DOREEN KERUBO MAGETO

COST FACTORS IN DIGITAL PRESERVATION

ABSTRACT

This study examines the cost factors using the Open Archival Information System (OAIS) in preserving digital records in organisations. The aim of the study is to use the OAIS functional model in identifying the cost factors that organisations need to be aware of in preserving digital records. The OAIS model breaks down the preservation process into smaller sub-processes which could then be assessed more easily with regard to their required resources and the associated costs factors. The cost factors will hopefully help organisations to budget better for the preservation of digital records.

Chapter one introduces various aspects of the study which include; background information to the study; statement of the problem; the purpose of the study; the aim and objectives, the scope of the study; the intended audience and finally the significance of the study.

Chapter two reviews literature relevant to the study. The focus of the literature was on the six entities of the OAIS functional model which include Ingest, Archival Storage, Administration, Preservation Planning, Data Management and Access.

Chapter three discusses the methodology adopted by the study. It specifically examines the research purpose, research approach, research strategy, data collection methods, population and finally the limitations that were encountered in the research. The study adopted a case study. Interviews, online questionnaires and documentation were used to collect data and the study population was the representative groups of professionals in Norway which include the EDOK (part of the Norwegian computer society), Verdiskaperne (information management network) and the LongRec project in Norway.

Chapter four presents the data collected and its analysis in relation to the study aims and objectives. Among the findings that the study came up with were: most organisations did not have a cost model for digital preservation; few organisations were using the OAIS model as framework in determining costs in digital preservation and finally there were different ratings on the cost factors decomposed from the OAIS model.

Chapter five presents a matrix for the cost factors in digital preservation and a cost model that provides a tentative formula for calculating the total preservation costs. The formula provides an outline which applies to different organisations based on their stability, nature and their needs.

Finally, chapter six presents a summary of the study findings, conclusions and recommendations for further research. The major conclusion from the study is that automation is a significant factor in controlling the costs of large-scale digital preservation.

ACKNOWLEDGEMENT

I am indebted to many people for their help in the last two years as I worked towards obtaining this masters degree. First and foremost I would like to thank my supervisors Thomas Sødring of Oslo University College and Inger-Mette Gustaven of Det Norsk Veritas (DNV) for providing me with the opportunity, advice and insight to write my thesis on this topic. Your guidance, valuable comments and support encouraged me to always remain focussed throughout all stages of this thesis.

I would like to thank Ragnar Nordlie of Oslo University College, for your advice and support in deciding the relevant methods of data collection for the thesis and your comments on my methodology section of this thesis.

I would also like to thank Thomas Mestl of DNV for your valuable insight during the last stages of my thesis. Thanks for taking your time to edit my work and your wise advice that greatly contributed to the quality of this thesis.

I am very grateful to DNV, for hosting me during my internship. This is where the idea of this thesis was born as a result of working with the LongRec project. Special thanks goes to Inger-Mette the initial contact with DNV, it was a pleasure working with you. My sincere thanks go to all the LongRec project case partners and especially the partners who willingly responded to my interviews and questionnaires amidst their busy schedules.

My special thanks go to all the teachers of DILL (Digital Library Learning) in Norway, Estonia and Italy and all the students for the support, encouragement and knowledge they have offered and shared in the last two years. I thank the Erasmus Mundus Programme that sponsored my studies which enabled me to achieve my dream of doing my masters.

People who made my stay in Norway are not to be forgotten. Many thanks to my friends: Eric Kahimano, Abigail Mutindi, Rahat Isaeva, Joseph Anin and Jemima Anin. Thank you all

Last but not least I would like to thank my parents who sowed in me the first seed of formal education. Thanks for the sacrifice that you made to educate me. I am truly and deeply indebted to both of you. Thanks to my loving husband for your support and love. You are my best friend.

Finally, my sincere gratitude goes to our almighty God who through His grace has seen me through the two years of study in three different countries.

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

ERP	ELECTRONIC RECORDS MANAGEMENT
CRM	CUSTOMER RELATIONSHIP MANAGEMENT
NAS	NATIONAL ARCHIVE OF SCOTLAND
IDC	INTERNATIONAL DATA CORPORATION
OAIS	OPEN ARCHIVAL INFORMATION SYSTEM
ERPA	ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK
CCSDS	CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS
ALA	AMERICAN LIBRARY ASSOCIATION
AIP	ARCHIVAL INFORMATION PACKAGE
SIP	SUBMISSION INFORMATION
DIP	DISSEMINATION INFORMATION PACKAGE

TERMINOLOGY

The following terms have been selected from the ISO reference model of Open archival information system because of their relevance to this study.

Access: The OAIS entity that contains the services and functions which make the archival Information holdings and related services visible to Consumers.

Administration: The OAIS entity that contains the services and functions needed to control the operation of the other OAIS functional entities on a day-to-day basis.

Archival Information Package (AIP): An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

Archival Storage: The OAIS entity that contains the services and functions used for the storage and retrieval of Archival Information Packages.

Archive: An organization that intends to preserve information for access and use by a designated Community.

Consumer: The role played by those persons or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISs, as well as internal OAIS persons or systems.

Data Management: The OAIS entity that contains the services and functions for populating, maintaining, and accessing a wide variety of information. Some examples of this information are catalogs and inventories on what may be retrieved from Archival Storage, processing algorithms that may be run on retrieved data, Consumer access statistics, Consumer billing, Event Based Orders, security controls, and OAIS schedules, policies, and procedures.

Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.

Digital Object: An object composed of a set of bit sequences.

Dissemination Information Package (DIP): The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS.

Finding Aid: A type of Access Aid that allows a user to search for and identify Archival Information Packages of interest.

Ingest: The OAIS entity that contains the services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage, and ensures that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.

Long Term: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future.

Long Term Preservation: The act of maintaining information, in a correct and Independently Understandable form, over the Long Term.

Metadata: Data about other data.

Open Archival Information System (OAIS): An archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities, as defined in 3.1, that allows an OAIS archive to be distinguished from other uses of the term 'archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.

Producer: The role played by those persons or client systems, who provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems.

Reference Model: A framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment. A reference model is based on a small number of unifying concepts and may be used as a basis for education and explaining standards to a non-specialist.

Submission Agreement: The agreement reached between an OAIS and the Producer that specifies a data model for the Data Submission Session. This data model identifies format/contents and the logical constructs used by the Producer and how they are represented on each media delivery or in a telecommunication session.

Submission Information Package (SIP): An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.

TABLE OF CONTENTS

1.0. INTRODUCTION

This chapter discusses the background information to the study, statement of the problem, the research purpose, the aims and objectives, the scope of the research, the intended audience and significance of the research.

1.1. BACKGROUND INFORMATION

According to (Gantz et al., 2008) in their survey of an updated forecast of world wide information growth through 2011, they found that the digital universe in the year 2007 was 2.25 x 1021 bits (281 Exabyte or 281 billion gigabytes) and they estimated that the compound annual growth rate between 2007 and the year 2011 is expected to be almost 60%. Figure 1 below illustrates the growth over time of digital information created, captured or replicated. (Gantz et al., 2008) attribute the growth to the fast-growing corners of the digital universe which includes those related to digital TV, surveillance cameras, Internet access in emerging countries, sensor-based applications, datacenters supporting "cloud computing," and social networks. They further state that the diversity of the digital universe can be seen in the variability of file sizes, from 6 gigabyte movies on DVD to 128-bit signals from RFID tags. Because of the growth of VoIP, sensors, and RFID, the number of electronic information "containers" files, images, packets, and tag contents is growing 50% faster than the number of gigabytes.

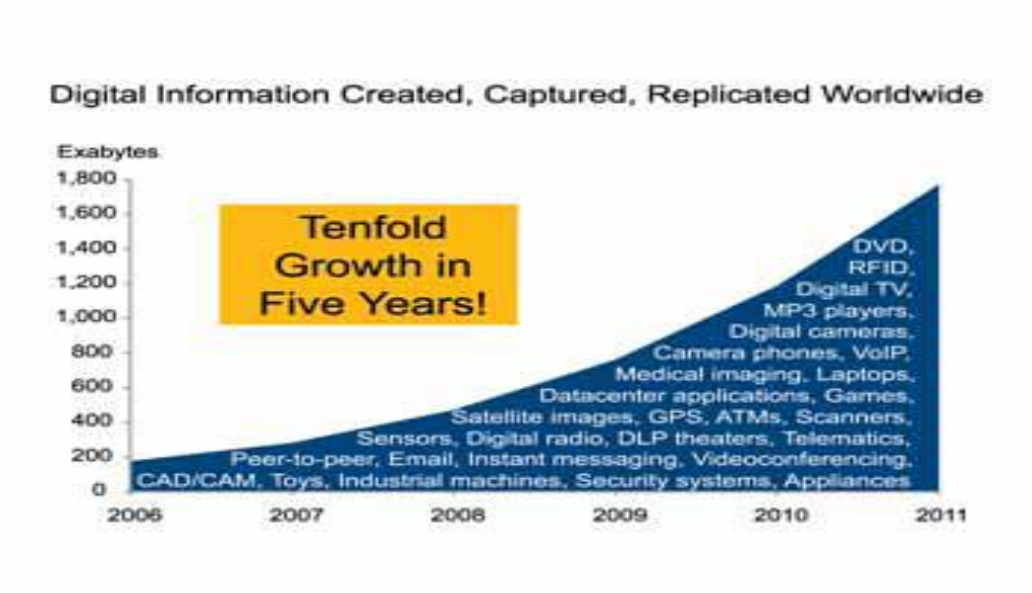


Figure 1: Forecasted growth in digital information being created captured and replicated worldwide from 2006 - 2011. (Gantz, et al., 2008)

However, (Lougee, 2002) attributes the explosion of information to distributed computing and the web which has democratized technology by bringing it to the desktop and as a result many individuals and institutions have the basic capability of publishing and generating information.

Organisations are facing major challenges due to the exponential growth of the amount of information generated every day. (IBM System Storage, 2008) points out that Information that surrounds us i.e. emails, instant and texting messaging, cellular communication, the internet and digital video has become part of our daily lives. These channels have become indispensable tools of business communications in virtually every organisation or industry. Business critical ERP (Electronic Records Management Programmes), CRM (Customer Relationship Management) and other transaction processing applications also collect and accumulate a wealth of business information. These applications empower new business initiatives, support daily operations and drive revenue generation. As businesses generate more and more information, the information need to be managed effectively and efficiently in order to meet business objectives, reduce risk and drive innovation.

Information is the most basic and essential asset in most modern organisations and like with any other business asset, recorded information requires effective management (National Archives of Scotland, 2005). If we believe that information assets are in need of management and preservation, then the value of the information needs to be clearly expressed to the decision makers. They should clearly understand the value or benefit that derives from it and the risks inherent in making or not making an investment. (Currall, Johnson & McKinney, 2006).

The greatest challenge at the moment as pointed out by (Gantz et al., 2008) is the issue of information overload. In their research they found that in 2007, according to their estimates, all the empty or usable space on hard drives, tapes, CDs, DVDs, and memory (volatile and non-volatile) in the market equalled 264 Exabyte which is very close to the total amount of information created or captured (see Figure 2 below), and from here on, the two numbers diverge. This mismatch between creation and storage, plus increasing regulatory requirements for information retention, will put pressure on those responsible for developing strategies for storing, retaining, and purging information on a regular basis.

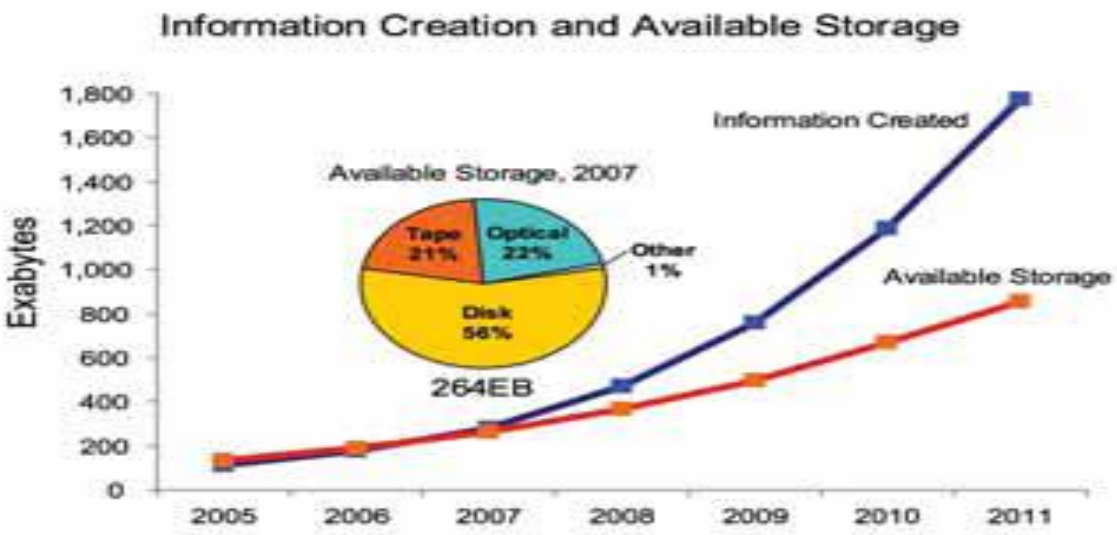


Figure 2: Available storage versus created information forecast from 2005 to 2011. (Gantz, et al., 2008).

(IBM System Storage, 2008) point out that the surge in criticality, value and volume of information is overwhelming, yet the value of information is not equal across applications and it changes over time. Identifying less frequently used data (inactive data) and then distinguishing between truly historical data and data that must remain immediately available (active data) is important for businesses. Managing inactive data appropriately will have bottom-line impact on productivity and operational costs.

In an environment of increasing regulatory measures and the need to comply with record retention requirements, business information must be appropriately stored, protected and disposed off when no longer required for compliance or business purposes. Failure to produce documents under litigation or regulatory audits in a timely manner can result in serious consequences, including hefty fines, even when information delay, loss or damage is accidental. As the information volume increases and required storage times lengthen, managing, storing and protecting business information becomes more challenging and costly (IBM System Storage, 2008).

The systematic management of digital records as pointed out by (NAS, 2005) allows organisations to:

- know what records they have, and locate them easily
- increase efficiency and effectiveness
- make savings in administration costs, both in staff time and storage

- support decision making
- be accountable
- achieve business objectives and targets
- provide continuity in the event of a disaster
- meet legislative and regulatory requirements, particularly as laid down by the Freedom of Information (Scotland) Act and the Data Protection Act and
- protect the interests of employees, clients and stakeholders.

Despite the knowledge on the importance and value of digital records by organisations, one thing that has been an issue of concern in organisations is the question of how much will it cost to preserve the digital records to ensure that they can be accessed as long as they are needed.

1.2. STATEMENT OF THE PROBLEM

The availability of cost data from ongoing digital preservation activities is scarce. This can be ascribed to the fact that the notion of preserving digital materials is relatively new. Consequently, most digital preservation initiatives are still in their infancy (Lavoie, 2003).The costs of preserving digital materials will be different from the cost of preserving other non-digital materials and will require resource commitments of a different nature on an ongoing basis (Cedar, 2002). (Russell &Weinberger, 2000) indicate that the ongoing costs of digital preservation are also likely to span a more extended timeframe than traditional preservation and it may be the case that different technical strategies will prescribe quite different costing timeframes and schedules.

