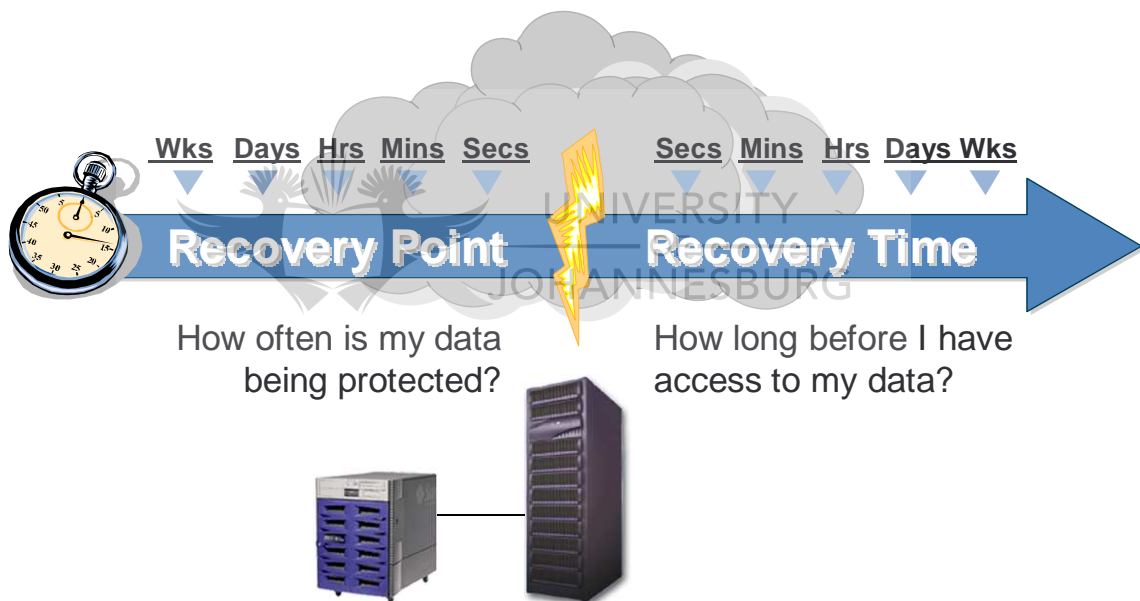


Figure 5: Recovery Timeline (Interview Supporting Documentation)

Cost of Downtime

The cost of downtime ranging from a few hours to two weeks.

Amazon.com	\$ 4.5 Million
Cisco	\$ 30 Million
Dell	\$ 35 Million
Intel	\$ 33 Million
Yahoo!	\$ 1.6 Million

Source: Forrester Research, Feb 2000
Quoted in USA Today, Feb 2000

Uptime

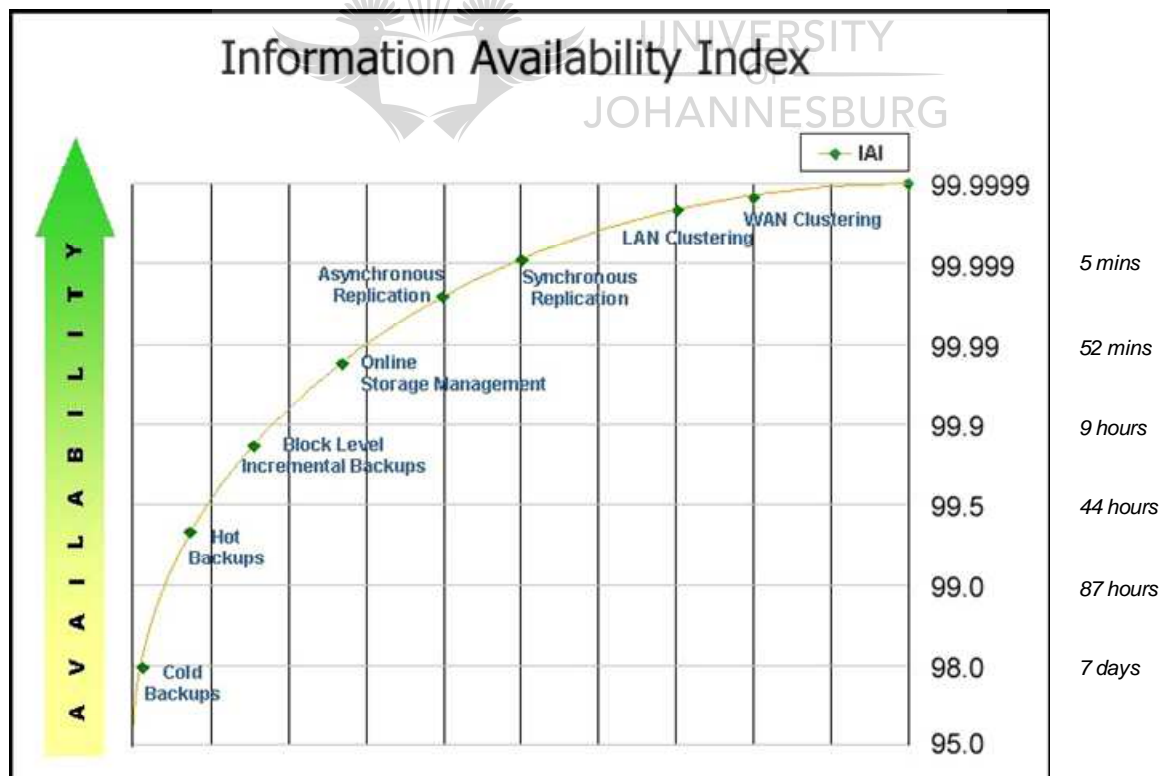


Figure 6: Information Availability Matrix (Interview Supporting Documentation)

Cost of Data Availability

The cost of data availability increases as the extent of availability increases. Therefore a system that needs to be 99.999 percent available is more expensive to run and maintain.

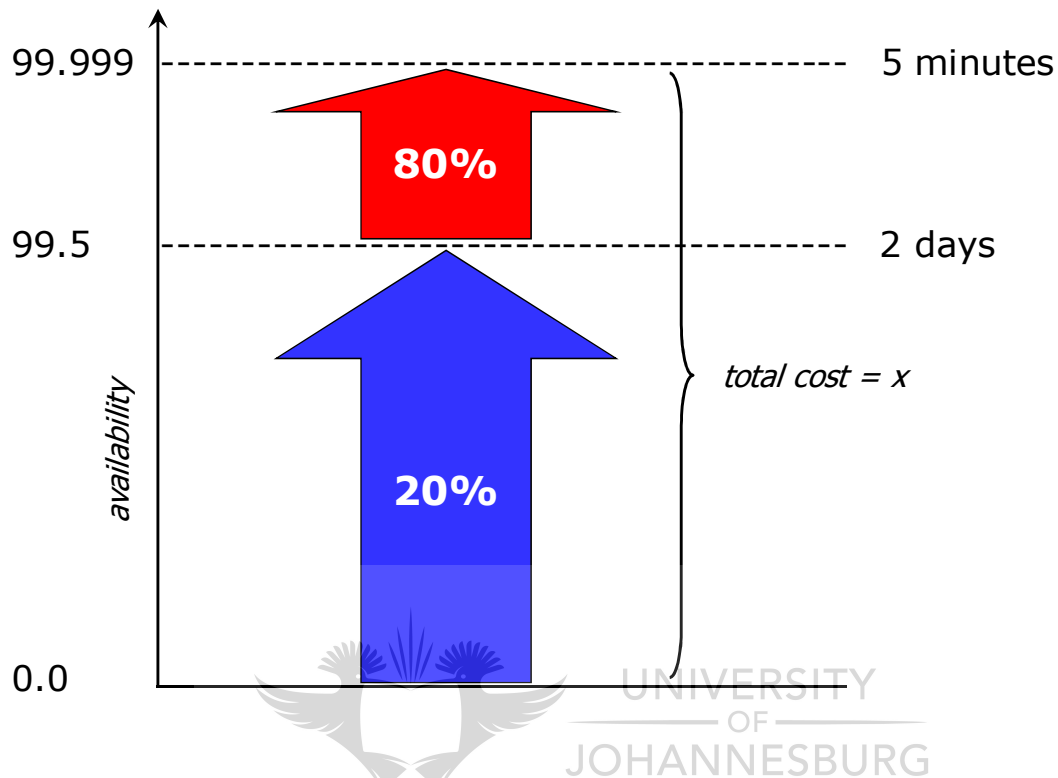


Figure 7: Cost of Data Availability (Interview Supporting Documentation)

The cost of backup

- What is the cost to the business if there is no data available to run the business?
 - Loss of business revenue
 - Loss of potential business
 - Loss of man-hours
 - Threat of business closing if critical data has been lost
- Determine the value of the data to the business
 - Backup does not always receive the necessary priority
 - What are the chances of your business surviving without its critical data?

Reasons to backup

- Mirroring and Replication does not protect against:
 - Data loss
 - Data corruption
 - Viruses
 - Human error
- Site disasters can:
 - Destroy servers and data storage
 - Destroy backup media left on site
- Backup is your last line of defense against total data loss
 - Backup tapes may represent the last valid copy of your data

General

When data is held on the production site, it needs to be protected from local failure, such as hardware failure, fire, flood, as well as logical failure, due to corruption and viruses. The hard disk is mirrored in order to cater for the threat of hard disk failure.

Single point of failure is a way to analyse the infrastructure and storage environment (Cost vs. Risk)



The risk of the server not being able to access data therefore your normal business operations will not work, such as exchange server etc.

For example: If an insurance company stored all their policies on one database, if it was out of use for one hour, they would lose 100 man hours. If they had an array of disks, there would be a problem if the whole array goes down.

Offsite Locations

Most of Lechabile's big customers have off site locations that they back up to. With one client being 150 metres away from their premises.

The concept of off-host backups is used by this client in particular. This process takes place during the night.

Methods of protecting data

5*9s Solution which is used in the developed world at financial institutions especially, but not used in South Africa as it should be.

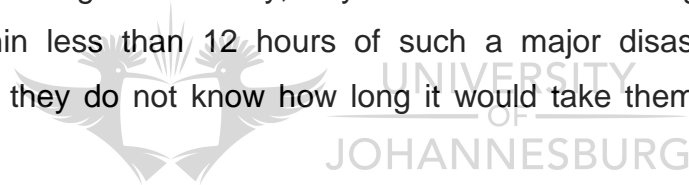
Cases

Approximately three years ago, a South African Telecommunication company had a burst water pipe in their computer room. They had no back-ups and restores in place which was a huge risk, even though business could still continue, they carried the risk of no back-up system.

According to the company survey:

When presented with a scenario where a fire completely destroys the company's main data centre, an alarming 43% of global companies with DR plans in place had no idea how long it would take them to achieve skeletal operations following such an event.

In terms of the next stage of recovery, only 19% would be able to get mostly back up and running within less than 12 hours of such a major disaster - but 45% of companies admit they do not know how long it would take them to operate again such a situation.



4.1.2 Structured Interviews with Computer Security Experts/Consultants

The second stream was extensive interviews with four consultants in the computer security specialist field. The consultants have been trained on best practices and trends as well as being exposed to advisory roles as well as creating architectures and solutions for corporations in terms of security and business continuity.

The results from the interviews are included in Section 4.1.2.1 through Section 4.1.2.6. (Interview questions in Appendix A).

4.1.2.1 Business Value of Data

Each of the consultant's clients understand what their business critical information is in their organisation.

50 % of their clients understand the different types of data the business has, while the other half are not aware of this. Some clients understand the different types of data but do not know the value it holds as they have not used any formula to calculate this.

According to the consultants interviewed, none of their clients define or categorise their data (e.g. Data Valuation). In some cases data valuation would be a regulatory requirement but even then some clients have not even done this. Neither have the consultants seen any models implemented in this regard.

Each of the consultants interviewed said it would be important to determine an organisation's value of data so as to understand what data needed to be protected. For certain industries this is more important, for example marketing where certain legal requirements are implemented under the Protection of Personal Information Act.

The following graphs are used to present findings related to interview questions, i.e. Figure 8: Results Graph – Business Value of Data – Client Perspective, Figure 9: Results Graph – Business Value of Data – Consultant Perspective, Figure 10: Results Graph – Threats – Consultant Perspective, Figure 11: Results Graph – Threats – Client Landscape, Figure 12: Results Graph – Impacts – Consultant

Perspective, Figure 13: Results Graph – Impacts, Figure 14: Results Graph – Alignment – Consultant Perspective, Figure 15: Results Graph – Data Storage/Preservation and Protection – Client Perspective, Figure 16: Results Graph – Data Protection, Figure 17: Results Graph – Data Storage, Figure 18: Results Graph – Data Protection Methods, Figure 19: Results Graph – Data Protection – Consultant Perspective, Figure 20: Results Graph – Data Retention – Client Perspective, Figure 21: Results Graph – Retention Period.

As noted, four computer security experts were interviewed who have consulted to many large South Africa organisations. The purpose of the interviews was to get a broad understanding of the South African context in terms of data protection/storage/preservation functions and perceptions thereof.



The question: *Does your client understand what their critical information is?*, was intended to get an understanding on how companies view the role of their business data within the organisation and how critical it is in terms of business function and revenue generation.