Several attempts have been made by researchers in coming up with cost models for digital preservation. A good example is the Hendley's cost model which is based on the seven stages associated with digital collection management. He expands this framework by enumerating a set of cost elements for each module, in order to describe the total costs associated with the long-term management of a digital collection. With this in hand, he then identifies the subset of costs directly or indirectly related to preservation. He then applies this general cost model to four categories of digital resources (Hendley, 1998). Other attempts to identify cost elements associated with digital preservation include (Russell & Weinberger, 2000), who organize cost elements according to the chronology of a collection manager's workflow.

However, a number of factors tend to make it difficult to analyse the costs of digital preservation. These factors have been referred to as 'uncertainties' by (Lavoie, 2003). He states that uncertainty arises from the very reason for digital preservation i.e. the constantly evolving technological environment in which digital materials exist. This creates an ever-present risk that archived digital objects will be "orphaned" when the hardware and/or software needed to store, render, or utilize them becomes obsolete or unavailable. The time

horizon within which technological obsolescence becomes a concern can be uncomfortably short, or even surprisingly long. It is difficult to anticipate, a priori, the future demise of current hardware and software environments, and therefore, difficult to forecast the timing, frequency, complexity, and ultimately, the cost of intervention to stave off obsolescence.

A second source of uncertainty that Lavoie states as associated with estimating the costs of digital preservation is that these costs are a function of many variables, including, but not limited to, the period of archival retention (ten years? one hundred years? "Perpetuity"?); storage technologies, including hardware and digital media; the level of access; the objectives of preservation; preservation strategies (migration? emulation?); type and variety of digital formats; richness of metadata description; and the relative mix of labour-intensive and automated processes. Accurate cost estimates must be predicated on a reasonably specific description of the preservation processes involved. This in turn suggests that there is little scope to generalize these estimates over heterogeneous digital preservation activities (Lavoie, 2003).

Despite all the above uncertainties, there has been progress in addressing some of the uncertainties facing digital preservation. The development of the OAIS reference model, which is currently an ISO standard has addressed a full range of archival information preservation functions including ingest, archival storage, data management, access and dissemination. The OAIS models provides a framework to be used in developing a broader consensus on what is required for an archive to provide permanent or indefinite long term, preservation of digital information (CCSDS, 2002).

This research will seek to use the Open Archival Information System (OAIS) as a frame work which can be used in determining the cost factors by decomposing the functional model and identifying possible cost factors that organisations should be aware of in digital preservation.

1.3. THE PURPOSE OF THE STUDY

The statement of the problem strives to show the importance of cost in digital preservation. (ERPANET, 2003) states that cost is one of the main criteria for people responsible for digital preservation when it comes to funding sustainable preservation infrastructure and related activities and hence there is a need for understanding the scope. The purpose of this research is to determine the cost factors using the OAIS reference model as a framework. This shall be investigated through the following research question;

What are the cost factors that can be identified from decomposing the functional model of the OAIS?

1.4. AIM AND OBJECTIVES

The aim of this research is to use the OAIS functional model in identifying the cost factors that organisations need to be aware of in preserving digital records. The cost factors will enable organisations to budget for the preservation of digital records.

OBJECTIVES

1. To establish how organisations identify cost factors for digital preservation.

2. To find out how organisations allocate resources towards the long term management of digital records.

3. To establish how the OAIS functional model can be used when determining cost factors in digital preservation

1.5. SCOPE OF THE RESEARCH.

The OAIS reference has got a functional model and an information model for long term preservation. In addition to the two models, they briefly introduce some preservation strategies and interoperability among OAIS archives. However, this research is concerned with determining cost factors in preserving digital records using the functional model of the OAIS.

The scope of this thesis is restricted to the representative groups of professionals in Norway which include the EDOK (part of the Norwegian computer society), Verdiskaperne (information management network) and the LongRec project in Norway. Most of the members of the EDOK and Verdiskaperne are also members of the LongRec project. The above groups were used to acquire information on cost factors in digital preservation due to time limit and lack of enough resources for multiple case studies.

According to the (Testbed Digitale Bewaring 2005), it is often difficult to specify the demarcations between the actual use of the records, their local storage, local preservation, archiving, and long term preservation. The concept of the 'records continuum', which is ideally suited to use in this context, can be defined as'; 'a consistent and coherent regime of management processes from the time of the creation of records, through to the preservation and use of records as archives'.

The focus of the research was based on the overall management of the digital objects to ensure that they can be accessed as long as they are required. In other words, the aim of identifying the cost factors is to ensure that the digital objects created or received in an organisation can be managed in such a manner that can ensure long term survivability and that their security can be guaranteed.

1.6. THE INTENDED AUDIENCE

The intended audience for the study therefore encompasses all individuals and organisations who have a role in the creation and preservation of digital resources, from the funding agencies, researchers, through to the organisations which may assume responsibility for their long-term preservation and use. Although it is not a technical report it will assume a basic understanding of the nature of digital objects and an overall awareness of the issues relating to their long-term preservation.

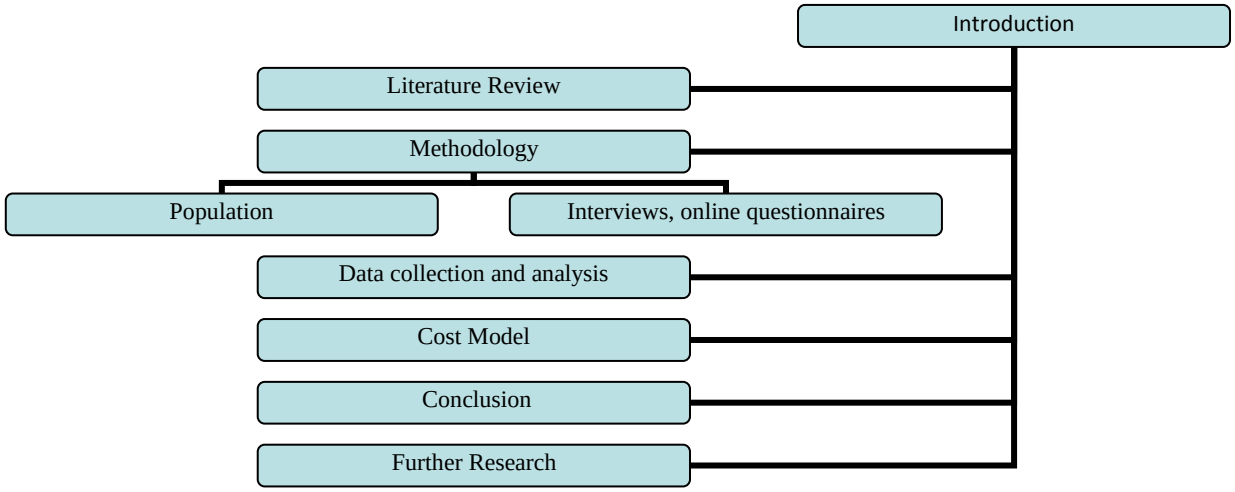
1.7. SIGNIFICANCE OF THE RESEARCH

- The outcome of the thesis shall inform organisations what cost factors they need to be aware of in long term sustainability of digital information. This gives them an opportunity to budget for the preservation of digital information.
- This thesis shall contribute to already existing research on the use of OAIS model as a framework for determining cost factors in preserving digital records.
- It will act as a source of reference for those who want to do further research on the same area.

1.8. DISPOSITION OF THE THESIS

This research shall consist of six chapters as shown in the figure below. Chapter one contains an introduction to the research area, statement of the problem, the research purpose, the aims and objectives, the scope of the research, the intended audience and significance of the research. The chapter two presents the literature review followed by the methodology used in this thesis in chapter three. The fourth chapter shall focus on the presentation and analysis of data gathered from interviews and questionnaires. Chapter five focuses on the cost model. Finally, chapter six presents a summary of study findings, conclusions and recommendations for further research.

THESIS STRUCTURE



2.0. LITERATURE REVIEW

2.1. INTRODUCTION

This chapter discusses the relevant theories needed to find answers to the research questions. The review focuses on literature addressing cost factors hidden in the Open Archival Information System (OAIS) functional model entities as a framework. The literature review will shed light on the various cost factors in the six functional entities of the OAIS model.

2.2. BACKGROUND ON COST FACTORS IN DIGITAL PRESERVATION

(Woodyard-Robinson, 2006) argues that calculating the cost of digital preservation is a complex task, but perhaps even more challenging is assessing the value of this work and securing the funding to perform it. He points out that key decision makers must be convinced that the value of the digital assets is equal to or greater than the cost of the services to maintain them in order to establish economically sustainable processes and business models. Few organisations are willing to pay for a preservation initiative without knowing how much it costs and how the costs are distributed. Hence, costs are a necessity, if not sufficient, component of a viable sustainability plan (Lorraine, Amy &Roger, 2008).

(Lavoie, 2004) describes economic sustainability as the ability to marshal sufficient resources, on an ongoing basis, to meet preservation objectives. He points out that there are many avenues by which sustainability can be achieved. The first one is an institutional commitment to budget a continuous supply of funds to support digital preservation. These funds might be used to extend a pilot project originally funded through seed money from a grant-giving organization. Another avenue he points out is that digital preservation activities might also be self-sustaining, generating revenues as a by-product of day-to-day operations. In these circumstances, economic sustainability might be defined in terms of cost recovery, or a minimum level of profitability.

(Lavoie, 2004) further points out that, strategies for attaining economic sustainability must be built on a sound empirical footing. Consequently, much more data on the costs of digital preservation is needed. Digital preservation is still in its infancy, and much of the available data is heavily skewed toward upfront costs; reformatting, setting up the digital repository, ingestion of materials, etc. As projects mature, empirical descriptions of digital preservation's complete cost trajectory will emerge. This data must be consolidated and synthesized to produce reasonable benchmark estimates of the cost requirements associated with various forms of digital preservation (Lavoie, 2004).

Using an established model as a basis for determining preservation costs can be helpful but one has to be aware of significant differences in collections and material types, organisation mission and the services they provide as all these aspects of an organisation can have significant effects on their costs. This was further emphasized by (Ashley, 2000) who pointed out in "*Digital Archive Costs: Facts and Fallacies"* that the primary factor influencing the magnitude of preservation costs is not the quantity of objects preserved. Rather, it is the range of services offered by the archival provider.

(Woodyard-Robinson, 2006) suggests that, a standard approach to determining costs is to break down the digital life cycle into processes based on workflow or a system model such as the OAIS reference model. Each stage or process, called a cost event, is then evaluated for likely cost sources. He points out that it may be helpful to use the OAIS reference model as a guide to enhance and inform the future of the long term digital life cycle. The OAIS discusses many processes that will be needed for long term preservation that may not yet be fully implemented within an organisation.

2.3. DEFINITIONS OF TERMS

Digital preservation

According to (Jones & Beagrie, 2002) digital preservation has been defined as:

> *All the activities employed to ensure continued access to digital resources which have retained properties of authenticity, integrity and functionality. The term "archiving" can be substituted for preservation provided this definition remains.*

According to (Cedar, 2002):

> *Digital Preservation is defined as: the managed activities necessary for ensuring the long-term maintenance and continued accessibility of digital materials and it involves the long-term maintenance of a byte stream and continued accessibility to its contents*

According to the American Library Association (ALA) (2007)

> *Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time."*

Throughout the thesis the term digital preservation will be used according to the following definition:

The series of actions, processes and / or interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed of value. This encompasses not just technical activities but also all of the strategic and organisational considerations that relate to the survival and management of digital materials.

Cost

According to (Ruiz, 2003),

Cost is the economic effort (the payment of salaries, the purchase of materials, the manufacture of a product, securing funds for financing the management of the company, etc.) that must be done to achieve an operational objective.

According to (Investor Dictionary, 2009),

Cost is something of value, usually an amount of money, given up in exchange for something else, usually goods or services. All expenses are costs, but not all costs are expenses. (An expense is the cost of resources used to produce revenue.) As a verb, cost means to estimate the amount of money needed to produce a product or perform a service.

According to the (Business Dictionary, 2009),

Cost is defined as, valuation in terms of money of (1) effort, (2) material, (3) resources, (4) time and utilities consumed, (5) risks incurred, and (6) opportunity forgone in production and delivery of a good or service. All expenses are costs, but not all costs (such as those incurred in acquisition of an income-generating asset) are expenses.

We shall use the term cost to refer to the definition given by the business dictionary since it is more comprehensive and covers all aspects that can be considered as cost factors in any setup.

2.4. THE OAIS MODEL

Beedham et al. (2004) discuss the history of the OAIS reference model. They state that it was first developed by the Consultative Committee for Space Data Systems (CCSDS).The CCSDS[1] was established in 1982 to provide an international forum for space agencies interested in the collaborative development of standards for data handling in support of space research. In 1990 the CCSDS entered into a co-operative agreement

[1] http://www.jisc.ac.uk/uploaded_documents/oaismets.pdf

with Subcommittee 13 (Space data and information transfer systems) of the Technical Committee 20 (Aircraft and space vehicles) of the ISO. At the request of the ISO, the CCSDS assumed the task of co-ordinating the development of archive standards for the long-term storage of archival data in 1995. Although the CCSDS was initially to address the problems of archiving data obtained from observations of the terrestrial and space environments and used in conjunction with space missions, it soon took an intentionally interdisciplinary view and ensured broad participation in the discussion of a reference model for the long term storage requirements of this digital information. The very first draft of the digital archive model was released after a year of work. The draft was then discussed by international and national working groups and at workshops, resulting in the publication of the first version of the OAIS model in 1999 and its update in 2001.

They further state that, development of the reference model began with the premise that one of the greatest challenges in accepting preservation responsibility within an organisation is finding a shared vocabulary for stakeholders with a variety of backgrounds to use for productive discussion of the issues. Thus, the model was first developed to establish common terms and concepts, to provide a framework for elucidating the significant entities and relationships among entities in an archive environment, and to serve as the foundation for the development of standards supporting the archive environment.

A broader task for the OAIS development has been defined as articulating the functionality and components of any system responsible for preserving any type of information over any length of time. The terminology used to describe the OAIS are often not the traditional archival or recordkeeping terminology since it is intended as a common language within which a diversity of communities can continue to implement and develop the OAIS model. The model has been very successful in one of its main goals to spur further interest and discussion of digital preservation and archiving issues and standards. The 2002 CCSDS version of the OAIS reference model was proposed and was accepted as an international standard in 2003: ISO 14721:2003 *Open archival information system Reference model.* (Beedham et al, 2004).

THE PURPOSE OF THE OAIS MODEL

The OAIS reference model[2] addresses a full range of archival information preservation functions including ingest archival storage, data management, access, and dissemination. It also addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives. It identifies

[2] http://public.ccsds.org/publications/archive/650x0b1.pdf

both internal and external interfaces to the archive functions, and it identifies a number of high-level services at these interfaces. It provides various illustrative examples and some "best practice" recommendations. It defines a minimal set of responsibilities for an archive to be called an OAIS, and it also defines a maximal archive to provide a broad set of useful terms and concepts (CCSDS, 2002).

This reference model (CCSDS, 2002):

- *Provides a framework for the understanding and increased awareness of archival concepts needed for long term digital information preservation and access;*
- *Provides the concepts needed by non-archival organizations to be effective participants in the preservation process;*
- *Provides a framework, including terminology and concepts, for describing and comparing architectures and operations of existing and future archives;*
- *Provides a framework for describing and comparing different long term preservation strategies and techniques;*
- *Provides a basis for comparing the data models of digital information preserved by archives and for discussing how data models and the underlying information may change over time;*
- *Provides a foundation that may be expanded by other efforts to cover long-term preservation of information that is NOT in digital form (e.g., physical media and physical samples);*
- *Expands consensus on the elements and processes for long-term digital information preservation and access, and promotes a larger market which vendors can support;*
- *Guides the identification and production of OAIS-related standards.*

This model is applicable to organizations with the responsibility of making information available for the long term this includes organizations with other responsibilities such as processing and distribution in response to programmatic needs. Also those organisations and individuals who create information that may need long term preservation and those that may need to acquire such information from such archives (CCSDS, 2002).

2.5. THE OAIS FUNCTIONAL MODEL

The OAIS functional model has six function modules. **Ingest**, **Data Management**, **Archival Storage** and **Access** providing operations for receiving, storing and retrieving data. On the other side we have **Preservation Planning** and **Administration** to manage the preservation system and preserved data. The following discussion of the six functions shall be related to cost. Figure 4 gives a clear picture of the six function modules of the OAIS.

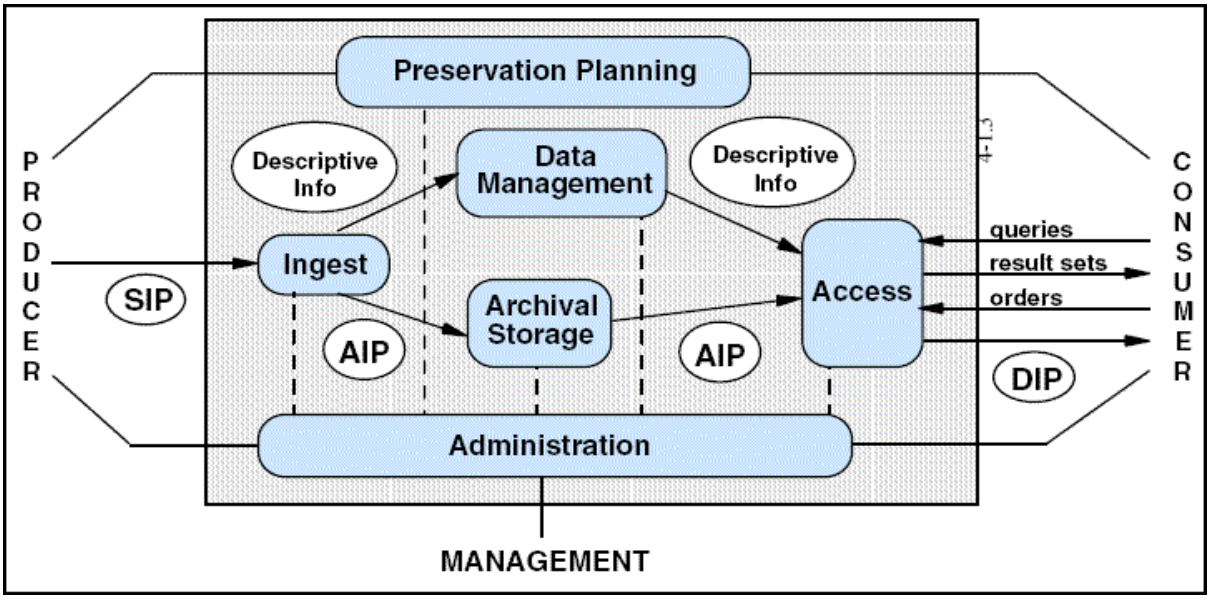


Figure 3: The OAIS Functional model is divided into six entities (here shown in blue). CCSDS (2002)

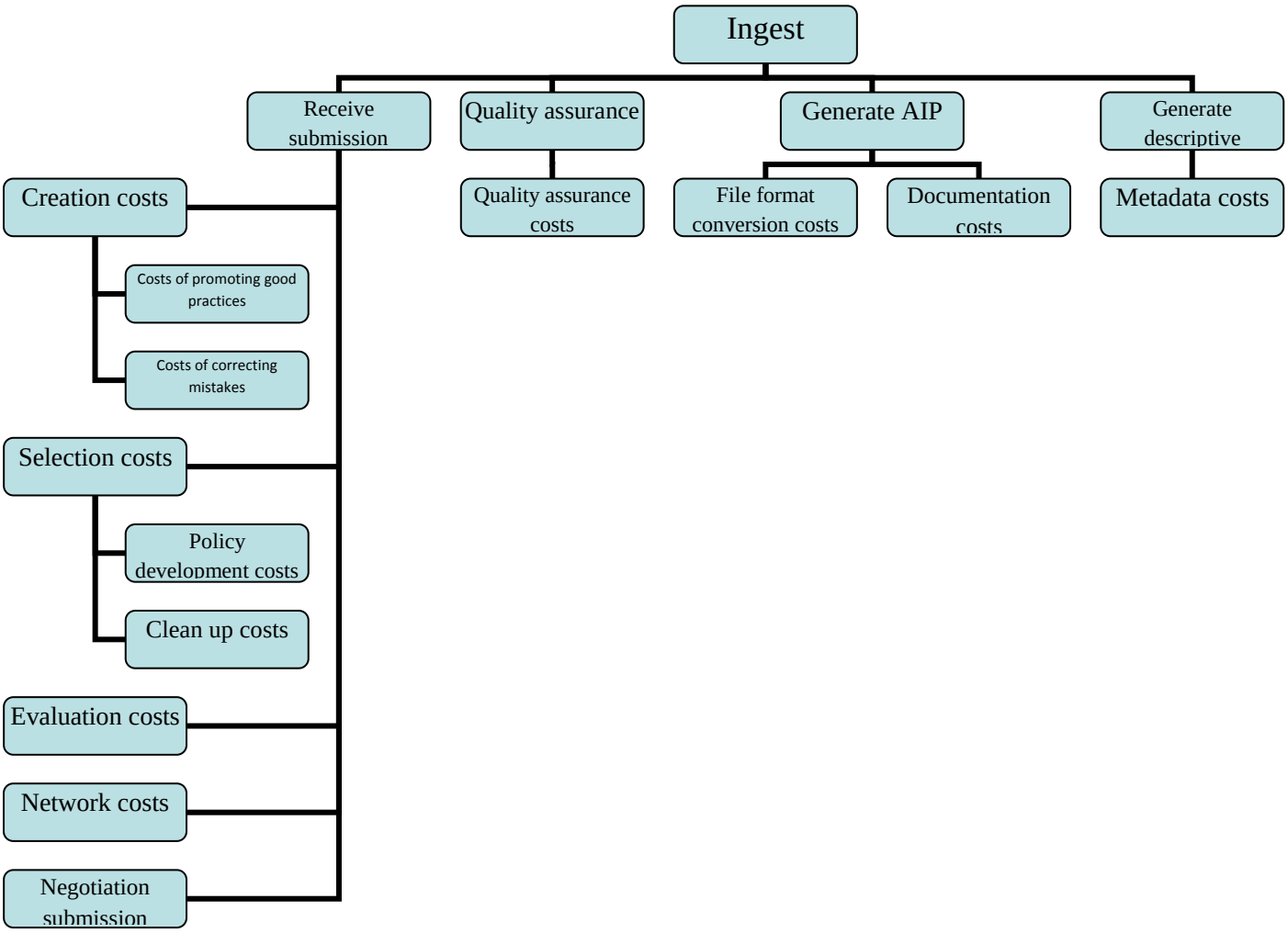
2.5.1. INGEST ENTITY

This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs, performing quality assurance on SIPs, generating an Archival Information Package (AIP) which complies with the archive's data formatting and documentation standards, extracting Descriptive Information from the AIPs for inclusion in the archive database, and coordinating updates to Archival Storage and Data Management (CCSDS, 2002).

According to (Rusaalepp, 2003) the ingest of data resources into an archive starting from negotiating the submission to storing the verified and documented resource for long-term preservation must be well defined, transparent and documented. The ingest process must be efficient and smooth, yet discriminating to reject any data and documentation that is not fit for long-term preservation. Therefore, a digital archive must ensure that the quality of digital resources it ingests for preservation is sufficient for them to be retained over long term and that access to them can be maintained.

Figure 5 gives a summary of the cost factors that will be discussed under the Ingest as a part of the OAIS functional model.

THE COSTS FACTORS IN THE INGEST



2.5.1.1 RECEIVE SUBMISSION FUNCTION

The Receive Submission function provides the appropriate storage capability or devices to receive a *SIP* from the Producer (or from Administration). Digital SIPs may be delivered via electronic transfer (e.g., FTP), loaded from media submitted to the archive, or simply mounted (e.g., CD-ROM) on the archive file system for access. Non-digital SIPs would likely be delivered by conventional shipping procedures. The Receive Submission function may represent a legal transfer of custody for the content information within the SIP, and may require that special access controls be placed on the contents. This function provides a *confirmation of receipt* of a SIP to the Producer, which may include a *request to resubmit* a SIP in the case of errors resulting from the SIP submission (CCSDS, 2002).

2.5.1.1 .1. CREATION COSTS

Even though, the creation of digital objects is not directly related to the OAIS model, it is of worth mentioning that organisations need to be informed of the creation costs in digital preservation. Decisions made when a digital resource is created will have a significant impact on the options subsequently available

for its future management, use and preservation. They will also significantly affect the cost of those options. Some organisations involved in preservation are able to exert considerable control over the creation process, while others have little or no influence. (ERPANET, 2003) indicates that within an organisation it is more possible to influence creation of documents information objects than from outside the organisation. This is because policies, standards and procedures can help their creation and management. On the other hand if there is less influence on the creation of digital records it means mostly higher costs incurred.

(Granger, 2000) states that for digital materials, the link between creation and preservation is much more important because decisions about the way a digital object are created influences how (or indeed whether) it can be preserved. Likewise, decisions taken at the time of preservation can impact on how (or indeed whether) the material can be accessed in the future. Therefore the "costs" of preservation start with the creation of the digital object. For example, open file formats are always fully documented and there is no licence or licence fee required and the users can freely modify the format structure compared to proprietary formats which are not public and in most cases modification is not allowed. Therefore, it is necessary to understand the costs involved in handling the various digital objects during creation.

(Hendley, 1998) points out that for the data centre, archive or library, the cost of promoting good practice to depositors in order to save money at the management and preservation stages can be expected to modify the cost associated with correcting mistakes and bad practice. The adoption of best practices during creation can help simplify the task of preserving and managing the digital resources and this may lead to reduction in costs.

Hendley further points out that, organisations therefore cannot ignore the creation stage even if it is outside their control. They face two major cost areas relating to the creation stage.

- The cost of promoting good practice to depositors. This can include posting guidance notes on web pages, running courses for defined user groups, educating funding bodies, and incorporating guidance notes in funding literature. The costs will vary depending on the type of data centre/archive/library and the clientele they serve. All data centres should invest at least some resources in this preventative measure.
- The cost of correcting mistakes and examples of bad practice at the creation stage. If much is spent in promoting good practices then less will be spent in correcting the mistakes.

In addition to these two, there is also the cost of supporting rare file formats.

2.5.1.1.2. SELECTIONS COSTS

When the SIP is received from the producer or administrator, there has to be a selection of digital objects. According to (Cedar, 2002) the collection management policy is an important tool for defining what materials are of long-term interest to the collection. As such it will need to specify what exactly will be preserved if a digital object is to be retained. What functionality or "look and feel" will be important to preserve? Such decisions will influence the level and method of access that will be necessary for the object as well as the level of preservation metadata required for long-term retention. Cedar continues to argue that, for digital materials, value judgements made by the archivist and/or collection manager will determine what level of functionality needs to be retained, in cases where they may not have the necessary background to do the judgement the top level management may have to be involved. The cedars project has coined the term "significant properties" to describe those components of a digital object deemed necessary for its long-term preservation. Determining the significant properties of a digital object (i.e. the acceptable level of functionality) will dictate the amount of information or "metadata" (including detailed technical metadata) that must be stored alongside the byte stream to ensure the object remains intelligible

(Hendley, 1998) offers the following two practical ways of limiting these costs:

- *Define a basic standard, where valuable resources which fall below that standard are brought up to that standard but not taken beyond it.*
- *Stipulate that digital resources which do not meet a minimum standard in areas such as documentation will be rejected. Both the promotional (best practice) and the corrective (clean-up) costs relate directly and indirectly to preservation.*

Selection is very important in the receive submission as each new acquired data resource means a new long-term responsibility Hence, assessing the future costs associated with each new submission (beyond the ingest operation time) is part of efficient collection management.

2.5.1.1.3. EVALUATION COSTS

(Hendley, 1998) points out that the costs associated with evaluating digital resources, assessing them against a series of technical and practical criteria relate directly to preservation. He further state that, the cost level involved will depend on the size and complexity of the digital resource and how well documented it is. These costs tend to vary depending on the level of stability of the organisation. For example organisations which are stable have automated the evaluation system while others use the staff to do the evaluation manually which tends to increase the costs.

2.5.1.1.4. NETWORK COSTS.

Archival repositories that receive large numbers of digital records from diverse locations may require a high-speed connection or a flexible connection capable of accommodating varying loads. (National Archief, 2005). The costs of acquiring a good network system that ensures successful submission of SIPs and one that is able to detect errors resulting from the SIP submission have to be taken into consideration as part of the costs in the ingest.

According to the OAIS model (CCSDS, 2002), the Network services provide the capabilities and mechanisms to support distributed applications requiring data access and applications interoperability in heterogeneous, networked environments. These services include the following:

- *Data communication includes API and protocol specifications for reliable, transparent, end-to-end data transmission across communications networks.*
- *Transparent file access to available files located anywhere in a heterogeneous network.*
- *Personal/micro computer support for interoperability with systems based on other operating systems, particularly microcomputer operating systems, which may not be formally specified in a national or international standard.*
- *Remote Procedure Call services include specifications for extending the local procedure call to a distributed environment.*
- *Network security services include access, authentication, confidentiality, integrity, and non-repudiation controls and management of communications between senders and receivers of information in a network* (CCSDS, 2002).

A good network system should be able to provide all the above services. This means that the costs of a good network system should include access, authentication, confidentiality, integrity control and management of communication between the sender and the receivers of information in a network. This may vary in most cases for example in organisations where already an IT network exists there might be no real costs directly related to long term preservation. However, in cases where it does not exist, it may involve sharing costs with other departments in an organisation.

2.5.1.2. THE QUALITY ASSURANCE FUNCTION

The Quality Assurance function validates (QA results) the successful transfer of the SIP to the staging area. For digital submissions, these mechanisms might include Cyclic Redundancy Checks (CRCs) or checksums associated with each data file, or the use of system log files to record and identify any file transfer or media read/write errors(CCSDS, 2002).