The question: *Does your client understand the different types of data it has and what value it holds?*, was related to different aspects of the data's value within an organisation.

The question: *Does your client define or categorise their data (e.g. Data Valuation)?* Was asked to get an understanding of whether organisations realised that business data could be evaluated and categorised using a Data Valuation Framework for the purpose of storing the data according to its value, as mentioned by Croy (2006).

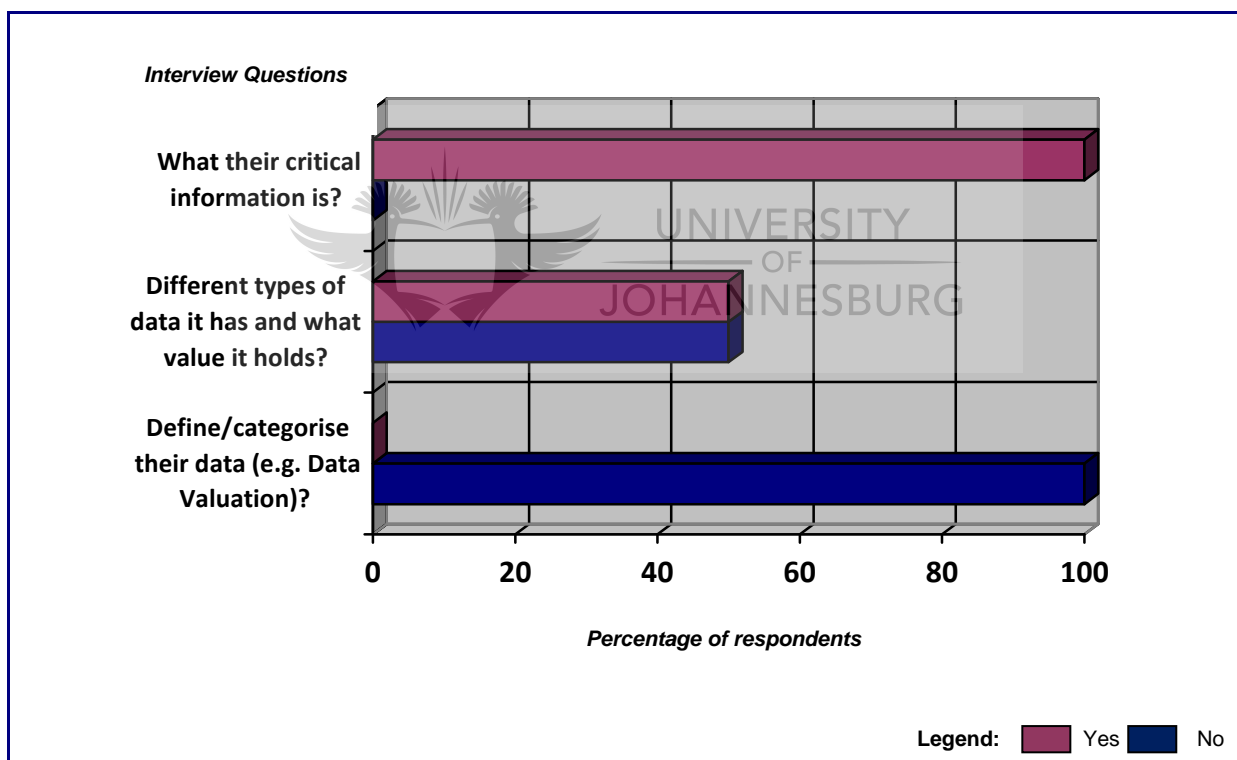


Figure 8: Results Graph – Business Value of Data – Client Perspective

Figure 8 clearly shows that companies generally understand the critical information in their possession but only half actually understand what value their data holds to the organisation as a whole. It is interesting to find that the consultant's clients do not define or categorise their data in order to store it accordingly.

From a consultant perspective, the question: *Do you as a consultant think it would be important to determine an organisation's value of data?*, was intended to understand whether a Data Valuation Framework may be required or favourable for an organisation to use in order to understand in more detail what value is attached to their data in terms of keeping business functions operating, flowing through systems to provide key business data for decision making, whether the data is required by customers for services and whether the data contributes to the organisations' revenue generation.

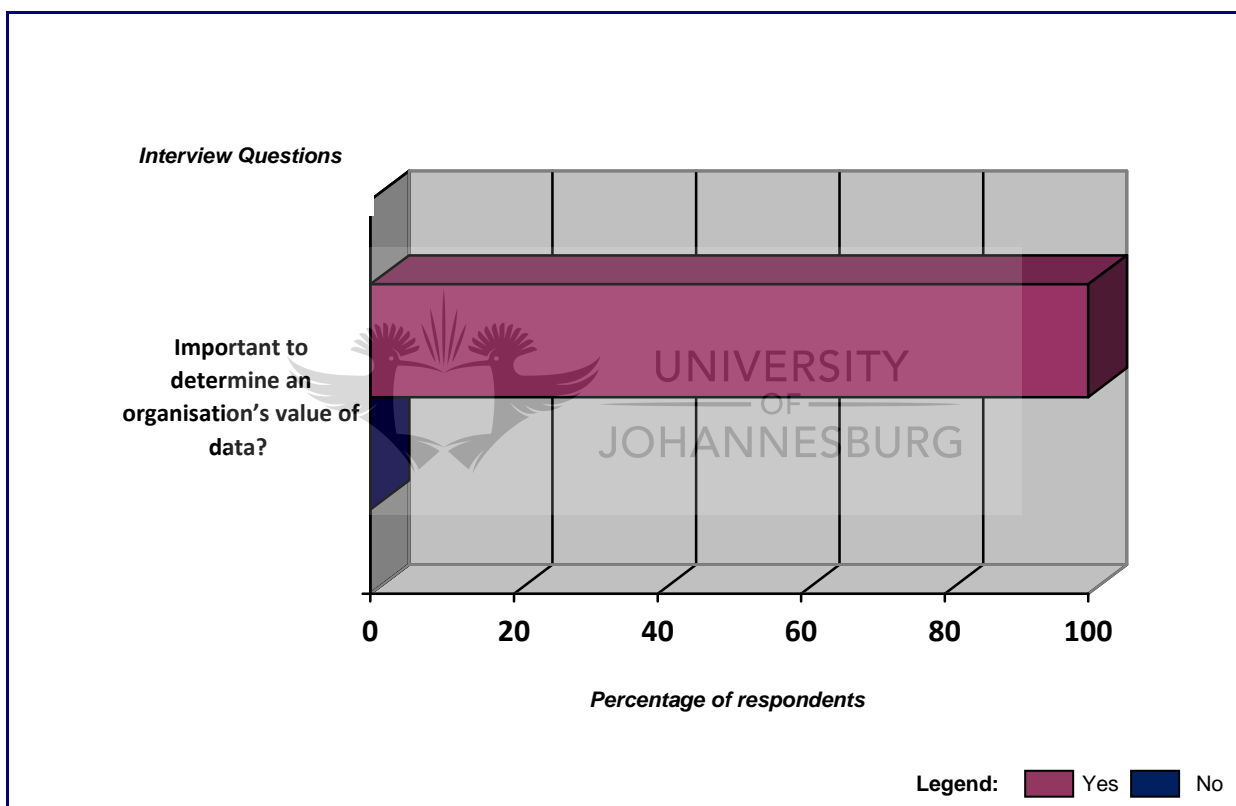


Figure 9: Results Graph – Business Value of Data – Consultant Perspective

Figure 9 indicates that the consultants who were interviewed strongly believe that from a best practice point of view that an organisation should determine the value of its business data.

4.1.2.2 Threats

All of the consultants interviewed said that different types of business data have different threats, for example, confidential data being accessed by competitors or credit card data being accessed by hackers.

Different companies have different threats depending on their location and information they hold. One of the consultants interviewed said that there was more organised crime and this would impact local businesses in Southern Africa.

Clients identify threats to their organisation by performing threat assessments and attack modelling or risk assessments.

The following have been identified as threats to South African enterprises by all consultants' clients: Natural disasters (floods etc), Man-made disasters (war, terrorism), Computer system failure (hardware, software), External computer threats (viruses or hackers), Internal com threats (accidental or malicious behaviour) and Change control issues (patches).

Each of the consultants would advise their clients to perform a Threat Assessment in order to understand the threats to the business as one cannot manage what one does not know.

From a consultant perspective, the questions: *Do different types of business data have different threats? (E.g. confidential data taken by competitors? Credit card data taken by hacker?)*, and, *Would you advise your clients to perform a Threat Assessment?*, related to understanding if there is a certain data that is targeted by specific threats. For instance, an organisation’s confidential and new business strategy and direction may be a target for a competitor to access in order for the competitor to strategically reposition themselves. Furthermore, the researcher wanted to understand whether the computer security expert would or has advised their clients of performing a Threat Assessment.

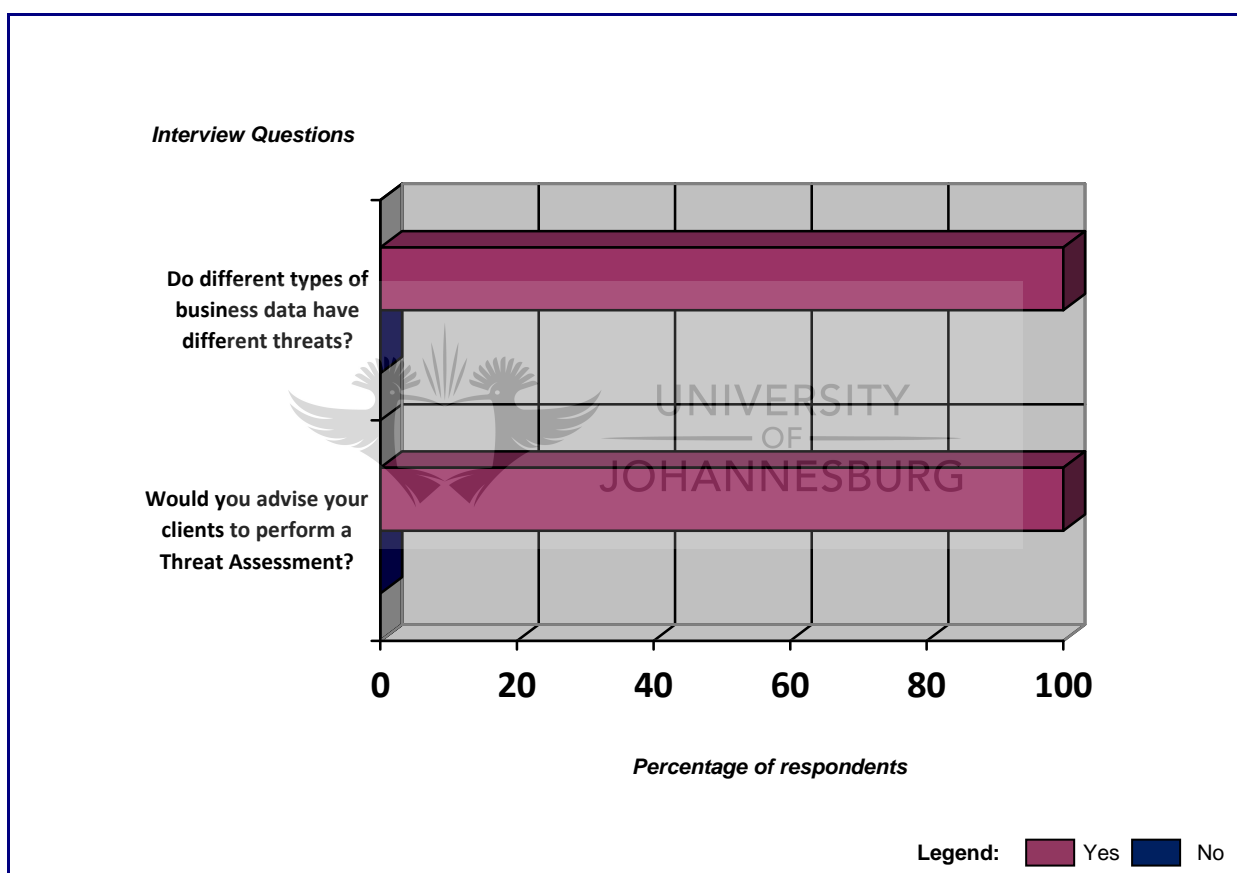


Figure 10: Results Graph – Threats – Consultant Perspective

Figure 10 indicates that the consultants would advise their clients to perform a Threat Assessment to understand the different threats that the organisation would face.

The question: *Which of the following threats has your clients experienced in the South African landscape?*, relates to a Threat Assessment to understand the threats to the organisation.