(Warren, 2003) in his book *Hacker's Delight* defines the cyclic redundancy check, or CRC, as a technique for detecting errors in digital data, but not for making corrections when errors are detected. It is used primarily in data transmission. In the CRC method, a certain number of check bits, often called a checksum, are appended to the message being transmitted. The receiver can determine whether or not the check bits agree with the data, to ascertain with a certain degree of probability whether or not an error occurred in transmission. If an error occurred, the receiver sends a "negative acknowledgement" (NAK) back to the sender, requesting that the message be retransmitted. These are basically computational costs in quality assurance.

(Rusaalepp, 2003) states that checking the conformity of the data resource to its model described in the accompanying documentation serves from the Service Provider's point of view, the purpose of a quality control. The Archive must ensure that it can provide the service it claims to, with the data resource that has been submitted. Thus, the archives usually carry out a number of checks to discover any problems, anomalies and potential difficulties with the files and data they have been given. He further states that the consistency check is also important for reducing risks of introducing errors through conversion and processing when creating the preservation and dissemination versions of the data resource.

QUALITY ASSURANCE COSTS

Data integrity services ensure that data is not altered or destroyed in an unauthorized manner, this service applies to data in permanent data stores and to data in communications messages (CCSDS, 2002). (Dollar Consultant, 2002) states that these integrity check verifications are likely to be labour intensive and cumbersome if they are manually performed, and in practice, are unworkable if thousands of integrity check verifications are undertaken. Automation of these procedures could substantially reduce the requirement for human involvement and make them more workable. This means that the costs might be higher if it is manually done, however if the services are automated it may be less costly.

2.5.1.3. THE GENERATE AIP FUNCTION

The Generate AIP function transforms one or more SIPs into one or more AIPs that conform to the archive's data formatting and documentation standards. This may involve file format conversions, data representation conversions or reorganization of the content information in the SIPs. The Generate AIP function may issue report requests to Data Management to obtain reports of information needed by the Generate AIP function to produce the Descriptive Information that completes the AIP. This function sends SIPs or AIPs for audit to the Audit Submission function in Administration, and receives back an audit report (CCSDS, 2002).

2.5.1.3.1. FILE FORMAT CONVERSION COSTS

The growing complexity of digital objects has led to the increase in file formats for example text files, audio files, video files etc. The challenge is how to handle the various file formats in which information is coded. This is because different file formats have different requirements in terms of long term preservation.

(Mestl et al., 2007) state that a file format specifies a bit stream that is read/ written from/to a non volatile storage medium, without the exact knowledge of the stream of 0s and 1s files cannot be interpreted correctly and the data content in files can in the worst case be useless. The file format must be seen as a way of decrypting the content of a file. They state that formats can be distinguished between two different groups, proprietary and open formats which they describe below:

- *Proprietary formats is where some companies or organisations owns the file specifications and do not want to make them public. The format code is usually not available to the end user and restrictions for using and modifying any proprietary file format may apply.*
- *Open formats or non proprietary are always fully documented; no licence or licence fees required and the user can freely modify the format structure.*

According to IANA[3], there are thousands of different file formats, which have been categorized into eight main groups and broken down into subcategories. The information on file formats will enable an organisation to make decision regarding preservation of the digital object. This is because complex objects or compound objects consisting of different types of formats may also have dynamic behaviours and the complexity of the digital objects entails more maintenance which means more costs.

Open file formats are preferred for long term preservation but many of the records created or received in the organisation may not offer any choices in regard to storing a file in different formats where as others may allow it .Also different file formats do not only store the data in different bit stream but may actually not allow storing all the desired information. Therefore in preparing the digital records for ingestion, the organisation should consider how the stored information shall be used in the future for example if its really necessary to have all the functionality encoded in the file then a format shall be chosen that encodes for that , where as if only viewing some presentation encoding will be suitable enough. The cost of file format conversion will depend on the selected file format. One recommendation that is agreed upon by (NLA, 2002) &Kenney, 2000) is the use of a standard format that is non-proprietary as these kinds of formats are more likely to have a preservation path in the future and software needed to read files.

[3] IANA(Internet Assigned Names Authority).http://www.iana.org/assignments/media-types/

Use of file formats which have been well documented, are widespread in use, have undergone thorough testing and are non-proprietary and usable on different hardware and software platforms minimises the frequency of migration and reduces the risk and costs in their preservation (Maggie & Beagrie, 2002). Narrowing the range of file formats handled streamlines the management process and reduces preservation costs and also reduces the ongoing cost of software licences required by the institution (Jones & Beagrie, 2001). This applies even if free open source software is being used, since most studies show that there is still an associated total cost of ownership and more applications will always cost more to support and maintain.

The complexity of the material submitted and number of objects acquired generally has more impact on costs than the total storage size. The type and variety of formats accepted into the repository will also affect cost, for example proprietary formats are likely to be more difficult and expensive to manage in the long term. It may be possible to reduce costs by limiting the formats the repository will accept, or transforming material into a standard common format. This can be done by reducing the number of file types and possibly reducing the storage size.

2.5.1.3.2. DOCUMENTATION COSTS

According to (Hendley, 1998) data documentation refers to the extent to which the digital resource's structure, content, provenance and history have been documented. He states that the first task in the ingest stage, involves checking the documentation supplied with the digital resource, editing it or adding to it, if required, and then managing the documentation and ensuring that it conforms to the documentation standards. The costs may increase if for instance the documentation is provided in paper form it will require a digitization of it (which will include the cost of time, manual costs, and hardware and software costs for digitization).

However, (Hendley, 1998) points out that if centres put more resources into promoting good practice to depositors this will better the documentation and hence lower the costs associated with reading, amending and managing the documentation. Even when a complete set of documentation is provided, the resources required to study the documentation is significant especially when it has to be done manually. When the documentation is poor then clearly the costs increase dramatically as the centre has to test the digital object and produce or request additional documentation.

The costs associated with reading, editing and managing the documentation for a digital object do relate directly or at least indirectly to preservation. Without documentation it might be difficult to know whether a digital object can be preserved and to determine the preferred preservation strategy for that resource (Hendley, 1998).

2.5.1.4. GENERATE DESCRIPTIVE INFORMATION FUNCTION

The Generate Descriptive Information function extracts Descriptive Information from the AIPs and other sources as a basis for Data Management. This includes metadata to support searching and retrieving AIPs (e.g., who, what, when, where, why), and could also include special browse products (thumbnails, images) to be used by Finding Aids (CCSDS, 2002).

METADATA COSTS

(Day, 2002), points out that since the mid-1990s there has been a growing awareness of the part that metadata can play in supporting the long-term preservation of digital objects and that preservation function has been integrated to some definitions of metadata. (Cunningham, 2000) defines metadata as "structured information that describes and/or allows us to find, manage, control, understand or preserve other information over time". (Calanag, Tabata & Sugimoto, 2001) have commented that extensive metadata is the best way of minimizing the risks of a digital object becoming inaccessible.

Metadata[4] is fundamental to preserving digital resources. Without quality metadata there is little or no value in preserving the resource. Preservation metadata includes a number of different types of metadata: administrative (used in managing information resources including rights and permissions), technical (describing hardware and software needed to maintain an information object) structural (identifying the relationships between objects such as part of, dependent upon that form intellectual entities), and provenance (metadata documenting the history of the object and any actions taken to maintain and provide access). The time and human level-of-effort involved in creating and maintaining preservation metadata is significant and carries a high cost (Irwin, Gewirtz, Glick, Novak & Pilette, 2006).

The more the descriptive information the higher the cost and this should be linked with the value of the records. (Crystal & Land, 2003) in their paper on metadata and search they found out that the estimates of the large scale metadata creation made it evident how important the process is. (Doanne, 2003) observed that his company typically charges from $195,000 to $275,000 to initially set up a metadata solution for a corporation (which will then face additional ongoing costs). Figure 4 below,illustrates the difference in terms of metadata costs. The diagram illustrates what it will cost a company on average per employee for file searching, verification, organization and backup as compared to hiring a consultant to do the initial setup of the metadata.

[4] http://www.library.yale.edu/iac/DPC/DigitalPreservationCostCentersFinal1.pdf

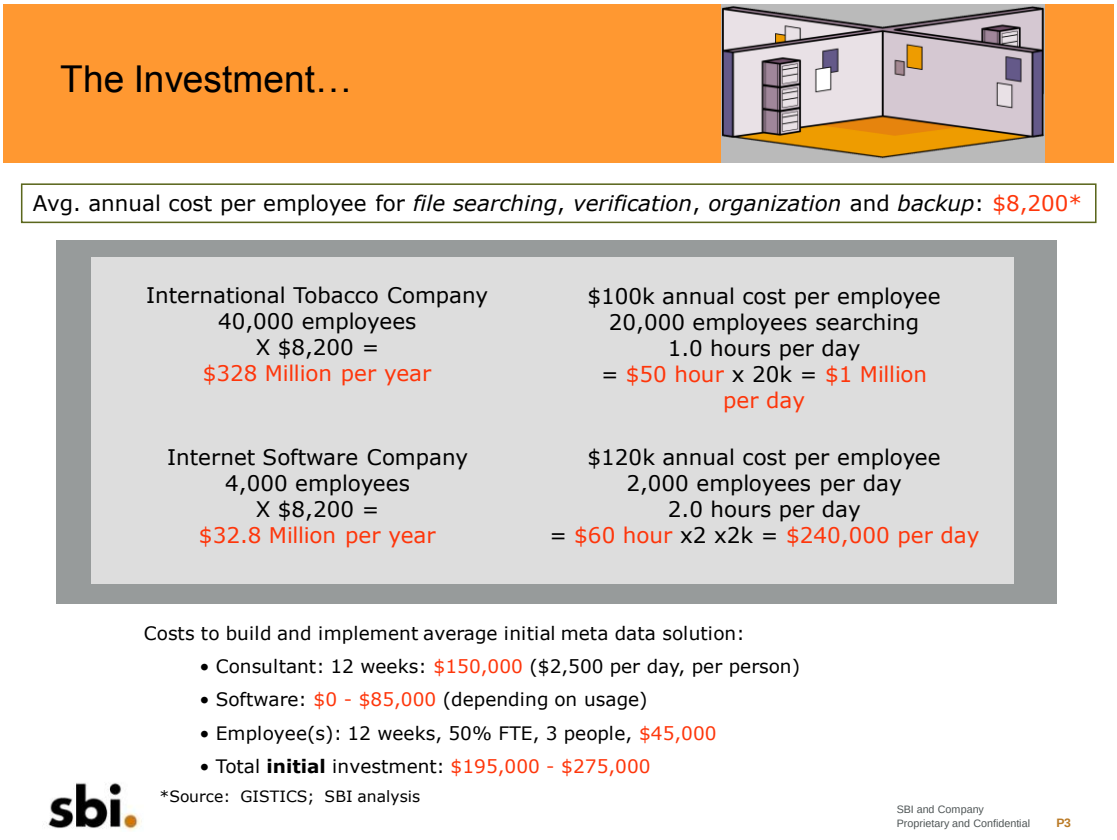


Figure 4: The difference in metadata cost. Doanne (2003). (http://dublincore.org/groups/corporate/Seattle/)

However, (Crystal & Land, 2003) point out that these enormous costs have tempted many organizations to consider by passing a central cataloguing operation in favour of resource authors creating metadata directly. But resource authors often create metadata records of poor quality, including incomplete or inaccurate information. It then falls on information specialists to identify and repair these "broken" records. The fixing of these mistakes can be very costly.

Development of appropriate representation information or detailed technical metadata should also be represented in the preservation metadata and will require specific technical expertise. Metadata costs will also need to accommodate the gathering of rights management information (Granger, 2000).

(Doanne, 2003) states that it is important to mention that the return on investment on metadata can be achieved and any improvement in the ability to find content reduces the overall cost to individuals, workgroups and the enterprise. He further states that using metadata in an intranet environment to reduce employee time spent finding and verifying files may save, at a conservative estimate, $8,200 per employee.

2.5.2. ARCHIVAL STORAGE ENTITY

(Lavoie, 2004) gives a detailed description of the archival storage entity as the portion of the archival system that manages the long-term storage and maintenance of digital materials entrusted to the OAIS. More specifically, the Archival Storage function is responsible for ensuring that archived content resides in appropriate forms of storage e.g., online, near-line, off-line and that the bit streams comprising the preserved information remain complete and render able over the long-term. To meet this responsibility, Archival Storage periodically undertakes procedures such as media refreshment or format migration. He further states that the Archival Storage function also implements various safeguard mechanisms, such as error-checking procedures, to evaluate the outcome of preservation processes, as well as disaster recovery policies to mitigate the effects of catastrophic events. Finally, Archival Storage retrieves items from the OAIS's storage systems in support of access requests by Consumers.

COST FACTORS IN ARCHIVAL STORAGE

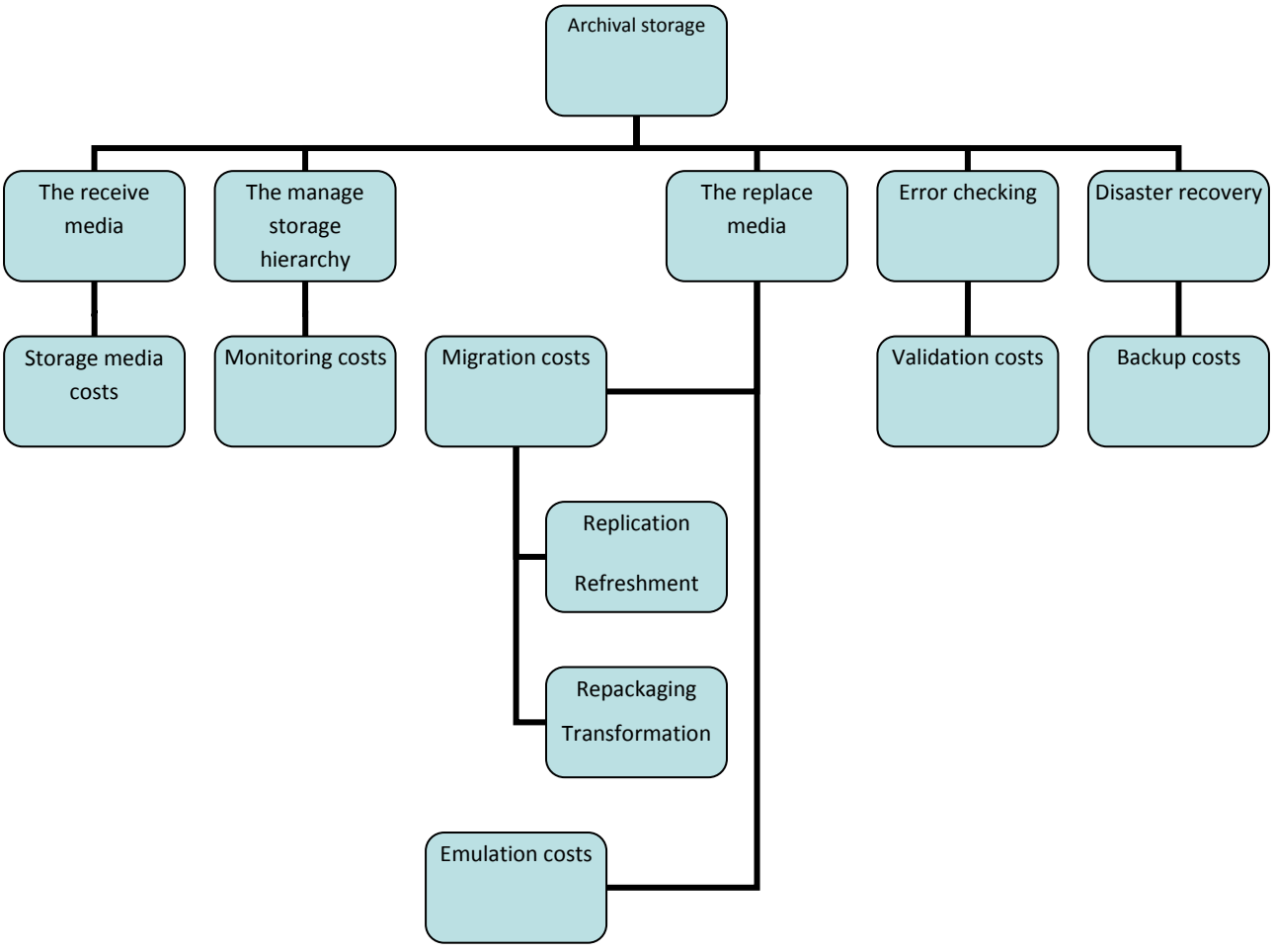


Figure 5: cost factors identified in archival storage of the OAIS functional model

2.5.2.1. RECEIVE DATA FUNCTION

The receive data function receives a storage request and an AIP from Ingest and moves the AIP to permanent storage within the archive. The transfer request may need to indicate the anticipated frequency of

utilization of the data objects comprising the AIP in order to allow the appropriate storage devices or media to be selected for storing the AIP. This function will select the media type usually discs and tapes, prepare the devices or volumes, and perform the physical transfer to the Archival Storage volumes. Upon completion of the transfer, this function sends a storage confirmation message to Ingest, including the storage identification of the AIPs (CCSDS, 2002).