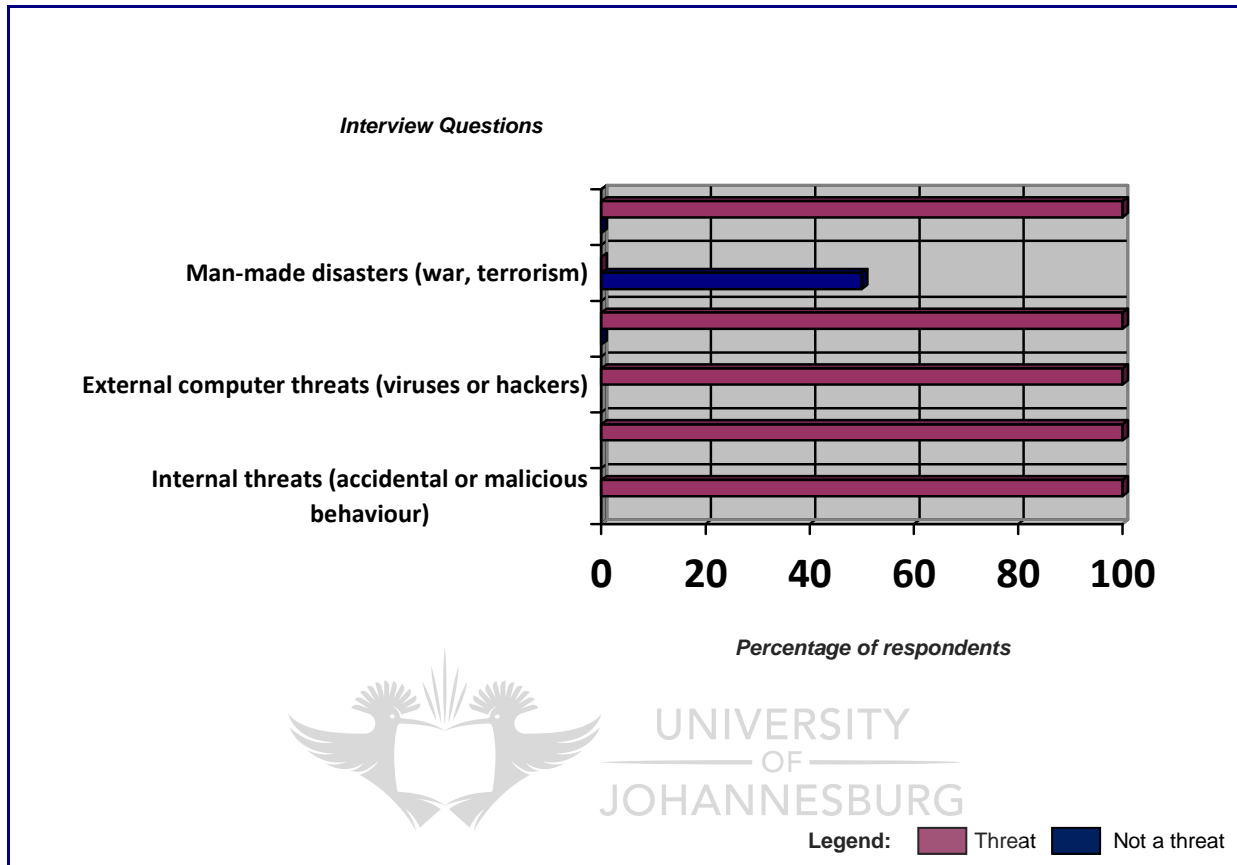


Figure 11: Results Graph – Threats – Client Landscape

Figure 11 highlights that the only area that is not perceived to be a threat to organisations in the South African landscape are man-made disasters such as war and terrorism.

It is interesting to note Fogleman’s statement in Section 1.4 that the conduct of warfare is changing and information as a tool must be protected against Information Warfare. Together with the speculation of the Library of Alexandria’s fire in Section 1.4 being caused due to Civil War opens a number of possibilities for a new threat to be on the horizon.

4.1.2.3 Impacts

The possible impact to the business of data loss is: revenue loss, financial loss via fraud or legislative fines, reputation loss and loss of stakeholder confidence.

The impact depends on the type of data lost, how confidential it is and the financial consequences to the organisation.

Each consultant highlighted different impacts, correlating to the type of data.

Each consultant said that it was important for a company to perform a Business Impact Assessment in order to understand what the impacts are to a business if a threat were to occur. They stated that this would assist in understanding the threats to the business and that it may be needed to prove that the business was an ongoing concern.



From a consultant perspective (Figure 12), there was a 100 percent agreement that different impacts correlate to different types of data and the importance of a company assessing the impact on business is required.

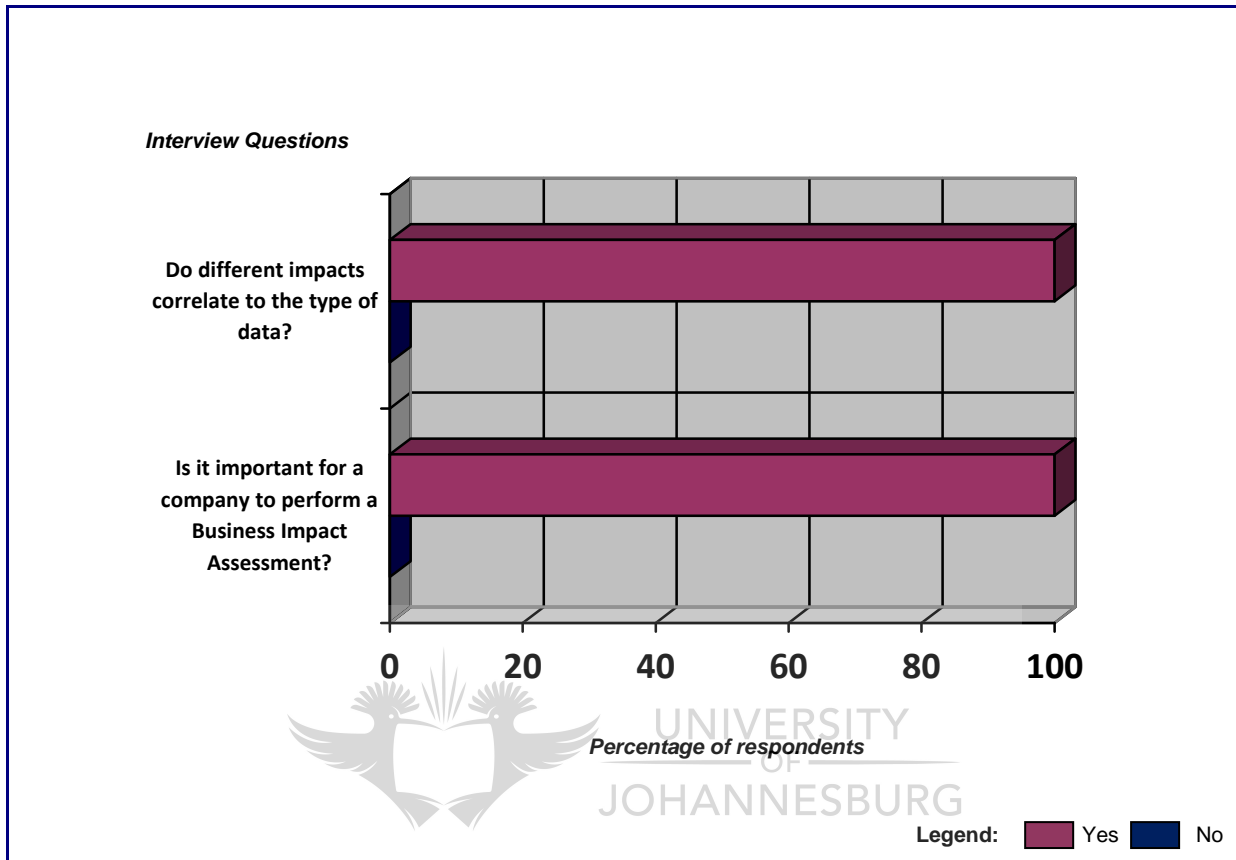


Figure 12: Results Graph – Impacts – Consultant Perspective

In answer to the interview question: *What do the impacts depend on?*, impacts are seen to be dependent on the type of data, how confidential it is and its financial aspects but not the value that the data holds for the organisation (Figure 13).

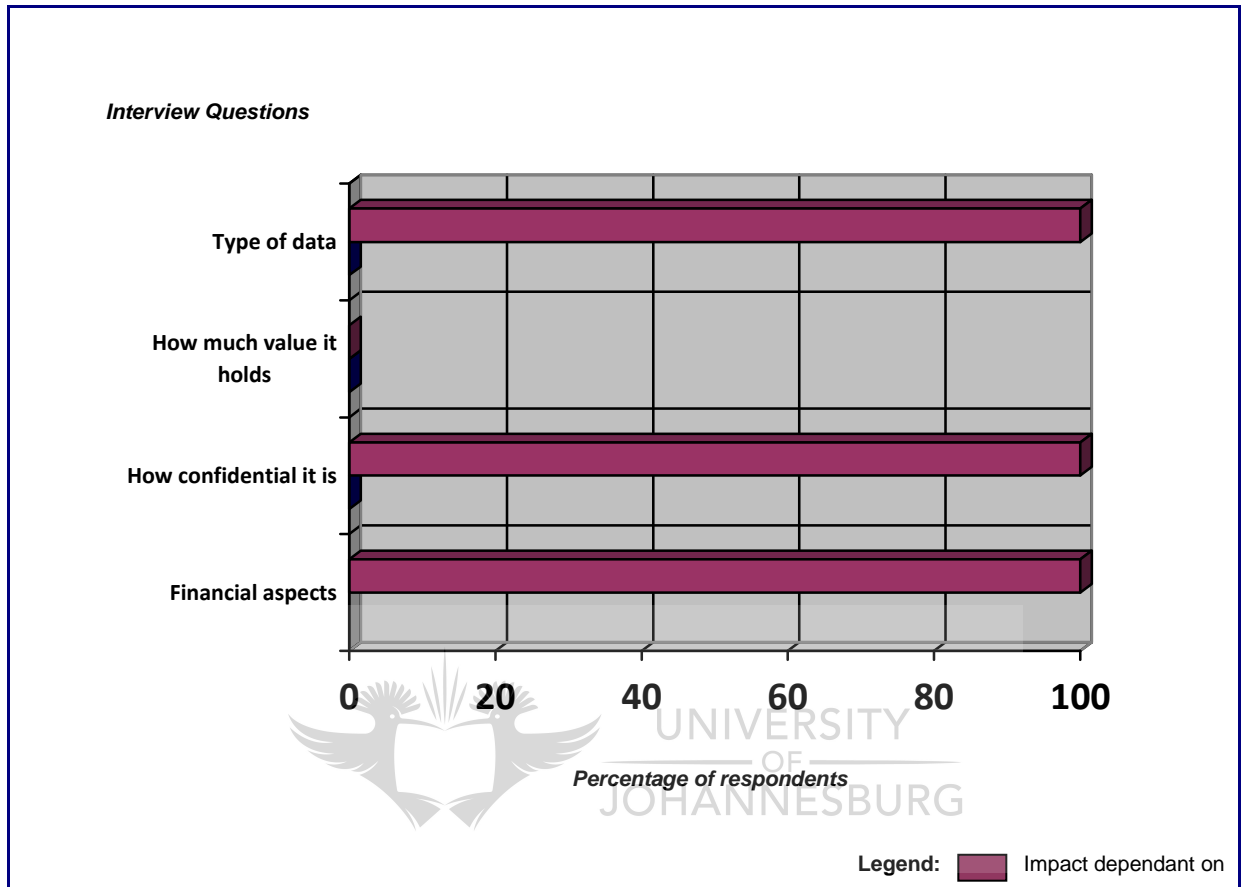


Figure 13: Results Graph – Impacts

4.1.2.4 Alignment

Each consultant said that there was a correlation between the value of data, the threats towards it and the impact on the organisation if it were to be destroyed or lost. This was despite their not having seen many clients conduct such an exercise.

Each consultant said that it would make sense to align the data category, along with its associated threats and impacts in order to protect data accordingly. This is however also dependant on the situation as it may be cost-effective for some companies to protect all data in the same way, i.e. to classify all data as sensitive and apply consistent controls. Meanwhile others said it would make data protection more cost-effective and help organise recovery priorities.

Each of the consultants said it would make sense to have different strategies and methods, depending on the value of data, potential threats to data, and impacts of information loss on the organisation. However, the strategies should be kept as simple as possible.



From a consultant perspective (Figure 14), there was a 100 percent agreement on the correlation between the value of data, its threats and the impact of the threat on that particular data set.