2.5.2.1.1. STORAGE MEDIA COSTS

(Brown, 2003) provides general advice on issues which should be considered by the creators and managers of digital records when selecting physical storage media for long-term preservation. He points out that Server-based hard disk storage is the most effective and secure storage regime for electronic records, provided it is well managed and includes an effective back-up strategy. He further states that any physical storage medium is, by definition, completely dependent upon very specific combinations of hardware and software for access and the accessibility of information stored on such media is therefore highly vulnerable in today's rapidly evolving technological environment. The need to periodically refresh electronic records onto new media is inescapable for the foreseeable future. Nevertheless, careful selection of appropriate media can maximise the periods between refreshment cycles and simplify the refreshment process, in addition to providing the securest storage environment possible.

The scorecard approach described here by (Bernett, 1997) provides a simple method for evaluating currently available media against the selection criteria. Each medium should be scored against the criteria on a scale of 1 (does not meet the criterion) to 3 (fully meets the criterion). As a general rule, no medium which scores less than 12 should be considered.

Media	CD-R	DVD-R	Zip Disk	3.5" Magnetic Disk	DLT	DAT
Longevity	3	3	1	1	2	1
Capacity	2	2	1	1	3	3
Viability	2	2	1	1	3	3
Obsolescence	3	2	2	3	2	2
Cost	3	2	1	1	3	3
Susceptibility	3	3	1	1	3	2
Total	16	14	7	8	16	14

Figure 6: An example scorecard, comparing some common media types

(Beagrie, Chruszcz & Lavoie, 2008) in their research on "*Keeping Data Research Safe"* they point out that storage media will be selected on the basis of service requirements e.g. data volumes, required speed of

access, or archival properties and cost. The selection of storage media will influence the frequency of future storage media migration and staff and equipment needed for this task. They further point out that, it is important to remember that the total cost of ownership of archive storage media and systems is substantially higher than the purchase cost alone.

Organisations will incur costs in acquiring the storage media that will be used in digital preservation. (National Archief, 2005) states that appreciable costs can be incurred in the purchase of storage media (tapes, CDs, DVDs) and that organisations have to make sure an adequate amount is budgeted for effective storage and backup media. The storage options are impacted by the resources available, by the volume of data to be stored and by how it will be used and preserved.

The cost of file storage will include: maintenance and purchase of hardware; transferring or refreshing files from generation to generation of storage media; and scheduled testing of the ability to use the intellectual content of the digital object. This will vary according to changes in the technology marketplace, and on current expectations is likely to decrease in relative terms, over time (Cedar, 2002).

2.5.2.2. MANAGEMENT OF STORAGE HIERARCHY

The management of data within the archive needs to take account of storage management policies, operational statistics, or directions from the Ingest stages. Cost will be affected by any special levels of service, or any special security / protection measures that are required. These include on-line, off-line or near-line storage, required throughput rate, maximum allowed bit error rate, or special handling or backup procedures. Monitoring is needed to ensure that no corruption of data occurs during transfers. However, the size and complexity of the archive will impact both the necessity and the cost of providing operational statistics summarizing the inventory of media on-hand, available storage capacity in the various tiers of the storage hierarchy, and usage statistics (Beagrie, Chruszcz &Lavoie, 2008).

2.5.2.2.1. MONITORING COSTS

All storage media need to be monitored for signs of data loss. The sample and frequency with which this is done will influence costs. This will be a more significant cost for storage media requiring manual intervention and inspection compared to automated systems (Beagrie et al., 2008). The costs may include the time and the manpower required in case monitoring is done manually, however if it is automated the hardware and software costs may be significant.

2.5.2.3. THE REPLACE MEDIA (MEDIA REPLACEMENT) FUNCTION

(Chen, 2001) states that digital preservation is plagued by the short media life, obsolete hardware and software, and slow read times of old media and the rapid technological advances do not solve the problem;

instead, there is need to migrate digital objects from one generation of technology to another every few years. He points out that for digital records, the preservation issues extend beyond media life considerations and devices for reading these media rapidly become obsolete; the various formats for digital documents and images introduce additional complications

2.5.2.3.1. MIGRATION COSTS

CCSDS (2002) defines digital migration as the transfer of digital information, while intending to preserve it, within the OAIS. It is distinguished from transfers in general by three attributes:

- A focus on the preservation of the full information content
- A perspective that the new archival implementation of the information is a replacement for the old
- Full control and responsibility over all aspects of the transfer resides with the OAIS.

(Kol & Oltmans, 2005) define migration as copying data, or converting data, from one technology to another, whether hardware or software, preserving the essential characteristics of the data. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. They further point out that with migration, file formats will be converted into new formats as soon as the original formats run the risk of becoming obsolete. They give the example of, if technology scans indicate that PDF version 1.1 will soon be inaccessible, all files in the digital archive of format PDF 1.1 will have to be converted into, for example, PDF format version 1.4. This way, the digital publications will be prepared for rendering for another period of time, until the format PDF version 1.4 runs the risk of becoming obsolete itself. At that time another migration procedure will need to be carried out.

(Kol &Oltmans, 2005) argue that with migration the electronic publications will always be available in the form that is generally accepted, e.g., PDF, and current hardware and software will be able to render these formats with little difficulty and older documents that are properly migrated will be available for some time in the present and the near future, and their electronic content can be used for copy and reuse. However they indicate that a major drawback might be that while converting documents from one form to another, some aspects of the document's layout or even worse data might get lost. Moreover, with migration it may be impossible to perform a conversion if the file format and the migration tool are no longer active. Therefor they advise that when applying the migration strategy we have to constantly study conversion programs and execute them when possible, so as to prevent digital information from getting lost

The OAIS model identifies the four primary digital migration types as ordered by increasing risk of information loss. They include:

- ***Refreshment:*** *A Digital Migration where a media instance, holding one or more AIPs or parts of AIPs, is replaced by a media instance of the same type by copying the bits on the medium used to hold AIPs and to manage and access the medium. As a result, the existing Archival Storage mapping infrastructure, without alteration, is able to continue to locate and access the AIP. It offers a short term solution for preserving access to digital materials by ensuring that information is stored on newer media before the old media deteriorates beyond the point at which the information can be retrieved,*
- ***Replication****: A Digital Migration where there is no change to the Packaging Information, the Content Information and the PDI. The bits used to convey these information objects are preserved in the transfer to the same or new media-type instance. Note that Refreshment is also a Replication, but Replication may require changes to the Archival Storage mapping infrastructure.*
- ***Repackaging****: A Digital Migration where there is some change in the bits of the Packaging Information during the transfer.*
- ***Transformation****: A Digital Migration where there is some change in the Content Information or PDI bits while attempting to preserve the full information content. The resulting AIP is intended to be a full replacement for the AIP that is undergoing transformation. The new AIP qualifies as a new version of the previous AIP. The first version of the AIP is referred to as the original AIP and may be retained for verification of information preservation* (CCSDS, 2002)

While it is difficult to project future migration costs, we do know that there will be recurring costs. This requires an ongoing and significant investment of resources in order to ensure that the expertise, financial resources, and equipment are available to meet this need as frequently as is necessary.

2.5.2.3.2. EMULATION COSTS

Emulation combines software and hardware to reproduce in all essential characteristics the performance of another computer of a different design, allowing programs or media designed for a particular environment to operate in a different, usually newer environment. Emulation requires the creation of emulators, programs that translate code and instructions from one computing environment so it can be properly executed in another (Kol & Oltmans, 2005).

(CCSDS, 2002) discusses the advantage of hardware emulation as the claim that once a hardware platform is emulated successfully all operating systems and applications that ran on the original platform can be run without modification on the new platform. However, this does not take into account dependencies on input/output devices. For example, it may not be possible to fully simulate all of the old hardware

dependencies and timings, because of the constraints of the new hardware environment. Further, when the application presents information to a human interface, determining that some new device is still presenting the information correctly is problematical and suggests the need to have made a separate recording of the information presentation to use for validation. Once emulation has been adopted, the resulting system is particularly vulnerable to previously unknown software errors that may seriously jeopardize continued information access. Given these constraints, the technical and economic hurdles to hardware emulation appear substantial (CCSDS, 2002).

Another advantage of emulation as discussed by (Oltman, 2004) is that the original "look and feel" of the publication can be preserved. As with preserving books, the authentic instantiation will be there to be rendered, in contrast to migration in which possible other instances are used rather than the original. However, a serious drawback is the complexity of developing and maintaining emulation tools. In the future, we will have to maintain several emulation tools, and it cannot be proven that these will always work on future computer platforms (Oltmans, 2004). He further states that emulation models require more initial investments compared to migration models. Emulation tools have to be developed, and this requires serious Research &Development, including technical skills to implement the concepts. Furthermore, emulation tools have to be maintained over time, which also requires investments in both researchers and implementations (Oltmans, 2004).

A very good example that has been illustrated by (Oltmans, 2004) in comparing costs between migration and emulation can best be used in deciding which preservation approach is suitable in any particular situation.

He gives the following formula when calculating storage costs, and in this cases the formula for calculating the costs for migration and emulation.

The formula for calculating migration costs is as follows:

$K(t,a) = h(t,a) + m(t,a)$

Where K (t) is the total cost of holding a objects for a period of t years, and

h=storage, m=migration. a= number of objects, t= time

The costs of migrating digital objects is dependent of time t (the longer we preserve the objects, the more often we have to convert them) and of the number of objects a (the more objects in the archive, the more conversion actions have to be executed).

The formula for calculating emulation costs is as follows:

K (t,a) = h(t,a) + E + e(t)

Where K (t) is the total cost of holding a objects for a period of t years, and h=storage, E=setting up initial emulation tool, and e (t) =emulation over time.

The costs for buying an emulation tool are expressed by E, while the yearly maintenance is expressed by e. The maintenance is not dependent of the number of objects: emulation tools apply to the entire collection, and no special action is needed when rendering an object in the digital archive.

(Kol and Oltmans, 2005) advise that, where the conversion of objects to other formats constitutes a considerable cost factor in migration, these costs can be saved when applying emulation. However, emulation requires more initial investments, which makes it inappropriate for short-term preservation. Therefore, for a proper cost comparison, the costs of each strategy should be specified in relation to the term for preservation. Emulation costs can be classified by the tasks that should be executed in order to realize access in the future

2.5.2.4. ERROR CHECKING FUNCTION

This is a function that provides statistically acceptable assurance that no components of the AIP are corrupted during any internal archival storage data transfer (CCSDS, 2002).

The costs could include software and hardware that monitors the records to ensure that they are not altered by sending notification when such errors occur. This will also include standard mechanisms for tracking and verifying the validity of all data objects within the archive.

2.5.2.4.1. VALIDATION COSTS

The cost of validating the completeness of an object according to (CCSDS, 2002) could be maintained for every individual data file. A higher level of service such as Reed Solomon coding to support combined error detection and correction, could also be provided. The storage facility should provide for random verification of the integrity of data objects using CRCs or other error checking mechanisms.

(Hendley, 1998) states that the key to validation is assessment which is done to ensure the following:

- That the resource is complete as documented
- That the resource is functioning properly and operates on the specified hardware and software environments
- That the resource is consistent

The costs associated with assessing digital objects all relate directly or at least indirectly to preservation. Without assessment it is impossible to know whether a digital object can be preserved and to determine the preferred preservation strategy for that resource (Hendley, 1998).

2.5.2.5. DISASTER RECOVERY FUNCTION

This function provides a mechanism for duplicating the digital contents of the archive collection and storing the duplicate in a physically separate facility. This function is normally accomplished by copying the archive content to some form of removable storage media but may also be performed via hardware transport or network data transfers (CCSDS, 2002).

One recommendation to handle media issues is to create backups (Cedars, 2002, Kenney, 2000 &IMLS, 2001) using more than one kind of backup software to write the copies so as to safeguard against software bugs. In this scenario, at least one copy should be maintained in an offsite location and the media periodically checked as per the refreshing methodology.

Media storage areas should be locked and accessible only to properly trained personnel. All media, no matter how reliable, needs to be backed up. Creation of multiple backups and the use of off-site storage for one set of copies provide the best protection against catastrophic loss (Cornell Library University, 2005).

2.5.2.5.1. BACKUP COSTS

According to (Cornell Library University, 2005) the maintenance of redundant copies of valued digital content is an essential component of any digital preservation program, and a key element in the prevention of catastrophic loss. They suggest that a great variety of backup solutions are available, which to use depends on:

- quantity of data
- rate of change
- degree of automation desired
- available budget

In addition to backing up data files, application software and operating systems may also need backup. In some cases, it may be necessary to purchase additional licenses or obtain special permission from the software vendor in order to back up applications (Cornell Library University, 2005).

In addition to testing backup media periodically to ensure the data is still readable and has not been altered, restore procedures should also be tested to ensure that the hardware, software and any outside vendors involved in maintaining backups are all functioning as expected (Cornell Library University, 2005).

(Cornell Library University, 2005) points out that a prudent backup strategy places at least one copy of all critical data at a sufficient distance from the main data store so that it is not likely to succumb to the same disaster. This is called off-site storage. They further advise that institutions should check with regulatory agencies for their records retention requirements and establishing a reciprocal storage arrangement with a cooperating institution may be a low-cost way to manage off-site storage. They indicate that if outsourcing, make sure you are getting true data management, not just warehousing as generic storage facilities are unlikely to know how to properly store and handle digital media.