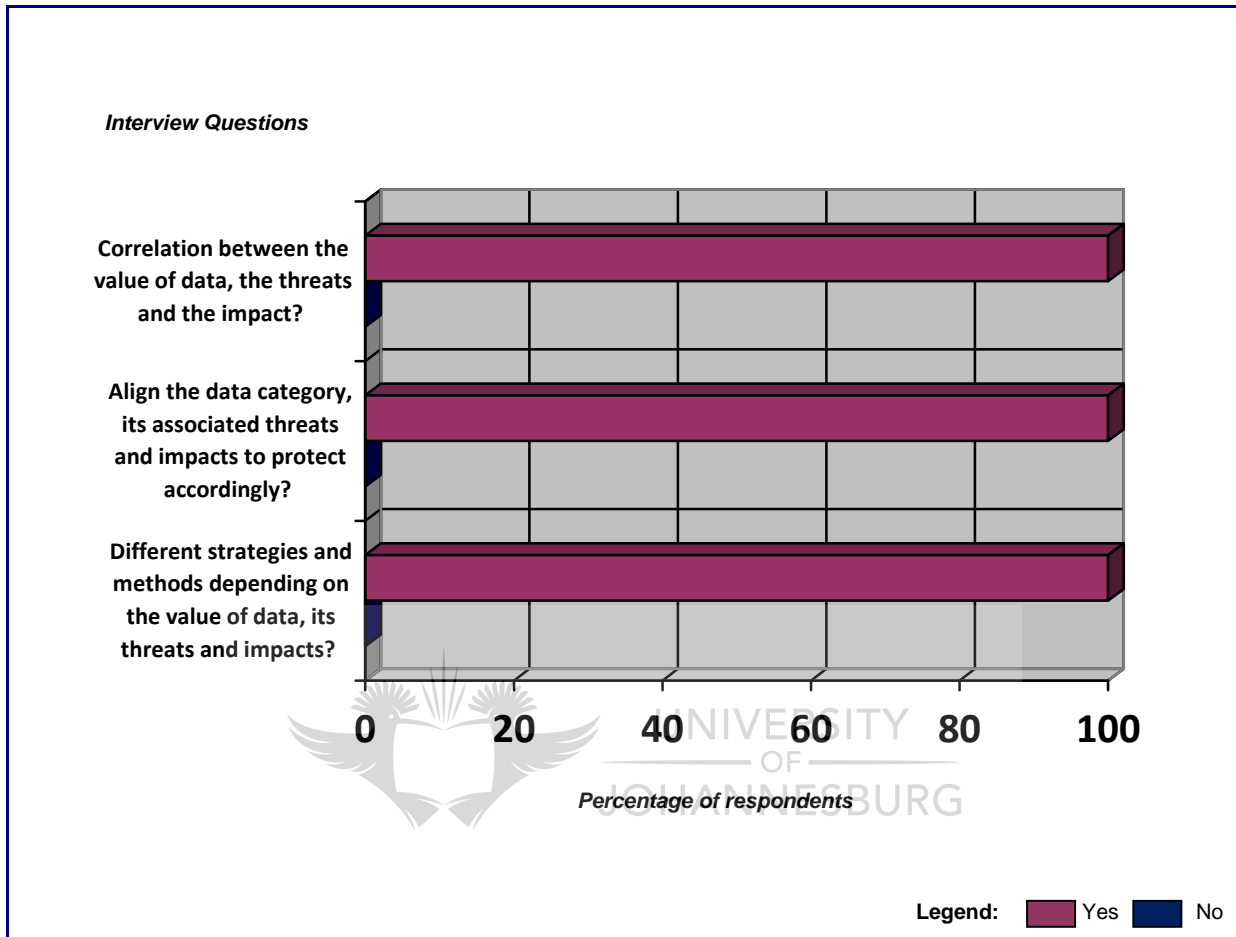


Figure 14: Results Graph – Alignment – Consultant Perspective

In addition different strategies and methods should be applied based on it differing value, threats and impacts.

4.1.2.5 Methods to protect/store/preserve data

A) Data Protection

The consultant's clients have protection methods, such as a Strategy and Policy in place, but only half have Procedures and Processes in place.

B) Data Storage or Preservation

The consultant's clients store and recover data according to its value (e.g. Mission-critical data stored differently from non-critical data).

The consultant's clients have data storage methods and preservation, and Strategy, Policy Procedures and Processes in place.

Data protection or data storage or preservation methods used by the clients being: Business Continuity Planning, Disaster Recovery, Risk Assessment, Threat Assessment, Business Impact Analysis, Storage architectures, but none have Information Lifecycle Management and a Data valuation framework in place.

The clients choose which protection methods and strategies to deploy through the Security lead in business or a consultant, but mainly this is via consulting and through an audit.

Each of the consultants said it would make sense to protect and preserve data according to its value, but this depends on the situation, e.g. financial results prior to release being protected differently from other data, such as financial results that have been released.

For the interview questions relating to data protection and storage, it can be noted that even though data protection and storage methods and general strategies are applied within organisations, the consultant's clients do not store and recover data according to its value as shown in Figure 15.

This would be interesting to note as the interviewee from the Storage and Business Continuity company mentioned that companies typically store all of their employees data (e.g. personal digital photographs, music files, personal emails) which are large in terms of space. This data is stored in the same storage locations and mediums as critical business data. In terms of recovery, there is no systematic way to recover the most critical data first. This means that a company will use critical time trying to recover its data while it is restoring all data including non-critical data. This would imply that a bottleneck of unrelated data will exist in the restoration process.

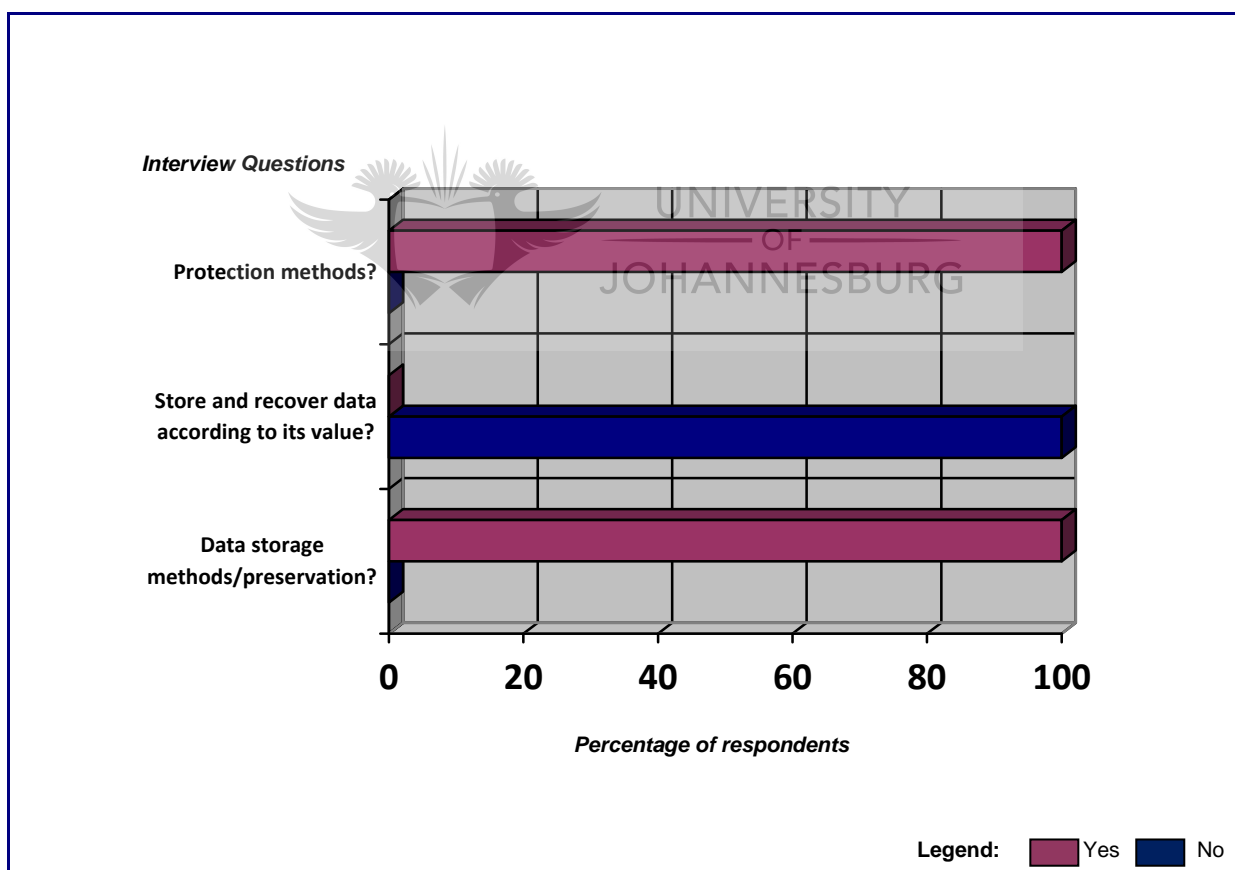


Figure 15: Results Graph – Data Storage/Preservation and Protection – Client Perspective

In terms of Data Protection, strategies and policies are typically in place within the client's environment; however processes and procedures are not fully designed and implemented across every single client. Figure 16 shows that half of the consultant's clients have procedures and processes in place for data protection.

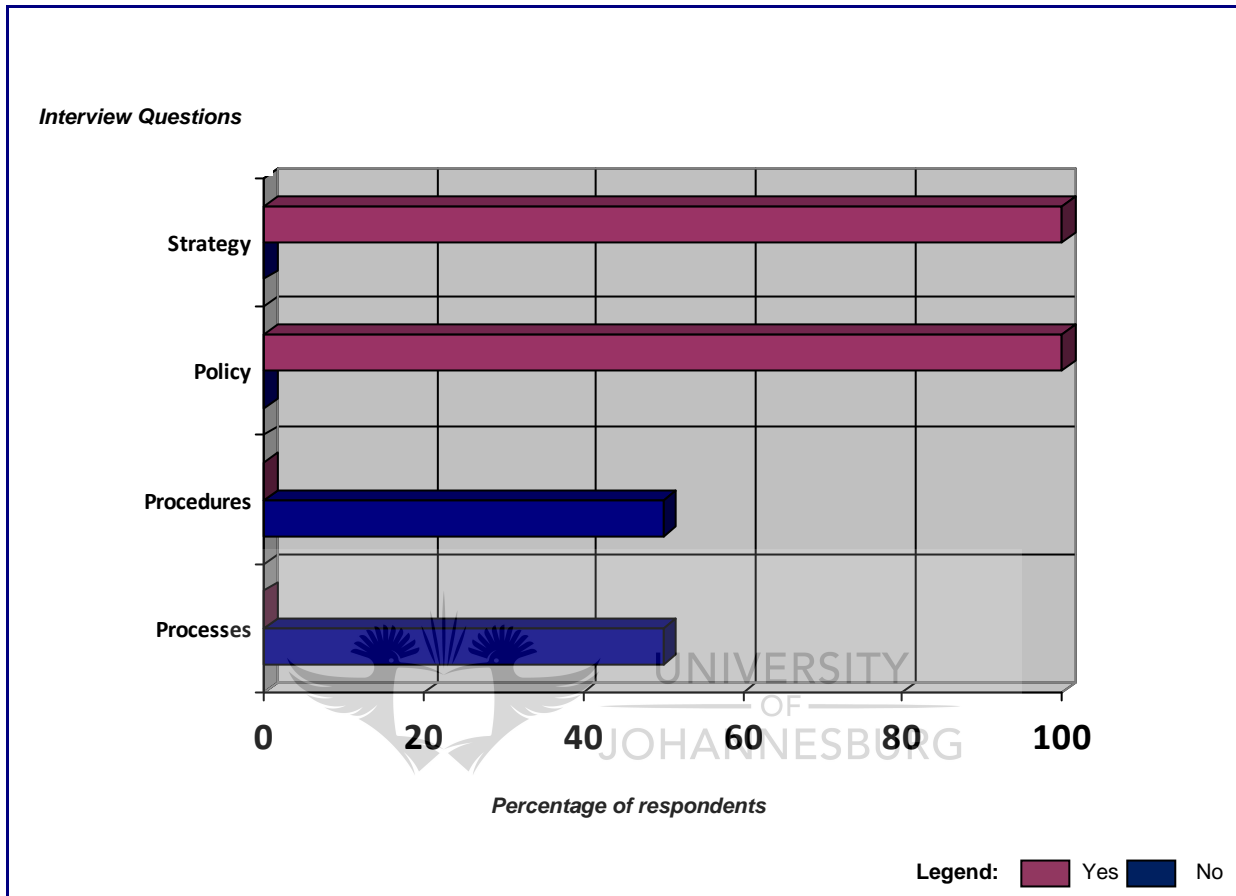


Figure 16: Results Graph – Data Protection

For Data Storage, from Figure 17 it can be seen that data storage strategies typically do not exist across the consultant's client landscape. However, 50 percent of the clients have data storage policies, procedures and processes in place.