2.5.3. DATA MANAGEMENT ENTITY

The Data Management function maintains databases of descriptive metadata identifying and describing the archived information in support of the OAIS's finding aids; it also manages the administrative data supporting the OAIS's internal system operations, such as system performance data or access statistics. The primary functions of Data Management include maintaining the databases for which it is responsible; performing queries on these databases and generating reports in response to requests from other functional components within the OAIS; and conducting updates to the databases as new information arrives, or existing information is modified or deleted. In managing these databases, the Data Management function supports search and retrieval of the OAIS's archived content, and administration of the OAIS's internal operations (Lavoie, 2004).

There have been fairly high costs associated with data management because this is where metadata management largely occurs. As automated tools are developed and implemented, data management costs should become more manageable. Other costs include technical skills, software, and storage. (Cornell Library University, 2005).

The other costs according to (National Archief, 2005) include the purchase of operating systems and the standard software for databases. There will also be a need for protection software (against viruses, unauthorised access, and tampering with the archives by unauthorised persons). There may also be a need for specific software for the receipt and storage of authentic digital archival records, such as Depot 2000 or the Digital Archive System of the United Kingdom National Archives (an example that has been given by the National Archief, 2005). Every organisation which works with digital records will, irrespective of its

size, have a need for a DMS (Document Management System) or an RMA (Records Management Application). Communications software and network and database licences are also factors that should be taken into consideration when budgeting however in most cases they are overlooked (National Archief, 2005).

2.5.4. ADMINISTRATION ENTITY

The Administration function is responsible for managing the day-to-day operations of the OAIS, as well as coordinating the activities of the other five high-level OAIS services. Other responsibilities include interacting with Producers (e.g., negotiating Submission Agreements), Consumers (e.g., providing customer service support), and Management (e.g., implementing and maintaining archive policies and standards). The Administration function is also responsible for overseeing the operation of the archiving and access systems, monitoring system performance, and coordinating updates to the system as appropriate. Administration serves as the central hub for the OAIS's internal and external interactions: it communicates directly with the five other OAIS high-level services Ingest, Archival Storage, Data Management, and Access, as well as the OAIS's external stakeholders Producers, Consumers and Management (Lavoie, 2004).

It relates to negotiating submission agreements, system configuration, physical access control, establishing standards and policies, auditing submissions and providing customer service. This is where the organizational infrastructure meets the technological infrastructure. Organizational resources need to be allocated to automatically (and appropriately) apply relevant policies and/or develop a stringent workflow that provides a manual or semi-automated and well-documented process. This requires policy development and programming skills as well as software and tools (Cornell Library University, 2005).

Every archival institution has a management that is responsible for establishing standards and policies, controls the physical access to the premises, provides customer service, negotiates with depositors for materials to be submitted and performs other management functions. These housekeeping functions are essential to the operation of any archival institution but the OAIS Administration entity is not only a co-ordinating entity, it has direct contact and oversight over all the other functional entities. Thus, Administration serves as the central hub for the OAIS's internal and external interactions: it communicates directly with the five other OAIS services Ingest, Archival Storage, Data Management and Access, as well as the OAIS's external stakeholders Producers, Consumers and Management (Beedham et al, 2004).

2.5.4.1. THE NEGOTIATION SUBMISSION AGREEMENT FUNCTION

This function solicits desirable archival information for the OAIS and negotiates Submission Agreements with Producers. This function also negotiates a data submission schedule with the Producer (CCSDS, 2002).

This may include rights, ownership, scope of content, obligations for management and delivery and deciding which Content Information must be negotiated.

2.5.4.1.1. INTELLECTUAL PROPERTY RIGHTS (IPR) COSTS

Copyright and other intellectual property rights (IPR) such as moral rights have a substantial impact on digital preservation. The preservation of digital materials is dependent on a range of strategies, which has implications for IPR in those materials. The IPR issues in digital materials are arguably more complex and significant than for traditional media and if not addressed can impede or even prevent preservation activities. Consideration may need to be given not only to content but to any associated software. Simply copying (refreshing) digital materials onto another medium, encapsulating content and software for emulation, or migrating content to new hardware and software, all involve activities which can infringe IPR unless statutory exemptions exist or specific permissions have been obtained from rights holders (Jones and Beagrie,2002). As both migration and emulation will involve manipulation and changing presentation and functionality to some degree (especially over any period of time) important issues of principle and practice are raised in negotiations. It is important to establish a dialogue with rights holders so that they are fully aware of these issues and the actions and rights required to ensure the preservation of selected items.

The IPR costs may include any licence fees charged for permission to preserve the digital object, and to allow continued access to it. Other rights expenses may include the cost of research to establish the relevant rights owners, the cost of time spent in the negotiation and drafting of the licence agreement, and any related legal costs (Cedar, 2002). Legal and negotiation skills will be required for rights negotiation. For major decisions (e.g. concerning a whole collection) staff will need to be senior and will involve collection managers. More junior staff may carry out day to day activities.**(**This also represents an ongoing cost; however, it will cost more at early stages until there are processes and workflows in place to deal with it on a day to day basis or through a central agency) (Cedar, 2002).

2.5.4.2. THE MANAGE SYSTEM CONFIGURATION FUNCTION

This function provides system engineering for the system for the archive system to continuously monitor the functionality of the entire archive system and systematically control changes to the configuration. This function maintains integrity and tractability of the configuration during all phases of the system life cycle. It also audits system operations, system performance, and system usage (CCSDS, 2002).

The costs include the software and the hardware required to monitor the functionality of the entire archive system. This system can be automated and may require experts who can recommend on the appropriate system and offer support.

2.5.4.3. PHYSICAL ACCESS CONTROL FUNCTION

The function provides mechanisms to restrict or allow physical access (doors, locks, guards) to elements of the archive, as determined by the archive policies(CCSDS, 2002).

The costs incurred will include the staff costs e.g. the security guards, the costs of physical infrastructure that ensure that the servers are secure, costs for air conditioning, uninterrupted power supply and non labour costs (like secure building). The physical infrastructure includes elements like power supply, cooling system, physical housing, security, fire protection and cabling.

Power is one area that has received great focus in the recent past as a crucial factor that needed to be taken into consideration in terms of the costs in the future. The digital objects can be accessed through hardware and equipments that use power e.g. computers, servers. The heating system and lighting system use power to provide a stable environment for the digital objects. (Gantz et al., 2008) points out that power costs are continuing to rise and to rise more quickly than the costs of new servers. Power consumption, estimated at 1kW per server rack in 2000 is estimated at 20kW per rack in 2007 and project to rise to 20kW per rack in the next few years (Gantz et al., 2008). They point out that greater capacity will not necessarily solve the problem for a custodial institution since the creation of new information is accelerating faster than our capacity to store it and the engineering of cooling large, complex servers is complicated. Hence, power requirements to operate and maintain such systems are considerable and the costs not only of upgrading but also of operating these systems require serious deliberation. At the least, it is essential to consider the costs of power and cooling requirements, especially given that the cost of energy will continue to rise in the future.

2.5.4.4. STANDARDS AND POLICIES

(Best, 2007) starts by posing a question on how much do standards cost to develop? He categorically states that there is no one number, given the wide range in size, scope, and the levels of technical complexity of various standards. As desirable as an accurate number would be, it is difficult even to generalize, but perhaps an example or two can help frame the answer and provide a means for determining the true cost of any particular project. He states that the cost of a standard depends, of course, on a number of variables including the size and complexity of the work, the availability of existing work upon which the standard can be based, the number and quality of resources available to contribute to developing the solution, etc. Therefore there is no “one size fits all” price.

(Best, 2007) outlines the following individual costs that go into developing and approving a standard, including

- Membership dues or participation fees paid by participants to the standards organization;

- Costs associated with the salaries and support for technical resources (people) involved in developing the work;
- Travel costs for attending face to face committee meetings;
- Infrastructure costs such as email lists, web pages, telephone conference calls; and
- Clerical and administrative costs of the standards organization, including for the administration and oversight of the technical process, conducting of public reviews and organizational ballots, and publication of the completed work.

Best further gives the example below to illustrate the costs of developing a standard which he has illustrated using a hypothetical case.

Let's consider a hypothetical standards activity consisting of 20 engineers employed by eight different companies, with each of the companies paying $15k in membership dues to the standards organization. To be fair we should take into account that each of these companies is participating in – let's say three – other technical activities within the organization.

$$\$15,000 \times 8 \times 0.25 = \$30,000$$

Let's also assume that these 20 engineers are working for three years to develop a specification, and are each devoting an average of 20% of their time to the project. We'll assume further that the fully burdened personnel cost of an engineer, including salary, benefits, office space and

Equipment, etc. is $150,000/year. This gives us the following cost for technical resources for the development project:

$$\$150,000 \times 20 \times 3 \times 0.2 = \$1,800,000$$

(You should notice right away that membership dues is a relatively minor cost of participating in Standards work, compared to the cost of participation.)

Additionally, the committee developing the standard meets face to face four times a year. We'll estimate the travel costs at $2000 per person per meeting, and throw in an extra $2000 for meeting space and food. The time cost for the participants will be assumed to be included in the figure above.

$$((20 \times \$2000) + \$2000) \times 4 \times 3 = \$504,000$$

As noted above, the infrastructure and clerical/administrative costs provided by the standards Organization are paid through membership dues, which have already been counted. The total cost for developing our hypothetical standard is

$30,000 + $1,800,000 + $504,000 = $2,334,000

That's just for the development effort, the first phase of the work. The second phase would include the costs associated with the approval of the work including public and member reviews and voting. This could add a cumulative hundreds or possibly thousands of hours of additional work by technical people, increasing the costs even further. There is also the administrative processing and preparation for publication, though as this is done by the standards organization's staff these costs are covered by the membership dues or participation fees already included.

This approach can be used by organisations to calculate the costs of participating in developing standards and policies for digital preservation in organisations (Best, 2007).

2.5.4.5. AUDIT SUBMISSION FUNCTION

This function will verify the submission (SIP and AIP) meet the specification of the submission agreement. The function receives reviews from preservation planning and may also involve an outside committee (e.g. science and technical review). The audit process must verify that the quality of the data meets the requirement of the archive and the review committee (CCSDS, 2002).

The costs involved include auditing costs especially where it involves external auditors who have to verify whether the SIPs and AIPs meet the required specification. The costs vary here from organisation to organisation and depending on the size of the organisation. The variation also may depend on internal archive policies.

2.5.4.6. PROVIDE CUSTOMER SERVICE FUNCTION

This function will create, maintain and delete customer accounts. It will collect billing information from access and will send bills and collect payments from consumers for the utilization of archive system resource (CCSDS, 2002). According to the study team as reported by (Beagrie et al.,2008), customer accounts are useful for reporting usage and restricting access as appropriate to closed collections with specific licence conditions.

The costs involved in this function include the costs of a acquiring a customer relationship management system (CRMS), this will include both the hardware and software costs. According to the (Free online dictionary of computing, 2009), CRM (Customer Relationship Management) are Enterprise-wide software applications that allow companies to manage every aspect of their relationship with a customer. The aim of

these systems is to assist in building lasting customer relationships to turn customer satisfaction into customer loyalty and to improve services provided to customers.

Other relevant cost factors will include dedicated staff that will manage the CRMs and handle customer requests and this will in addition include training costs for the staff to ensure continued support of the system, retraining and skill upgrading costs.

2.5.5. PRESERVATION PLANNING ENTITY

This includes monitoring user groups, emerging technologies, standards and platforms. The objective is to protect against obsolescence, develop preservation strategies and migration plans as required (CCSDS, 2002).

This service is responsible for mapping out the OAIS's preservation strategy, as well as recommending appropriate revisions to this strategy in response to evolving conditions in the OAIS environment. The Preservation Planning service monitors the external environment for changes that could impact the OAIS's ability to preserve and maintain access to the information in its custody, such as innovations in storage and access technologies, or shifts in the scope or expectations of the Designated Community. Preservation Planning then develops recommendations for updating the OAIS's policies and procedures to accommodate these changes. The Preservation Planning function represents the OAIS's safeguard against a constantly evolving user and technology environment. It detects changes impacting the OAIS's ability to meet its responsibilities, designs strategies for addressing these changes, and assists in the implementation of these strategies within the archival system (Lavoie, 2004).

Preservation Planning is perhaps the core OAIS function. It defines and manages strategies to enable digital objects, stored as Archival Information Packages (AIPs), to move through time without suffering unacceptable losses and changes to content or functionality. It starts long before objects are moved into a formal long-term preservation system and continues throughout their existence (Cornell University Library, 2003).

Preservation planning demonstrates more clearly than other functions that digital preservation is a shared responsibility, both within and between institutions. An organization needs to determine the contributions it can make to community-based developments of preservation strategies and the ongoing technology watch and how it can benefit from external research and development. The organization needs to identify and adopt appropriate preservation strategies (Cornell University Library, 2003).

2.5.5.1. THE MONITOR DESIGNATED COMMUNITY FUNCTION

This function interacts with archive Consumers and Producers to track changes in their service requirements and available product technologies. Such requirements might include data formats, media choices, and preferences for software packages, new computing platforms, and mechanisms for communicating with the archive. This function may be accomplished via surveys, via a periodic formal review process, via community workshops where feedback is solicited or by individual interactions. It provides reports, requirements alerts and emerging standards to the Develop Preservation Strategies and Standards function. It sends preservation requirements to Develop Packaging Designs (CCSDS, 2002).

The costs in this function will include dedicated staff who will be involved in monitoring the designated community which varies from one organisation to another depending on the clients they offer services to. The other costs are those incurred in organising community workshops, carrying out the surveys and the periodic reviews.

2.5.5.2. MONITOR TECHNOLOGY FUNCTION

The purpose of this function is to identify technologies which could cause obsolescence in the archive's computing environment and prevent access to some of the archives current holdings. This is achieved by tracking emerging digital technologies, information standards and computing platforms (i.e., hardware and software (CCSDS, 2002).

2.5.5.2.1. TECHNOLOGY WATCH COSTS

An implication of the rapid evolution of storage media and file formats and the risks of technology obsolescence is the necessity of maintaining a register of hardware and software capacity in the institution and preservation metadata to enable a formal process of "technology watch". A failure to implement an effective technology watches or IS strategy incorporating this will risk potential loss of access to digital holdings and higher costs. A retrospective survey of digital holdings and a risk assessment and action plan may be a necessary first step for many institutions, prior to implementing a technology watch. (Jones & Beagrie, 2002).

2.5.5.3. DEVELOP PRESERVATION STRATEGIES AND STANDARDS FUNCTION

This function is responsible for developing and recommending strategies and standards that enable the archive to better anticipate future changes in the Designated Community service requirements or technology trends that would require migration of some current archive holdings or new submissions. This function

receives reports from the Monitor Designated Communities and Monitor Technology functions, and it receives performance information, inventory reports and summarized consumer comments from Administration. This function sends recommendations on system evolution to Administration. This function also receives external data standards from Monitor Technology and produces profiles of those standards that are sent to Administration as proposals on their potential usage. This function also receives issues from Develop Packaging Designs and Migration Plans in the case of unanticipated submission requirements, and responds with advice to handle the new requirements (CCSDS, 2002).

The costs include the costs of developing the preservation strategies and standards, outsourcing costs in case the regular staff in organisations lacks the competency to develop the strategies and standards. This can be costly at initial stages but save times and involves experts who are in a better position to develop proper preservation strategies.

2.5.5.4. THE DEVELOP PACKAGING DESIGNS AND MIGRATION PLANS FUNCTION

This function develops new IP designs and detailed migration plans and prototypes, to implement Administration policies and directives. This activity also provides advice on the application of these IP designs and Migration plans to specific archive holdings and submissions. This function receives archive approved standards and migration goals from Administration. The standards include format standards, metadata standards and documentation standards. It applies these standards to preservation requirements and provides AIP and SIP template designs to Administration. This function also provides customization advice and AIP/SIP review to Administration on the application of those designs (CCSDS, 2002).

The migration goals received by this function tend to involve transformations of the AIP, including transformations of the Content Information to avoid loss of access due to technology obsolescence. The response to the migration goals may involve the development of new AIP designs, prototype software, test plans, community review plans and implementation plans for phasing in the new AIPs. This process may call on expertise or resources from other functions within Preservation Planning, such as prototype development from the Monitor Technology. This effort also will require consultation from the other functional areas and from the Designated Community. Once the migration plan, associated AIP designs, and software have been tested and approved, this function will send the entire migration package to Administration, which will schedule and perform the actual migration (CCSDS, 2002).

The costs include the expertise or resources from other functions to oversee the development of the designs and migration plans. The costs of developing new AIP designs, prototype software, test plans, community review plans and implementation plans for phasing in new AIPs.

2.5.6. ACCESS ENTITY

Access function manages the processes and services by which Consumers and especially the Designated Community locate, request, and receive delivery of items residing in the OAIS's archival store. Typical services provided by Access in support of the Consumer include processing queries of the OAIS's holdings specifically, forwarding the request to Data Management and presenting the response (e.g., a result set) to the Consumer; and coordinating the retrieval and delivery of requested content by forwarding the request to Archival Storage, receiving the requested items, and performing any necessary transformations that must occur prior to delivery to the Consumer. Access is also responsible for implementing any security or access control mechanisms associated with the archived content. The Access function represents the OAIS's interface with its Consumers (and Designated Community): as such, it is the primary mechanism by which the OAIS meets its responsibility to make its archived information available to the user community (Lavoie, 2004).

From a digital preservation perspective as pointed out by (Cornell University Library, 2003) an OAIS must be able to find and deliver objects effectively in response to requests from Consumers (users and user systems). Preservation without the means for access is of debatable value, but access is still not the primary purpose of a preservation system. Access generally demands the use of current technology to render objects that will meet user expectations. The organizational responsibility for access is to ensure that the requirements to deliver can be met, that the user communities are well defined, and that adequate resources are available to support ongoing development in support of access.

Access entails organizational costs (e.g., policy development and maintenance, interface development, requirements analysis) and technological costs (e.g., programming to process and respond to queries, generate DIPS, develop rights management protocols). OAIS acknowledges the organization's option to charge users for copies, which may offset costs (Cornell University Library, 2003).

The Access function represents the OAIS's interface with its Consumers: it is the primary mechanism by which the OAIS archive meets its responsibility to make its archived information available to the user community. It manages the processes and services by which consumers and especially the designated community locate, request, and receive delivery of items residing in the OAIS's archival store. Access functions include communicating with Consumers to receive requests, coordinating the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers. Access is also responsible for implementing any security or access control mechanisms associated with the archived content. (Beedham et al, 2004).

2.5.6.1. FINDING AIDS COSTS

These are costs related to providing everything that is associated with reliably locating and accessing an acceptable presentation of the digital objects e.g., the public interface, web server, indexing, managing appropriate emulation tools, etc. Systems and/or application interfaces are expensive to develop and maintain. There are substantial savings from maintaining a small number of standard interfaces and a proportionately high cost to each custom interface the archive needs to develop. (Beagrie, et al, 2008).

2.5.6.2. SECURITY COSTS

Access to online servers is controlled through the use of password and/or address port filtering. Online data that is fully validated and approved for release is placed on publicly accessible servers (CCSDS, 2002).

Requirement for access control will add costs on a sliding scale depending on the level of control and methods required. Simple closure of a data collection for a specified time period before access to users is relatively trivial to automate in existing systems. In contrast manually checking and removing personal information in an access copy can involves a significant cost (Beagrie et al, 2008).

The security costs incurred in providing access include putting in place a system that will ensure that access to the digital object is restricted and this is especially important to digital objects that are not publicly accessible or which are sensitive and thus needs controls.

2.5.6.3. CUSTOMER SERVICE/SUPPORT COSTS

This involves dedicated staffs who provide user support for example reference service to the designated community and they respond to the queries from the customers either through telephones, emails or in person. The costs include the staff and any CRM system that supports users in accessing the digital objects.

(Beagrie et al, 2008) argues that the demands on user support increase with the volume of users, number of user communities, and proliferation of data types, data sources, and user tools. They state that it will be important to define the levels of support at the onset as this has a direct bearing on costs and therefore can impact on the archives policies regarding supported formats etc. The capacity will increase as more automated user support aids become available (beginning with on-line documentation, FAQ, etc.). They indicate that user support may also include variable potential levels of outreach, education, and training workshops for users and users increasingly expect high-speed access to be an inherent part of online systems. Maintaining and configuring access services to consistently meet these expectations will incur higher costs particularly for large volumes of users and accesses.

2.6. OTHER FACTORS

- The number of users has a great influence on the costs of preservation. The size, knowledge base, and number of individual users and user communities will have particular influence on costs and are a significant additional factor in costs incurred by community and reference level data collections. The broader the range of researchers supported the higher the investment will be in user support. Typically large community and reference data collections will involve staff with subject knowledge of the discipline(s) to support designated user communities. (Beagrie et al., 2008).
- The time between preservation actions. The time between preservation actions is a critical cost factor. The more preservation actions, the higher the costs. In addition, more preservation actions increase the risk of affecting the authenticity and integrity of the records, and there may also be a need for additional tests. The costs can be reduced with longer periods of time between the preservation actions. However, preservation actions carried out at excessively great intervals of time can increase the risk of problems with digital preservation and the cost of preservation (National Archief, 2005).

- Volume of the records. The expected volume of the records to be stored and managed will have substantial consequences for the costs. The storage costs increase linearly with the volume. Moreover the required space will increase even more rapidly when the records need to be stored in a variety of formats (for example, the original file format and two migrated formats) (National Archief, 2005).

3.0. METHODOLOGY

This chapter discusses the research methodology adopted by the study. It specifically examines the research purpose, research approach, research strategy, data collection methods, population and finally the limitations that were encountered in the research.

3.1. RESEARCH QUESTION/ RESEARCH PURPOSE

This study shall focus in exploring and describing how cost factors can be determined empirically using the OAIS model in digital preservation. Since this area has not been widely researched on, the study is primarily exploratory and conclusions shall be drawn from what will be explored. (Yin, 1994) states that, the purpose of an exploratory research is to gather as much information as possible about a specific subject. The expected outcome of the research is a cost table that will contain cost factors identified from the research and a ranking of the individual cost factors from the research findings.

3.2. RESEARCH APPROACH

There are two main research approaches when conducting research in social science, quantitative and qualitative research. As presented by (Pickard, 2007) with reference to (Gorman & Clayton, 2005), quantitative methodology assumes the objective reality of social facts and qualitative methodology assumes social constructions of reality.

In quantitative approach results are based on numbers and statistics that are presented in figures where as the qualitative approach the focus lies on describing an event with the use of words. The approach one chooses depends on the problem definition together with the kind of information that is needed. The two methods can in cases where it is suitable be combined (Gray, 2004).

However, (Pickard, 2007) points out that whichever paradigm you associate your research with, whichever methodological approach you take, demonstrating the value of your investigation is essential and that all we want is our findings to be believed and we are responsible for ensuring that they can be believed.

This research combines both the quantitative and qualitative approach based on the research purpose and the research questions. This requires us to study the phenomenon in-depth. This approach is suitable because we want to gain a better understanding of how we can determine cost factors by using the OAIS functional model in digital preservation. The choice of the approaches was made because of the analysis of the statistical figures from the online questionnaires and the in-depth discussions from the interviews which are basically qualitative to support the findings and strengthen the validity of the results.

3.3. RESEARCH STRATEGY

In this research we shall use case study which is generally preferred when answering "how"and"why" questions about a particular topic (Yin, 1994).

> *A case study has been defined as an empirical inquiry that investigates a contemporary phenomenon in depth and within its real life context especially when the boundaries between the phenomenon and context are not clearly evident. (Yin, 2009).*

(Yin, 2009) states that a case study is preferred in examining contemporary events but when the relevant behaviours cannot be manipulated. The choice of using the case study was also motivated since we have little control over the behavioural events and the events under investigation are contemporary. Reviewing the objectives of our research all of them are stated as "how" questions which makes the selection of a case study the most appropriate one. According to Pickard (2007), the purpose of a case study is to provide a holistic account of the case and in-depth knowledge of the specific situation, through rich descriptions situated in context. This was one of the reasons for choosing on a case study because we wanted to get an in-depth knowledge and understanding of the topic under research.

(Yin, 1994) proposes four main types of case study design each of which need to be selected on the basis of a particular set of conditions. This include: single case study holistic, single case study embedded, multiple case study holistic and multiple case study embedded.

A single case embedded is where within a single study, there may be a number of different units of analysis (Yin, 1994). In this research we have chosen to conduct a single case study embedded, the case in this research includes representatives of professionals in Norway which include EDOK (part of the Norwegian Computer Society), Verdiskaperne (Information Management Network) and the LongRec project. Using this single embedded case study will make our findings more robust and possibly make it easier to detect similarities and differences since the members of the project are from different case partners.

3.4. DATA COLLECTION METHODS

When collecting primary data the researcher has specific purpose for doing it, the same applies to those collecting secondary data. In this research the primary data was collected through interviews and online questionnaires. Secondary data was also collected through documentation by gathering information from the LongRec project website, other documentations were recommendations by the interviewees for example the NOARK standard, the provision concerning transfer of electronic records to the national archives, the Digital Black hole which were recommended by the National archive service of Norway.

Interviews as a collection method are one of the most significant sources of obtaining case study information (Yin, 1994). According to (Gray, 2004) if the objective of the research for example is largely exploratory involving say the examination of feelings or attitudes then interviews may be the best approach. As they allow the researcher to probe for more detailed responses where the respondent is asked to clarify what they have said .The research is primarily exploratory and the aim of using interviews as a means of gathering in-depth information from the respondents was to probe for more information and attain highly personalized data.

Semi-structured interviews were used because they are non-standardized and are often used in qualitative analysis. The interviewer has a list of issues and questions to be covered, but may not deal with all of them in each interview. The order of questions may also change depending on what direction the interview takes. Indeed, additional questions may be asked, including some which were not anticipated at the start of the interview, as new issues arise. Responses will be documented by note taking or possibly by tape-recording the interview. The semi structured interviews allows for probing of views and opinions where it is desirable for respondents to expand on their answers. This is vital when a phenomenological approach is being taken where the objectives is to explore subjective meanings that respondents ascribe to concepts or events (Gray, 2004). The questions were open ended and few closed ended, in order to preserve the flexibility of the interview so that in-depth information could be gathered.

There are different ways of conducting interviews which include face to face interviews, telephone interview and online interviewing. Face to face interviews have always been preferred and are mostly widely used as it involves a direct meeting between the interviewee and the interviewer by personal communication. It is possible not only to obtain much more information but also one can use visual tools to encourage responses. However in some situations, one can use telephone interview especially due to geographical distance between the respondent and the interviewer. (Gray, 2004) states that the advantages of conducting telephone interviews is that the costs are low and it can be conducted faster and the interviewer can help the respondent with any misunderstanding or difficulties they have. In this research we gave respondents the options to choose from, how they preferred to be interviewed. Three face to face interviews, two telephone interviews and one online interview were conducted.

The interview was guided by an interview guide which can be accessed in the APPENDIX 2. The purpose of using an interview guide according to (Ellis, 1993, 475) as reported by (Pickard, 2007, 173) is to ensure that each interview covers basically the same ground but gives the interviewer considerable discretion in the conduct of the interview. The same interview guide was used all through the interviews. We conducted six interviews. Before conducting the interviews we contacted the individuals who are involved in the case by

mail and asked those who were willing to be interviewed. The introduction about the project had already been done by the project coordinator to all the members of the case. Those who accepted to be interviewed a list of the interview questions was sent to them to enable them prepare adequately in advance. The respondents were later contacted by mail to book the dates and the time that was convenient for them to be interviewed. The respondents were given the option to choose how they wished to be interviewed. The options included; face to face, telephone and online chatting (Skype, MSN, and yahoo messenger). The interview guide functioned as a support tool during our interviews, during the interview the respondents were allowed to talk freely with the interview guide as the base. The interviews took relatively between one to three hours.

Notes were taken during the interviews. In addition mobile phone was used to record in order to register the empirical data more accurately. (Yin, 2004) supports recording as it gives the researcher the possibility to go back and double check the received data.

In addition to interview, primary data was collected through online survey questionnaires. The questionnaire was constructed in a combination of grading scales, multiple choices, and open ended questions (Yin,2004). I provided a brief introduction of myself, the research topic and the aim of the research. This was done to enable the respondent to be aware of the research topic and the aim of the research. The questionnaire had similar questions like those in the interview only that most of the questions were closed ended. Most of the questions required that the respondent use grading scales and multiple alternatives. However, where the respondents were required to provide their own opinion they were still limited on discussing in greater depth on the different questions. The link to the survey was sent to all the members in the case by mail. The questionnaire can be accessed in APPENDIX 1.

3.5. POPULATION

Qualitative research often works with small samples of people, cases or phenomena nested in particular contexts. Hence, samples tend to be more purposive than random. Again, in contrast to more quantitative approaches, samples may not always be pre-planned in advance, but may evolve once fieldwork has begun. So an initial choice for informants may lead to a decision to select a more contrasting set of deviant subjects (cases) as a comparison (Gray, 2004).

The population included all representatives of professionals in Norway which included EDOK (part of the computer Society), Verdiskaperne (information management network) and the LongRec project. The population comprised of information- rich case for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research. The

population included members who are information professionals and the assumption was that they had knowledge on digital preservation and that the knowledge from their organizations would be of great significance in this research. The population was proposed by the key informant who is the project manager of LongRec.

The questionnaire was sent to the entire population. However, the respondents were requested to indicate if they wanted to be interviewed. In the process, some respondents indicated that they were interested in being interviewed and they in turn pointed out to other potential respondents who had knowledge on the field of research. The key informant also turned to be of great help in pointing out to information rich cases for example the members who were from the National Library and National archive of Norway contributed greatly in providing information related to the research questions. They had experience in issues of digital preservation and therefore provided valuable information.

According to (Holme & Solvang, 1991) selecting respondents with the right knowledge about the research area is crucial for qualitative research. The population provided valuable information through the questionnaires, the interviews and some of the documentations that they recommended.