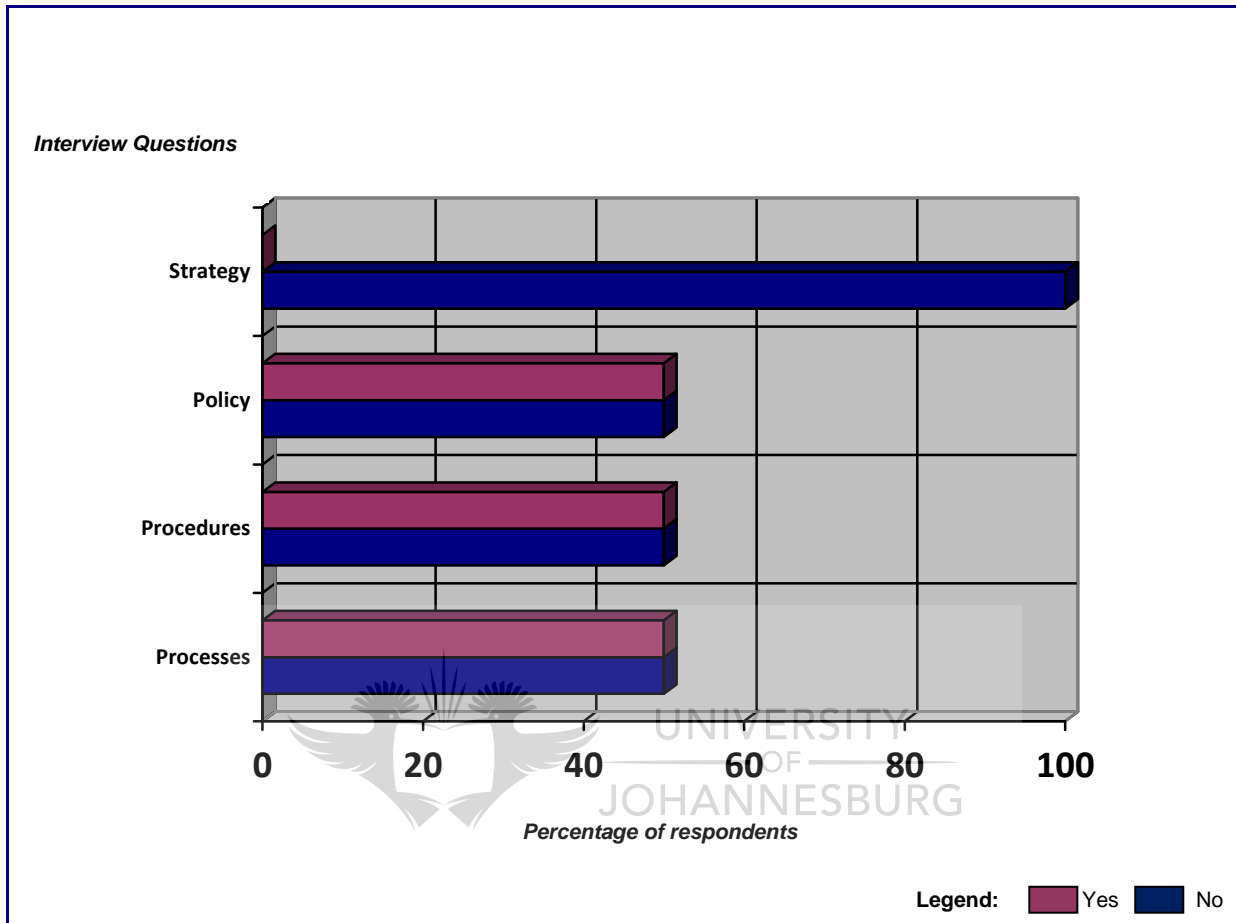


Figure 17: Results Graph – Data Storage

The question: *Which of the following data protection or data storage or preservation methods does your client use?*, related to general methods based on best practices.

From Figure 18 Information Lifecycle Management and a Data Valuation Framework are not in place within organisations, while Business Continuity, Disaster Recovery and the like are in place.

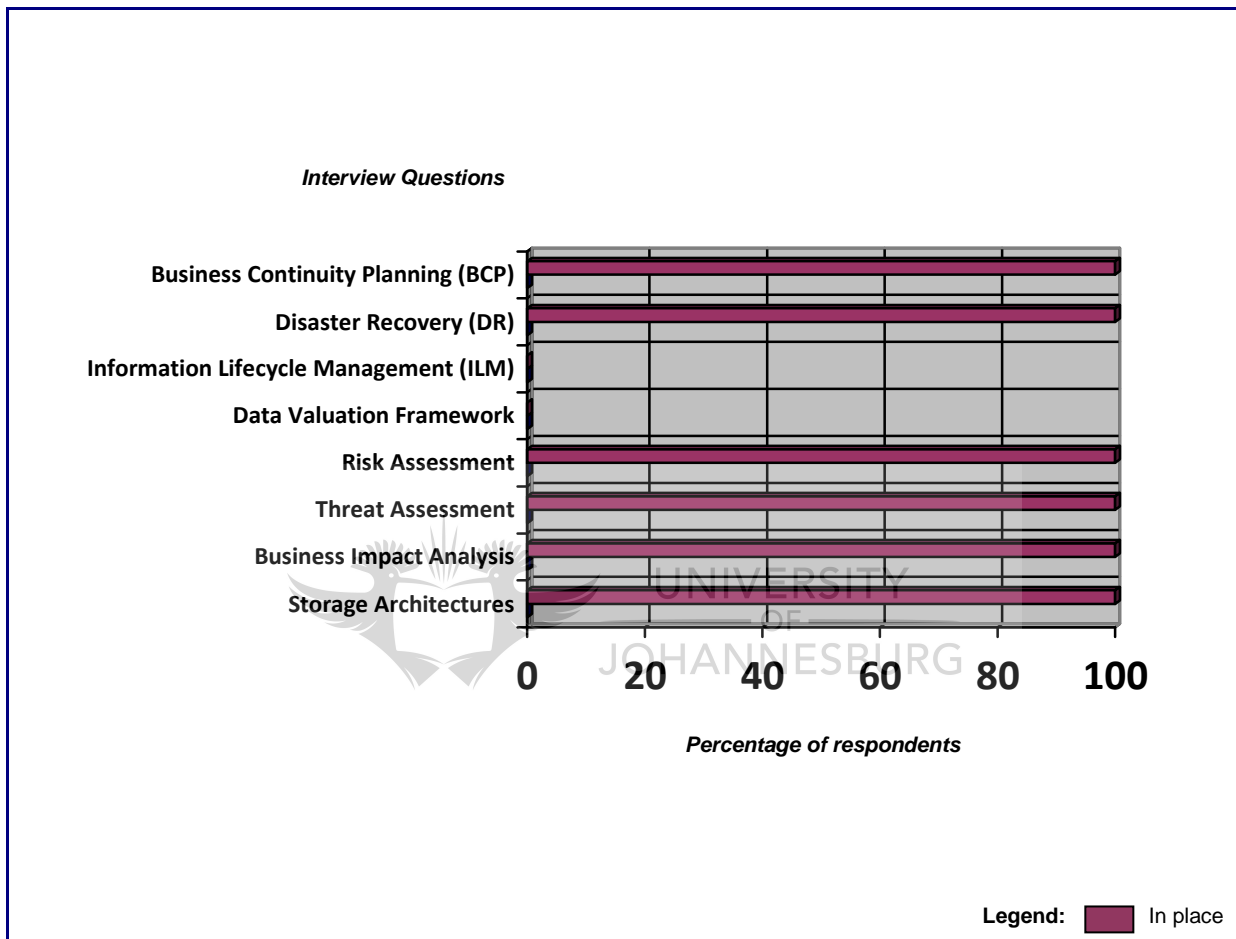


Figure 18: Results Graph – Data Protection Methods

From a consultant perspective, there was an agreement that it would make sense for the client to protect and preserve data according to its value, in addition to have data storage and preservation methods.

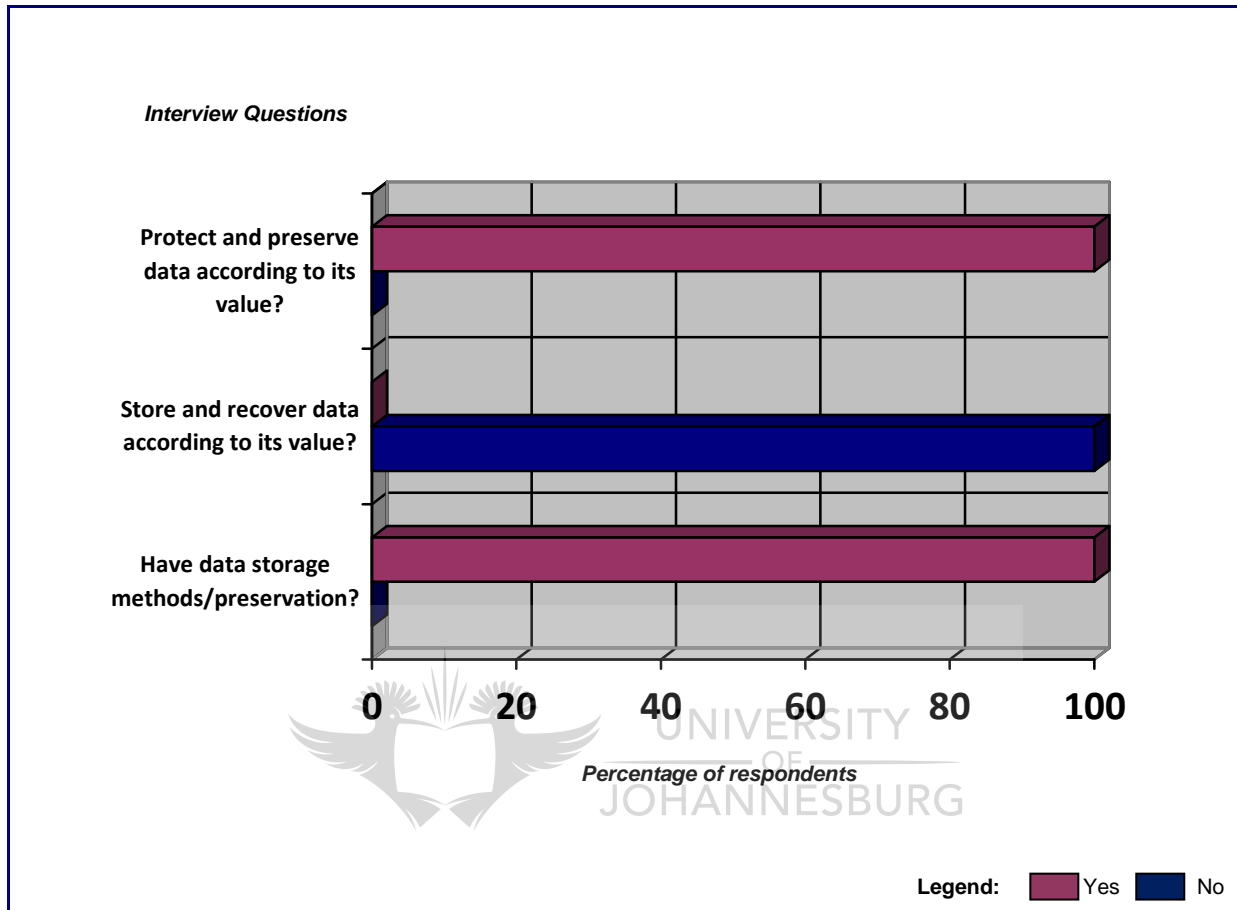


Figure 19: Results Graph – Data Protection – Consultant Perspective

Figure 19 highlights that the consultants do not think that data should be stored or recovered according to its value.

4.1.2.6 Data Retention

All clients have data retention policies to adhere to, but not all keep different types of data for different periods of time.

Regulatory or Legal data retention policies determine the period and not type of data. A company data retention policy is based on regulatory and legal policies.

All consultants said retention periods do not differ for different categories of data (e.g. mission-critical = 5 years) – this is based on legal requirements only.



From Figure 20, in terms of data retention it can be seen by the graph that from a client perspective there are data retention policies to adhere to in order to keep data for different periods of time.

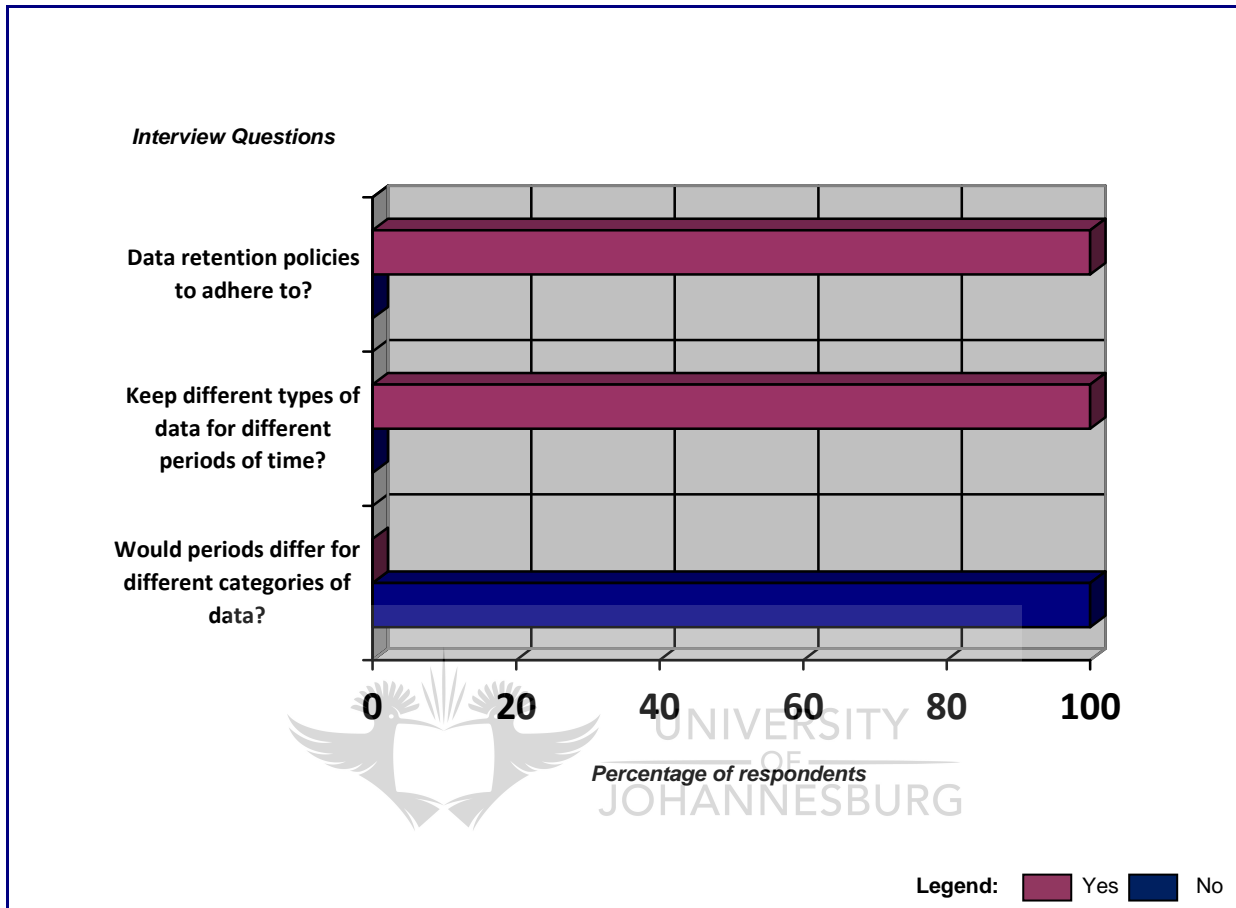


Figure 20: Results Graph – Data Retention – Client Perspective

However, it can be noted that these periods of retention are not different for categories of data.

Furthermore, from the question: *What determines the retention period?*, these retention periods are determined mostly by regulatory or Legal data retention policies and not the type of data or company retention policies as shown in Figure 21.

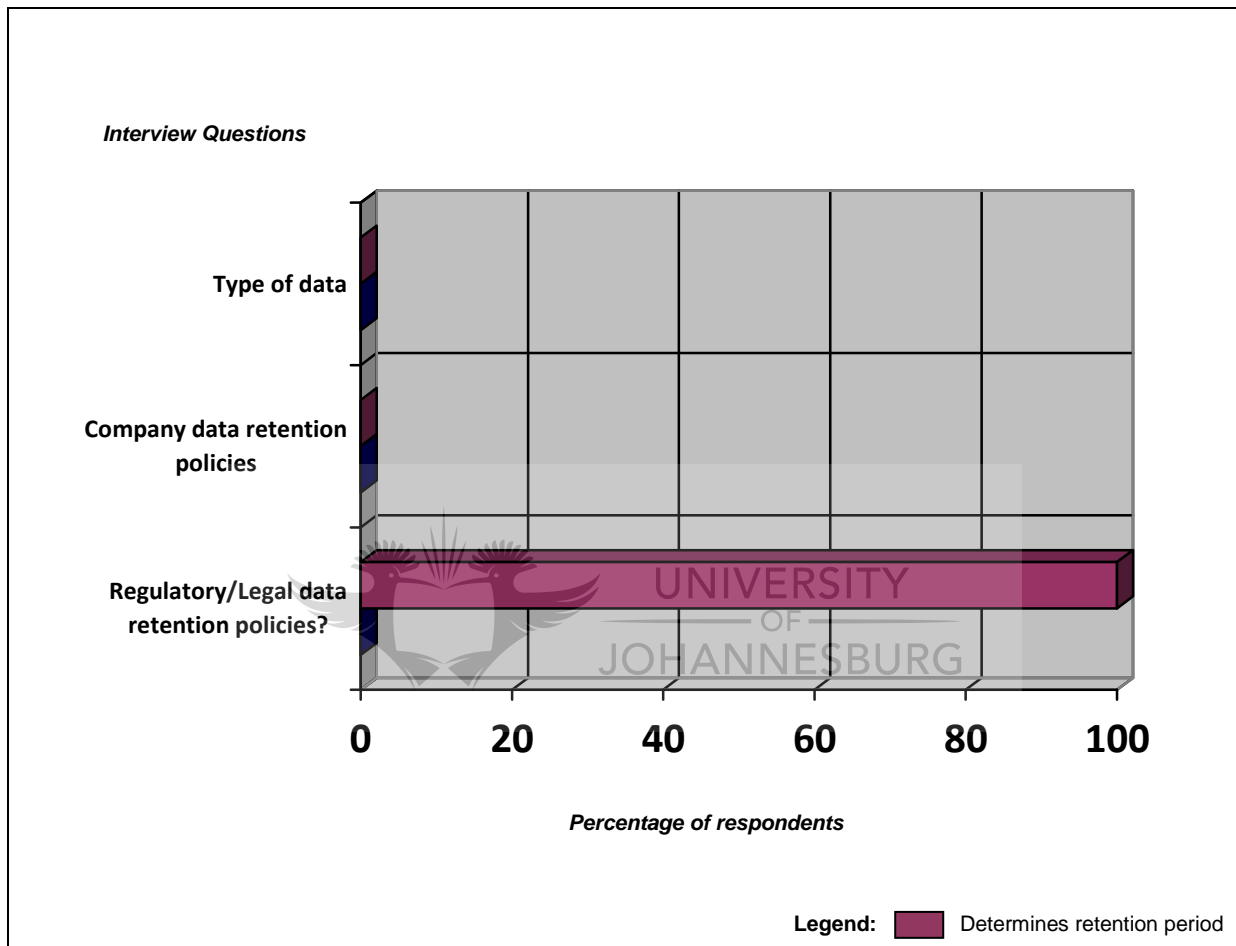


Figure 21: Results Graph – Retention Period

4.3. Case Studies

Two Case Studies were reviewed: World Trade Centre: 9/11 Disaster (in the USA) and the other a South African enterprise's Business Continuity Analysis.

Table 6: Background and findings of International Case Study

Case Study:	International Case Study Results – A Resilient Enterprise
	The New York Board of Trade (NYBOT)
Date:	January to June 2006
Source:	Massiglia, P (2003), The Resilient Enterprise Book
Purpose:	To understand the impact of the 9/11 attack on business in the USA.
Background	<p>NYBOT trades in physical commodities. All information about these activities must be available to any of its members at any given time. There are 1000 members.</p> <p>The NYBOT makes money on commissions, receiving fees for every contract of every trade that is completed. It also makes revenue by distributing commodity pricing information to customers and to ticket service bureaus.</p>
Findings:	<p><u>Event 1: Creation of the Disaster Recovery Plan</u></p> <p>Date: February 26, 1993</p> <p>Bomb exploded in basement parking lot beneath the World Trade</p>

Centre

Coffee, Sugar and Cocoa Exchange (CSCE) and New York Cotton Exchange were located on the 4th floor, (4 WTC).

Explosion caused the facility to lose power, heat, data centre cooling and the building was evacuated.

Cotton Exchange – outsourced its IT processes

CSCE – maintained an IT disaster recovery plan in 1993 (migration to cold backup site at SunGard Recovery Services in Philadelphia). Plan called for computer services and technical personnel to travel to Philadelphia, install an operating system on the cold systems, and install applications and data. The recovery process was expected to take between 24 and 48 hours to complete.

Upon evacuation, IT personnel went to Philadelphia cold site with backup tapes.

Weakness in the recovery plan was that there was not a backup site from which to trade from.

Therefore, even if the CSCE could get its computer systems back online, if it could not regain access to its building, it would take at least 30 to 60 days before CSCE could acquire a new site and install everything necessary to resume trading.

Pat Gambora, senior vice president of Trading Floor Operations and Systems realised this was not appropriate for a major exchange to operate. CSCE required a solid plan in the event that another disaster occurred. Pat developed a thorough disaster recovery plan that included dedicated, hot recovery site for the computer systems and space for trading if the primary space was unavailable (all data copied to the site almost immediately).


	<p>The Board was reluctant to spend the amount of money required but Deloitte & Touche seconded Pat's proposal and did a business case. The Board members were then happy to go ahead with the project.</p> <p>In 1995, CSCE contracted with SunGard for the development and ongoing maintenance of a Business Continuity Plan (BCP) document. It described the recovery of every process and activity (both manual and automatic) within every department in the CSCE should a disaster occur. An Incident Management team was put together to oversee the plans.</p> <p>In 1998, CSCE merged with the New York Cotton Exchange to create the NYBOT. One of the challenges is that the CSEC did its information processing in-house, whereas the Cotton exchange outsourced this function.</p> <p>Pat and Steve Bass (vice president of IT) brought the IT function in-house for the merged NYBOT.</p> <p>For Y2K preparation, they prepared a comprehensive "Member Disaster Recovery Packet". The packet included key radio stations, web sites (DR plan was also on web site), failover to other site if necessary, home phone numbers, addresses, vendor contact details, hotel numbers.</p> <p><u>Event 2: Disaster Struck</u></p> <p>When September 9/11 occurred, NYBOT was able to failover to their site at Queens. Just a few minutes after 8:00am on September 11th, NYBOT had completed its disaster recovery process and its systems were ready to begin trading again.</p>
<p>Conclusion:</p>	<p>The disaster recovery efforts of NYBOT were a success and the company was prepared. There were some problems with</p>

	<p>telephones and lack of space around the trading floors for staging. Nobody had expected the total loss of the WTC site and so NYBOT's disaster recovery plans did not take this into account. Therefore the lesson being to prepare for the worse and prepare for a complete loss.</p>
--	---



Table 7: Background and Findings of South Africa Case Study

Case Study:	Business Continuity Management (BCM)
Client:	Corporate Consumer
Date:	January to June 2006
Source:	Project
Purpose:	Analyse Business Continuity the within department and make recommendations
Background:	<p>A case study was created during a Business Continuity Analysis project at one of the corporations to analyse Business Continuity within a specific department and across the enterprise. The department containing engineering, development, maintenance and support functions of the key systems within the company as well as housing these systems and platforms. The services on these systems provide the company's customers with live services generating revenue for the company on a "per second" basis.</p>
Findings:	<p>Lack or absence of:</p> <ul style="list-style-type: none">▪ Business Continuity is tackled in isolation (no integrated and end to end business continuity plan across the enterprise).▪ Business Continuity and Risk Management processes and Business Continuity procedures not in place.

	<ul style="list-style-type: none"> ▪ Process to update or review plans not scheduled for. ▪ BCP plans were two years out of date ▪ Insufficient resources to work on disaster recovery within the departments for each system and business continuity specialists across the company. ▪ Lack of ownership of business continuity planning outputs (actual plans). ▪ Absence of clearly defined roles and responsibilities for Individuals, Business Continuity Team, Between departments or business functions & 3rd party Vendors. ▪ Poor communication in terms of continuity between Individuals, Business Continuity Team, Between departments or business functions & 3rd party Vendors
<p>Anticipated Results</p>	 <ul style="list-style-type: none"> ▪ Limited company or department budget ▪ Lack of sufficient or skilled resources ▪ Lack of commitment from Management ▪ No roles and responsibilities defined within department or enterprise wide
<p>Preliminary Results</p>	<ul style="list-style-type: none"> ▪ Isolation ▪ Limited company or department budget ▪ Lack of sufficient or skilled resources ▪ Lack of commitment from Management

	<ul style="list-style-type: none"> ▪ No roles and responsibilities defined within department or enterprise wide ▪ Disaster Recovery and Business Continuity concepts used interchangeably ▪ High concentration on storage itself and not necessarily on other protection concepts
<p>Significant Results</p>	<ul style="list-style-type: none"> ▪ Budget ▪ Definition of terms ▪ Implementation ▪ Commitment ▪ Data protection concepts or techniques are used in isolation
<p>Conclusion</p>	<p>Based on the results, the company would not be able to recover from a large scale disaster as Business Continuity is seen in isolation within the departments. In the event of a large-scale disaster one or two departments would perhaps be able to get online to recover information or systems but this is not sufficient as the company needs to operate as a whole and not in isolation. In order to bring in revenue if one department is down the complete service would not be available and its customers would not be able to utilise the company's "per second" services. This particular department is responsible for services and the platforms or servers they sit on as well as for the operations of the systems. Since the completion of the analysis the department has put Disaster Recovery Plans in place for each system.</p> <p>From the Business Continuity Analysis, awareness has been</p>

	<p>created within the department to create disaster recovery plans at a department level. It is now a requirement that each third party vendor and every new system has a disaster recovery plan.</p> <p>This however, does not solve the problem of an end-to-end Business Continuity solution for the business.</p>
--	---



4.4. Conclusion for Results and Analysis

From the results gathered and the appropriate data analysed, the researcher has come to a conclusion that currently there are best practices and trends used by organisations such as business continuity and disaster recovery. From the case studies analysed it can be stated that even though these functions exist within the organisation in general, there are however, challenges, such as business or executive level buy-in and budget for this.

Following this, the analysis highlights that threat and risk assessments are performed within the business as well as different storage practices.

In general, organisations understand the business value of their data but not the actual impact of downtime or unavailability of information. For instance, companies do not know what the calculated cost is in terms of a financial impact if critical data was unavailable for a week? Companies do not seem to have mechanisms or calculations to determine this. Therefore this is perhaps the root cause of business not fully understanding the criticality of having an end-to-end and more cost-effective data protection strategy and solution in place to protect their organisational business data.

Business data is not stored or retrieved according to its value. This seems to be as a result of organisational departments each claiming that all of their data is critical.

In terms of actual preservation of digital business data, there is a gap regarding strategies in terms of the medium used for storage or preservation.

From the results and analysis of the data, the researcher has created a Data Protection Model which addresses and proposes that data is categorised and then stored, protected, retrieved and preserved according to its value within the

organisation, its impacts due to data unavailability and retention period requirements.

The results from this chapter and those of the literature reviewed will form the basis of the Data Protection Model addressed in Chapter 5.



CHAPTER 5: FORMULATION OF GENERIC DATA PROTECTION MODEL

From the research results and the conclusions, we derive the following requirements to be included in the Data Protection Model.

Model creation points:

- Understand where the data resides in the organisation;
- Categorise data and understand its value;
- Threat Assessment must be performed in order to understand the threats to each organisation;
- Business Impact Assessment is required in order to understand what impact a disaster or event as such has on the organisation; and
- Create and define strategies and apply methods according to different data categories and their business value within the organisation.

The model proposed is a generic one to be used as a framework or guideline by an organisation. This model will enable an organisation to protect and preserve its information, according to its business data value within the organisation.

The model was developed by the researcher also based on the ideas, findings and feedback generated from the results analysed in the research for each of the objectives:

- To define the business value of data.
 - Data Discovery Planning (understanding where critical data resides, such as on an employee's laptop).
 - Data Categorisation to evaluate and categorise data according to its value in the organisation. For instance, source code in the Software Development house, billing data for a Telecommunications operator, booking information for an airline.

Once the key business data has been determined, the following areas can be addressed and applied appropriately, based on the type or category of business data.

- To identify threats to digital data and the impact of data unavailability or loss on business operations.
 - Threat Assessment
 - Business Impact Assessment
- To identify best practices and current trends deployed to protect critical business information.
 - This section to be utilised by organisations to deploy best practices and emerging trends, such as Business Continuity and their strategies, processes, procedures forming the basis of the data protection strategy, but having a different angle and focus on the critical data and not on the technologies themselves.
- To evaluate the lifecycle of data preservation.
 - Data would therefore be stored and protected according to its value but preservation is key to the retention and therefore its long-term retrieval. This function would require the organisation to include a data preservation strategy which would need to be assessed and updated frequently as technology and medium for storage and preservation change.
 - A Data Preservation Lifecycle Management strategy to help in understanding the changing lifecycle of storage and preservation media, and therefore being proactive in transferring data from medium to new, updated medium.
- To propose a Data Protection Model for preserving critical information.

The model expresses a holistic view of the topic and does not indicate lower levels of the study. It therefore serves as a framework which would be used by an

organisation, taking into account best practices and trends within the industry. The model was developed to illustrate data as “capsules of business value” and thereby protecting, storing, preserving and recovering each capsule in the appropriate manner, leading to a cost-effective solution.

5.1. Generic Data Protection Model

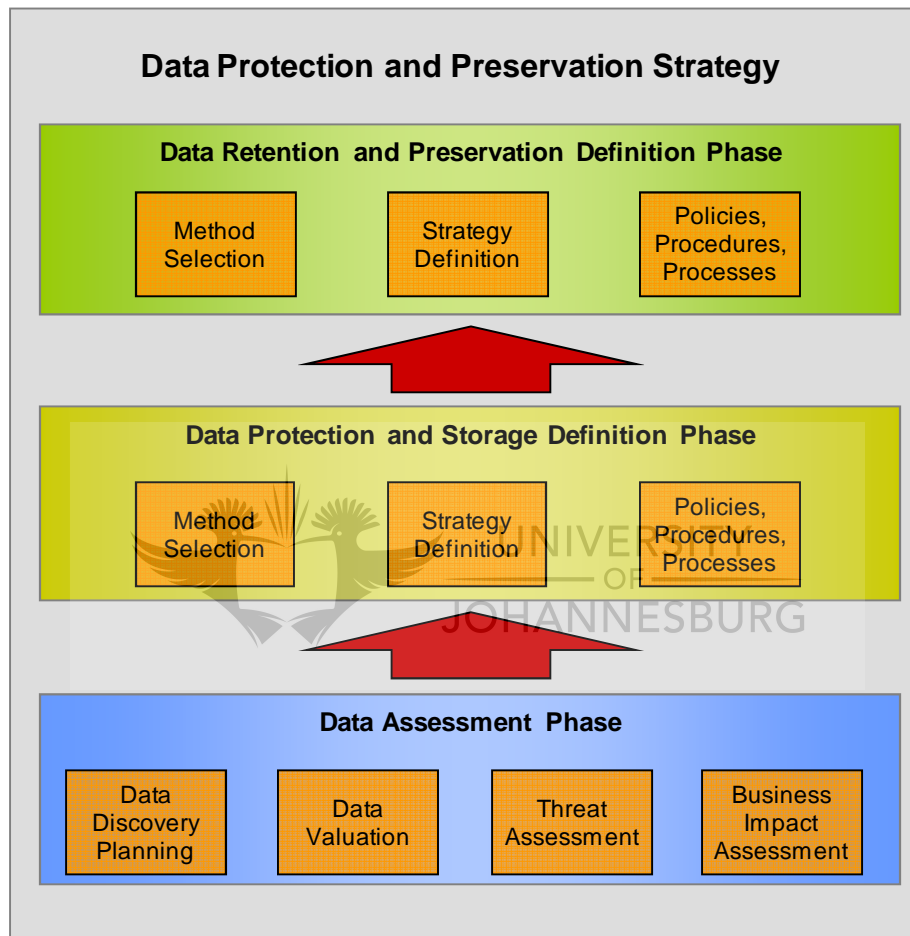


Figure 22: Proposed Data Protection Model

The Data Protection Model is split into two sections: the Assessment Phase and the Definition Phase

The Assessment Phase:

Data Discovery Planning: to understand where data resides.

Data Valuation: to categorise and understand the value the data holds to business.

Threat Assessment: to understand the different threats to the organisation and on specific information.

Business Impact Assessment: to understand the different impacts to the business if certain data or systems are unavailable or downtime.

The Definition Phase:

Protection Strategies: creating a cost effective data protection strategy from the information gathered in the Assessment Phase.

Data Protection Methods: identifying different protection methods and techniques to protect data.

Storage Methods: identifying different storage methods and techniques to protect data.

Policies, Procedures and Processes: creating policies, procedures and processes for data protection, business continuity, backup and restore.

Updates: Protection Strategy with information obtained.

Data Retention and Preservation: identify retention and preservation strategies and methods based on the type of data and its retention period.

Data Preservation Lifecycle Management: understand the changing storage and preservation technologies in order to transfer data from outdated technologies to new technology mediums.

5.2. Example: Data Category A

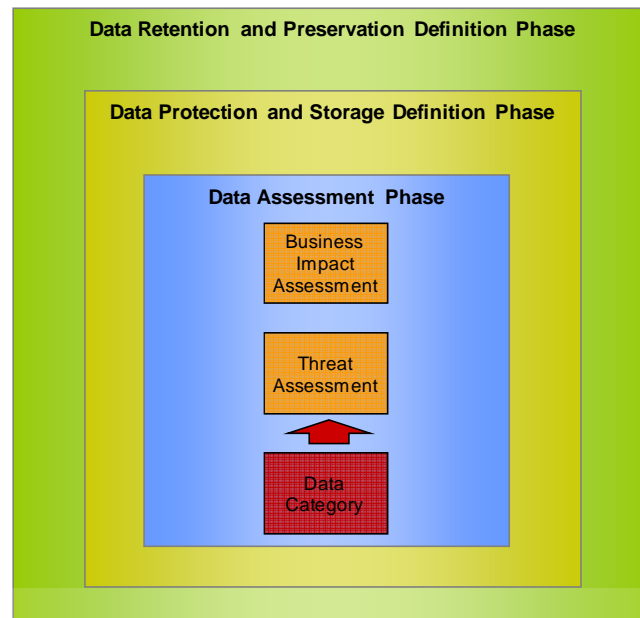


Figure 23: Data Category



Data Category A: Billing data (Telecommunications Organisation)

Threat: loss of billing data due to system downtime, cause: flood.

Impact: Decreased customer confidence (incorrect bills sent to customers with incorrect amounts), increase of complaints to Call Centre.

Storage/protection Strategy: make backups of data, replicate data to Site X.

Storage Method: Backup and replication, Information Lifecycle Management for management of old customer data.

Storage Medium: Tape drives.

Retention Period: 18 months (due to regulation and data retention act 18).

Preservation Strategy: Identify medium to be used, threats to the medium e.g. bit rot. Strategic plan of moving data onto latest mediums of storage and preservation.

Strategy Formulation

An end to end Data Protection and Preservation Strategy can then be created from the information obtained for each Data Category.

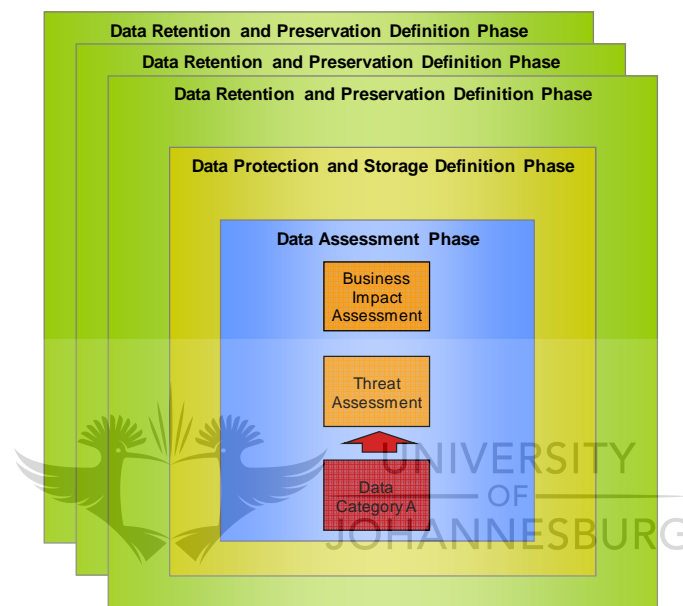


Figure 24: Data Protection and Preservation Strategy Formulation

The outputs of the strategy formulation process would be:

Information

- Data Discovery Plan: plan stating where business data resides.
- Data Valuation Matrix: matrix of data categories.
- Information Lifecycle Management Strategy: strategy dealing with the changing lifecycle of information.

- Knowledge Management Plan: plan to retain business and company knowledge.
- Enterprise Information Architecture: architecture of information required for each business operation and how this fits end to end in the company's value chain.
- Information Recovery Plan: plan to recover information to the state it was in before an interruption occurred.
- Information Cost Analysis Report: report for the CFO (Chief Financial Officer) indicating the costs associated with information loss and the cost of recreation.

Storage Media

- Data Preservation Lifecycle Management Strategy: strategy to manage the changing lifecycle of storage media on which data is stored and preserved.
- Data Preservation Plan: plan of changing storage media against new technologies and media and trends.
- Data Storage Strategy: strategy for data storage practices and cost effective solutions adapting to the changing lifecycle of storage media.
- Enterprise Storage Architecture: architecture of end to end storage solution .
- Storage Cost Analysis Report: report for the CFO (Chief Financial Officer) indicating the costs associated with storage technologies, backup medium.

Disaster

- Threat Assessment Report: report of the identified threats with their likelihood to occur.
- Business Impact Assessment report: report of the list of impacts.

- Disaster Recovery Plan (per company business unit): plan for preventing disasters or mitigating risks as they occur.
- Enterprise Business Continuity Plan: plan to ensure that business functions are continuous with minimal to no interruption.
- Disaster Cost Analysis Report: report for the CFO (Chief Financial Officer) indicating the costs associated if a disaster were to occur.

Preservation

- Retention period plan and strategy: this would be done against the Data Preservation Management Strategy and Information Lifecycle Management Strategy to ensure that the changing lifecycle of information is managed against the changing lifecycle of preservation or storage media.

Once each of these plans, strategies and reports are put together this can be consolidated and form a Data Protection and Data Preservation Strategy and Information Recovery Plan for each data category.



CHAPTER 6: DISCUSSIONS

6.1 Scope Limitations

Initially the population size for this research was too large, being all commercial enterprises within Southern Africa and its industries of Financial Services, Telecommunications, Energy, Government, Resources. This proved to be difficult in terms of determining which South African enterprises would be chosen to be interviewed, as well as who would be the most suitable employee to interview upfront without been directed to an alternative interviewee for specific questions that could not be answered by the initial interviewee. It was therefore imperative that the researcher look at a smaller population size with a sample being representative of a larger group. Therefore for this reason computer security consultants were selected based on their experience within the industry and across the different industry sectors, having been exposed to general best practices and challenges within organisations and having a holistic view of a thread of common trends and issues within an organisation in this regard.

The questions posed for the interviewees would be sensitive security questions such as:

- Where is data stored, where does it reside?
- Has your organisation experienced any downtime?
- What have the impacts been on your organisation?
- Do you have disaster recovery plan in place that is up to date and tested?

This could possibly lead to organisations and its employees being interviewed feeling vulnerable and exposed with such questions asked. Therefore the researcher chose computer security consultants as interviewees as they would have a general understanding and view of trends and issues in this study and would be able to give

a broad view for the research questions without being exposed as a company or supplying any client specific information. In this way it also allowed the researcher to receive a more complete set of feedback.

6.2 Research Method Weaknesses

Since the research methodology was qualitative in nature, with interviews taking place, a weakness for this has been that the findings could be slanted by opinions and not necessarily a quantified fact. This could result in the Data Protection Model that was proposed being idealistic in nature, more than being practical for organisations. The model would have to be tested within an organisation by creating a data protection strategy using the model in order to ascertain if it would be a guideline or an absolute best practice and to determine its practicality.

6.3 Limitations of proposed model

Limitations of the model could be potentially be limiting for usage for companies with a large scale of data who really feel that all their business data is critical. Although for this reason it can be argued that for these companies in particular a business case for an expensive but reliable data protection solution exists and be motivated with business. It would therefore be key in this case that the organisation performs data loss or unavailability or downtime calculations to back a decision that all data is critical and therefore required to be stored, protected and preserved using a more costly solution.

As the model is generic, it does not provide a copy/paste solution for companies therefore they would need to research best practices etc in their own capacity. It does not provide specific answers but rather a general guideline for how an organisation could go about creating a cost effective data protection strategy if that is a business requirement or a goals of their business and IT strategy.

CHAPTER 7: CONCLUSION

In conclusion, the proposed Data Protection Model serves as a generic model of potential assistance to organisations in determining the value of their data and the best way to store, protect and preserve. It is aimed at ensuring a cost-effective data protection strategy which allows for the data that is required to be available to be accessible on a day-to-day basis, and to be protected in the event of a disaster.

It can be noted in the research that companies are generally involved in business continuity practices, but data protection methods and business continuity as well as storage medium are typically expensive for the organisation as a whole. It is therefore necessary to store and protect information more effectively, as there are many more real and relevant threats to an organisation.

The Data Protection Model can be used to store, protect, preserve and recover critical business data.

7.1 Key Findings Summary



UNIVERSITY
OF
JOHANNESBURG

7.1.1 Key Findings

A number of key findings emerged from the research:

Organisations generally know what their critical data is but not understanding the true value of business data. Organisations do not know where all key data resides and the impacts of downtime are generally known but companies are not aware of full impact of information not being available. Furthermore, costs associated to data unavailability are not been known as no formula being used in organisations to calculate this.

In terms of best practices, organisations are required to perform a Threat Assessment to understand the threats that each organisation has as well as a Business Impact Assessment to understand what impact a disaster or event as such has on the organisation.

Best practice concepts are used but perhaps in isolation. There are also budget constraints regarding business continuity/disaster recovery in the organisation as it is not viewed as a critical business function. This could be due to the organisation not understanding the costs associated to data unavailability not been known as no formula being used in organisations to calculate this. Criticality of certain aspects is not taken seriously.

Business continuity plan is typically not tested and therefore fail proof. Concepts such as Information Lifecycle Management and Data Valuation are not implemented by organisations.

Data preservation strategies and methods are not of focus in organisations with digital preservation not being advanced in general. There is currently no Data Preservation Lifecycle or an understanding of this lifecycle.



7.1.2 Implications of findings

The broader implications of the findings suggest that a Data Protection Model is required to store, protect, retrieve and preserve critical business data for organisations that have distinct and different data categories. The findings imply that organisations deploy storage and protection methods that are perhaps technology or vendor drive and not necessarily a solution that is flexible to protect and store data appropriately.

Furthermore, the findings indicate that digital data preservation and its strategies are not delved into or even perceived as an issue within an organisation. This could pose potential issues in the future should data be required to be retrieved and the storage/preservation medium is obsolete or has gradually degraded. However findings suggest that a data preservation strategy is required to deal with technology changes for storage medium and challenges for each.

7.2 Proposed Model Summary

7.2.1 *Practicality of proposed model*

The model could be practical for organisations with issues with large-scale data with varying or very particular business value. The purpose of the model is to be used as a framework or guideline for an organisation seeking for a cost effective solution of protecting and storing its data. It provides a generic and holistic view to taking an approach to create a cost-effective data protection strategy and therefore solution.

The model goes one step further by indicating that preservation of data is key and very relevant and this is required to be addressed for adding the ability to retrieve data as required (even if its value has changed in the organisation) but is now put aside for retention purposes on a storage medium for a length of time. This preservation highlights the needs and challenges for different medium and the changing technology and the lifespan of the medium such as CDs.

7.3 Future Research



Since the model is generic and of a high and holistic level, a future research project could be used to further determine lower levels of the model and so provide solutions for organisations that may not have the expertise, resources or capacity to research and create a data protection strategy based on a generic model only.

REFERENCES

BACKUP 2004: The Backup Book: Disaster Recovery from Desktop to Data Center (Accessed 10/02/2004)

BAKER & KEETON 2005: Why Traditional Storage Systems Don't Help Us Save Stuff Forever (Accessed 07/08/2007)

<http://digitalpreservationstrategies.blogspot.com/2006/08/threats-to-digital-preservation.html>

BANNED BOOKS 2003 (Accessed 12/02/2003)

www.ancienthistory.about.com/library/weekly/aa091598.htm?terms=Alexandria+Library

BOGOSSIAN, M 1998: Sorting through data in the field of data-storage formats (Accessed 31 05/2004)

<http://albany.bizjournals.com/albany/stories/1998/08/10/smallb4.html>

BONNETTE, C 2003: Assessing Threats to Information Security in Financial Institutions (Accessed 28/09/2005)

http://www.sans.org/reading_room/whitepapers/threats/1143.php

BURNIE, M 2002: "Protect your data: Network Appliance outlines better ways of protecting and recovering your precious electronic data." Journal of Banking and Financial Services. Academic OneFile. Thomson Gale. University of Johannesburg (Accessed 13/08/2007)

CANE, D 2002: Preventing the Great Data Loss Disaster: Data Protection Shouldn't Stop at the Server (Accessed 9/02/2004)

<http://www.technologyreports.net/securefrontiers/?articleID=1007>

CHIN, K 2007: "Use a Digital Preservation Plan to Manage Content for the Long Term" Gartner Research Report (Accessed 26/02/2009)

COOPER, D.R. and EMORY, C.W. 1995: Business Research Methods, 5th Edition, Richard D. Irwin, Homewood, IL.

COUNCIL OF EUROPE 2009: Data Preservation Checklists (Accessed 10/01/2009)
http://www.coe.int/T/DG1/LegalCooperation/Economiccrime/cybercrime/Documents/Points%20of%20Contact/24%208%20DataPreservationChecklists_en.pdf

CROY, M 2006: The Business Value of Data, Forsythe Solutions Group, Inc., Skokie, IL (Accessed 18/12/2006)
<http://www.forsythe.com/na/aboutus/news/articles/2004articles/thebusinessvalueofdata>

DIGITAL PRESERVATION STRATEGIES 2006: Wednesday, August 02, 2006 (Accessed 22/12/2008)
<http://digitalpreservationstrategies.blogspot.com/>

FOLGEMAN, R. R 2003: Cornerstones of Information Warfare (Accessed 12/02/2003) www.af.mil/lib/corner.html

HENRICHSEN, L 1997: Taming the Research Beast, Research Methods: Panning (Accessed 16/01/2009)
http://linguistics.byu.edu/faculty/henrichsen/researchmethods/RM_0_01.html

ILM, 2004: Information Lifecycle Management Manual (Accessed 31/06/2004)

KAHLE, R 2003: Spreading the Digital Word (Accessed 9/02/2004)
http://www.extremetech.com/print_article/0,3998,a=41089,00.asp

KAOMEA, P 2003: Beyond Security: A Data Quality Perspective on Defensive Information Warfare (Accessed 9/02/2004)
<http://web.mit.edu/tdqm/papers/other/kaomea.html>

KRAYTON, M 2003: Chapter 2: Cost of Justifying your Backup Plan (Accessed 10/02/2004)

LEEDY, P.D. and ORMOROD, J. E. 2001: Practical Research: Planning and Design, Seventh Edition, Upper Saddle River: Merrill Prentice Hall.

LEWIS, J. A. (2006): Where Data Resides - Data Discovery from the Inside Out Digital Mountain

<http://digitalmountain.com/fullaccess/Article3.pdf>

LINCOLN, Y.S. and GUBA, E.G. 2000: Paradigmatic controversies, contradictions, and emerging confluences. In Denzin, N.K & Lincoln, Y.S (Eds.). Handbook of qualitative research. 2nd edition. Thousand Oaks. California: Sage Publications

MASSIGLIA, P 2003: The Resilient Enterprise VERITAS Software Corporation, California

MOORE, F 2003: Storage, New Game New Rules, Storage Technology Corporation, Louisville



NSS 2003: Network Support Services Training Manual (Accessed 01/03/2005)

Peers, E 1985: "Data Protection Act." Accountancy. Academic OneFile. Thomson Gale. University of Johannesburg (Accessed 13/08/2007)

Poker, A.M. 1996: "Computer system failure: planning disaster recovery." Nursing Management 27.n7 (July 1996): 38(2). Academic OneFile. Thomson Gale. University of Johannesburg (Accessed 13/08/2007)

Protecting data from disaster 1995: Academic OneFile. Thomson Gale. University of Johannesburg (Accessed 13/08/2007)

SEARCHSTORAGE.COM 2009: SearchStorage.Com Definitions (Accessed 10/01/2009)

http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci211633,00.html

http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci551320,00.html

http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci963635,00.html

http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci801381

SEINER, R 2001: Business Impact of Knowledge Management (Accessed 07/08/2007)

[http://www.tdan.com/view-articles/4943/:](http://www.tdan.com/view-articles/4943/)

SYNERGISTIC ONLINE SOLUTIONS (2009): Business Continuity Issues to Consider for AS/400 Enterprises (Accessed 12/01/2009)

http://www.synergisticonline.com/as400_business_continuity.html

TSAMAIDIS, H 2003: Great Library (Accessed 12/02/2003)

<http://www.straightdope.com/columns/read/2233/what-happened-to-the-great-library-of-alexandria>

UCERTIFY 2009: The fastest way to IT Certification. Planning a backup and restoration of files for disaster recovery (Accessed 20/01/2009)

<http://www.ucertify.com/article/planning-a-backup-and-restoration-of-files-for-disaster-recovery.html>

WRENN, G 2005: Data Center Management Advisory Newsletter: Ten steps to a successful business impact analysis (Accessed 07/08/2007)

http://searchdatacenter.techtarget.com/tip/0,289483,sid80_gci1094071,00.html

YIN, R. 2003: Case Study Research, Design and Methods, Third Edition, Sage Publications.

APPENDIX A

Interview Questions

Name:

Expertise:

Length in industry:

Company:

Clients consulted to:

When you consult to a company:



Business value of data

1. Does your client understand what their critical information is?
2. Does your client understand the different types of data it has and what value it holds?
3. Does your client define/categorise their data (e.g. Data Valuation)?
4. How does your client deal with the changing value of data?
5. Do you as a consultant think it would be important to determine an organisation's value of data?
6. If yes, why?

Threats

7. Do different types of business data have different threats? (E.g. confidential data taken by competitors? Credit card data taken by hacker?)
8. Do different companies have different threats?

9. How does your client identify threats to organisation?
10. Which of the following has been identified as a threat:
- a. Natural disasters (floods etc)
 - b. Man-made disasters (war, terrorism)
 - c. Computer system failure (hardware, software)
 - d. External computer threats (viruses or hackers)
 - e. Internal com threats (accidental or malicious behaviour)
 - f. Change control issues (patches)
 - g. Other (Please specify)
11. As a consultant would you advise your clients to perform a Threat Assessment?
12. If yes, why?

Impacts

13. What are the possible impacts to the business if there is a data loss?
14. What does the impact depend on?
- a. Type of data
 - b. How much value it holds
 - c. How confidential it is
 - d. Financial aspects
 - e. Other (Please specify)
15. Are there different impacts correlating to the type of data?
16. Is it important for a company to perform a Business Impact Assessment?
17. If yes, why?

Alignment

18. Is there a correlation between the value of data, the threats towards that data and the impact on the organisation if that data is destroyed/lost?

19. Would it make sense to align the data category, its associated threats and impacts in order to protect accordingly?

20. If yes, why?

21. Would it make sense to have different strategies and methods depending on the value of data, its threats and impacts to the organisation to protect that specific data?

Methods to protect/store/preserve data

Data Protection

22. Does your client have protection methods?

23. Which of the following does your client have in place for data protection?

- a. Strategy
- b. Policy
- c. Procedures
- d. Processes
- e. Other (Please specify)

Data Storage/Preservation



24. Does your client store and recover data according to its value? (E.g. Mission-critical data stored differently to non-critical data)?

25. Does your client have data storage methods/preservation?

26. Which of the following does your client have in place for data storage/preservation?

- a. Strategy
- b. Policy
- c. Procedures
- d. Processes
- e. Other (Please specify)

27. Which of the following data protection/ data storage/preservation methods does your client use?

- a. BCP
- b. DR
- c. ILM
- d. Data valuation framework
- e. Risk Assessment
- f. Threat Assessment
- g. Business Impact Analysis
- h. Storage architectures
- i. Other (Please specify)

28. From the different protection methods and strategies, how does your client choose which one to deploy?

29. Would it make sense to protect and preserve data according to its value?

30. If yes, why?



Data Retention

31. Does your client have data retention policies to adhere to?

32. Does your client keep different types of data for different periods of time?

33. What determines the period?

- a. Type of data?
- b. Company data retention policies?
- c. Regulatory/Legal data retention policies?
- d. Other (Please specify)

34. Would periods differ for different categories of data (e.g. mission-critical =5 years)?