3.6. LIMITATIONS OF THE STUDY

- The population for the study was relatively small because it comprised of 30 members who represented information professionals in Norway. We received 16 responses from the online questionnaire which was a very small number to draw a comprehensive conclusion regarding the research questions.
- There were only six interviews conducted , which also would affect the results of the research since most people did not want to be interviewed or did not have enough information in regard to the research questions, this had an effect on the results of the research.
- The assumption was that the population had knowledge on digital preservation and cost related issues, however it turned out that majority of the respondents did not have sufficient knowledge in regard to cost related issues in digital preservation.

4.0. DATA PRESENTATION AND ANALYSIS

In this chapter, the empirical data is presented as collected from representatives of professionals in Norway which included EDOK (part of the computer Society), Verdiskaperne (information management network) and the LongRec project. Both interviews and questionnaires are taken into consideration to bring out the complete picture of the research.

4.1. ESTABLISHING HOW ORGANISATIONS DETERMINE COST FACTORS

4.1.1. COST MODEL

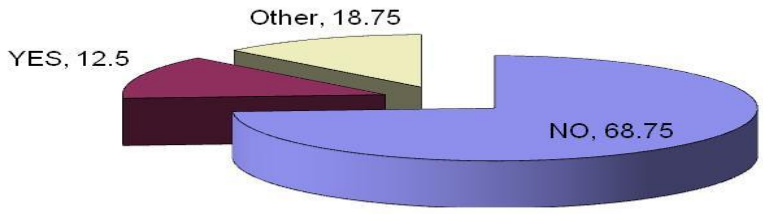


Figure 7: The pie chart is a %representation on "do you have a cost model for digital preservation? "

The response from the online questionnaire showed that 68.75% confirmed that their organisation did not have any cost model while 12.50% confirmed that they did have a cost model while 18.75% confirmed that they were not aware if any cost model existed.

One of the interviewees when asked if they had a cost model said, "I am afraid, we are not good at costs," and confirmed that they did not have any cost model for determining cost factors in digital preservation and that they received grants from the government and allocated the funds to different units and one of the units was digital preservation and the allocation of funds was based on compromise and was tradition based. This meant that the allocation of funds was done with no proper laid down policies or procedures.

The results indicated that the majority of the people indicated that their organisations did not have a cost model specifically for identifying cost factors in digital preservation. The fact that only 12.5% of the respondents indicated that they have a cost model is a critical issue as it is a threat to the long term preservation of digital objects. (Palm, 2006) points out in his article the *Digital Black Hole* that:

In the excitement about the solutions digitization offers, the right questions about costs are often not asked, especially about long-term costs for keeping the digital files alive. This enthusiastic attitude is risky, for the conversion process to create the digital files may well be quite expensive to start with, and these investments may turn out to be wasted if planning for the future is ignored and no structural funding for maintenance is secured.

This indicates the criticality with which organisations need to adopt appropriate cost models in planning for long term financial commitment towards preserving digital objects.

4.1.2. ORGANISATIONAL BUDGET

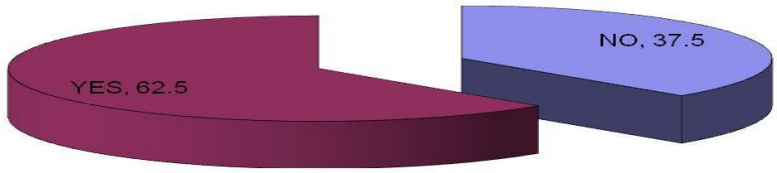


Figure 8: This is % representation on "is digital preservation included in the organisation's budget?"

62.5% respondents of the survey indicated their organisations had allocated ongoing budgeted resources for the long term preservation of digital objectives while 37.5% had not allocated ongoing funds for the sustainability of digital preservation. However, the people interviewed indicated that some of the organisations like the National archive of Norway and the National Library receives annual grants from the government and since its the responsibility of the government to ensure that these two main organisations manage the digital objects in their custody they are obliged to allocate annual funds which will ensure continuous sustainability in digital preservation.

4.1.3. ORGANISATION'S OBJECTIVES

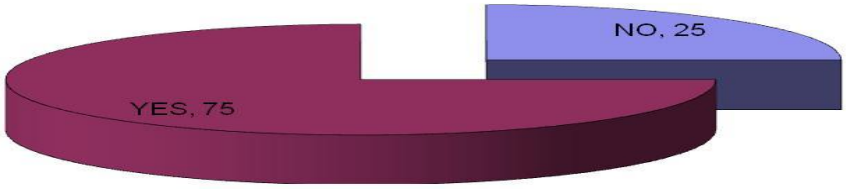


Figure 9: This is % representation on "is digital preservation part of the organization's objectives?"

75% indicated that digital preservation was included in the organisation's overall strategic objectives while 25% of the respondents indicated that the digital preservation was not included in the overall objectives of the organisation. Some of the respondents in the interview indicated that the issue of digital preservation which for example in some organisations was referred as general storage or database management was considered an IT function and presumably the funds were allocated under the IT department or unit.

One of the interviewees pointed out that digital preservation should be included in the overall organisational objectives because it is a kind of asset for the organisation.

At the national archives, Digital preservation is included in the overall organisational objectives. It is one of the core functions of the organisation and thus has got a greater placement in that it has a unit that addresses the issues of digital preservation.

The same case applies to the National library where it was pointed out that digital preservation is a priority and that the Norwegian government has recognised the Library as a cornerstone for the nation and therefore efforts are being put in place to ensure long term sustainability of the digital records.

However, others pointed out that digital preservation was not part of the overall organisation's objective rather in most organisations it was considered an IT function which involves data management and backup management. This meant that it was treated as an IT function.

4.1.4. GENERAL COST FACTORS IN DIGITAL PRESERVATION

The following general factors were highlighted by the respondent as those they take into consideration when preserving digital objects. In the questionnaire some factors were highlighted even though they never had

any cost model that guides them. The outcome of the survey from the questionnaire indicated that storage costs received a higher rating followed by technology, security, staff, legal issue received the least rating and some of the other factors that were given included organisation costs, retention costs and selection costs.

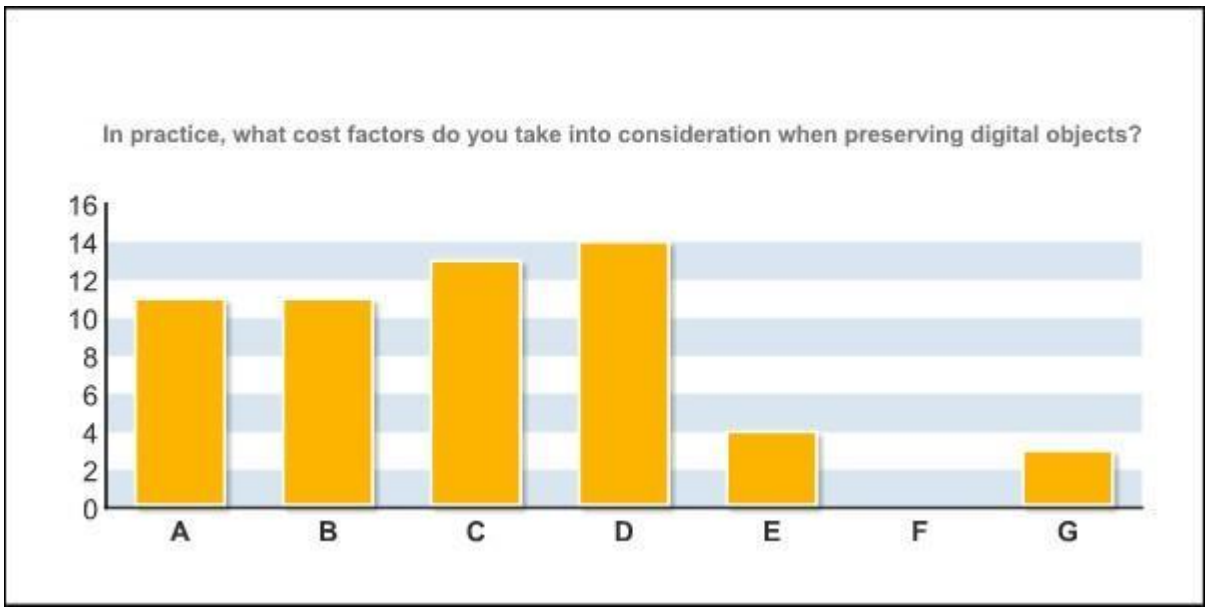


Figure 10: This is a representation on "what are the general cost factors in digital preservation?"

A) Staff costs
B) Security costs
C) Technology costs
D) Storage costs
E) Legal costs
F) Don't know
G) Other

Figure 11 below illustrates the general cost factors that were highlighted in the online questionnaire and it came out clearly that storage costs were highly regarded in terms of digital preservation. However, the interviewees gave the opinion that these factors varied from organisation to organisations. This depended on the functions of the organisation, the maturity of the organisation and the organisational sustainability.

When the interviews were conducted other factors emerged as a result of probing the interviewees and most of them were able to identify other costs that they did not consider initially. Figure 12 shows some of the cost factors that were identified from the interviews conducted as general cost factors those organisations considered in relation to digital preservation.

DIAGRAM 1: GENERAL COST FACTORS IN DIGITAL PRESERVATION

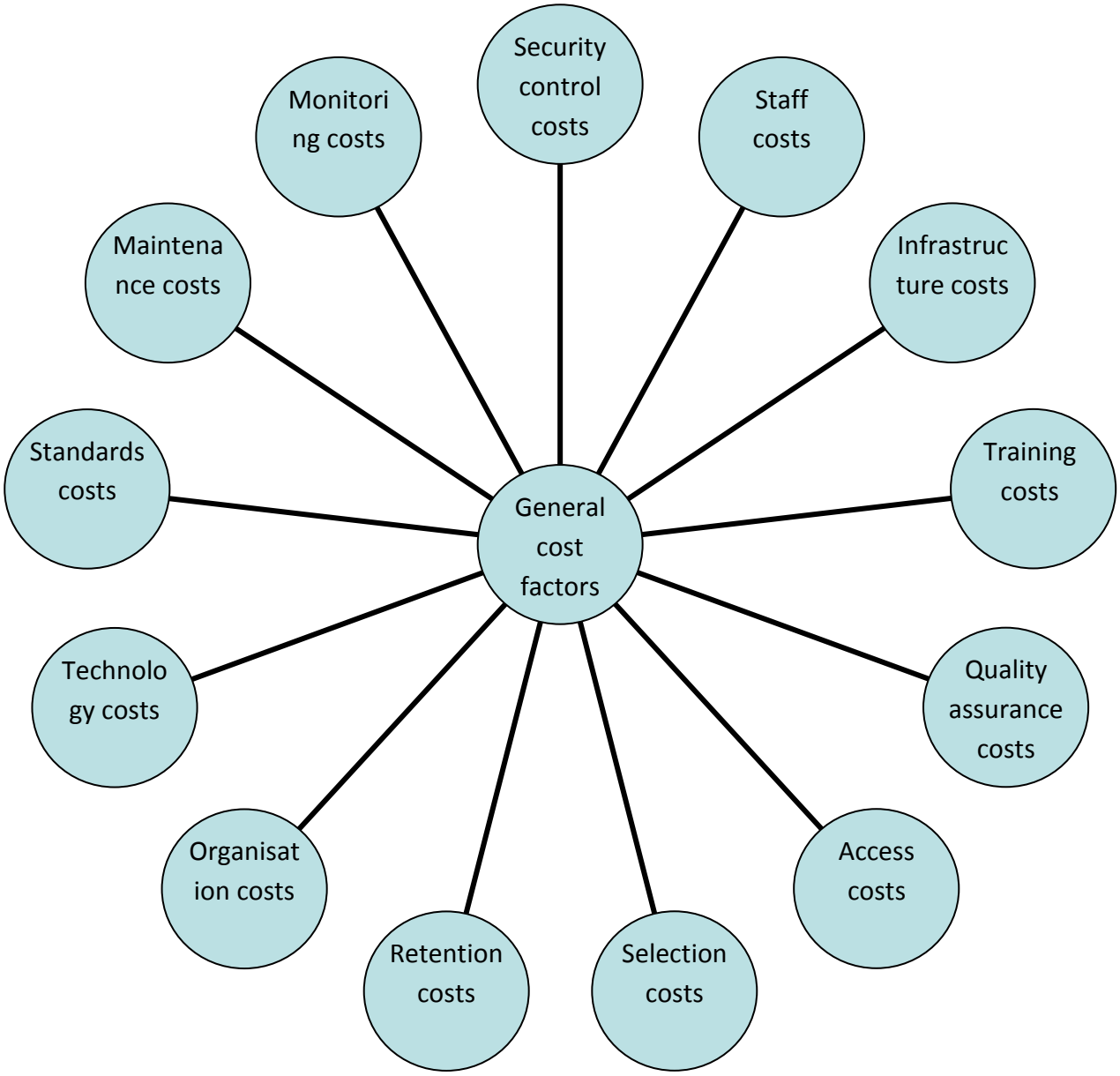


Figure 11: General cost factors highlighted from the 6 people interviewed.

4.1.5. CHALLENGES WHEN DEALING WITH COST ISSUES IN DIGITAL PRESERVATION

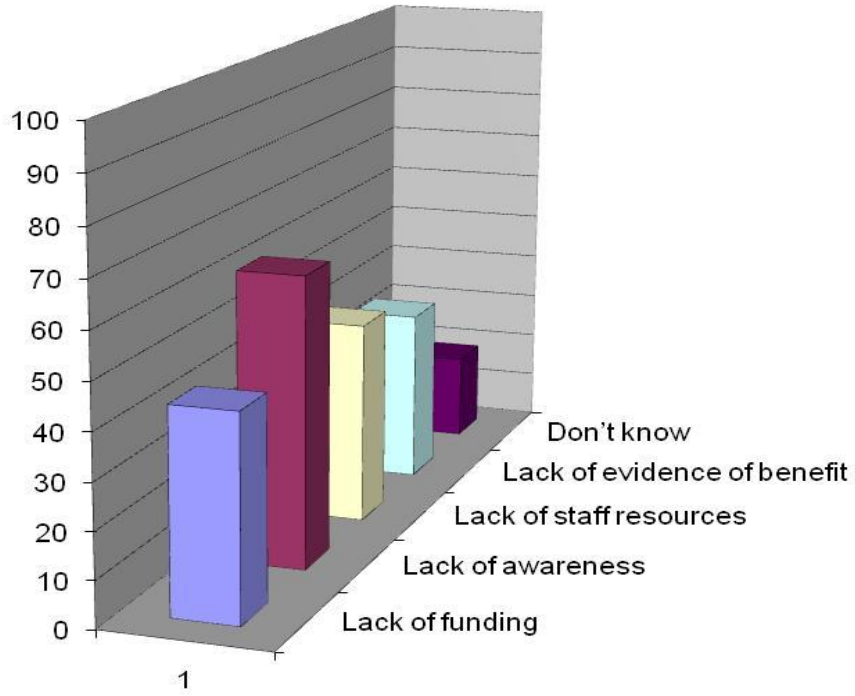


Figure 12: The figure shows the rating of challenges of costs in digital preservation.

The above challenges were highlighted in the survey as those that organisation face when determining costs in digital preservation. The survey indicated that a lack of awareness received the highest rating followed by lack of funds, lack of staff resources and lack of evidence for benefit was least rated. Other challenges that were highlighted in the survey were the corporate view and continuous focus and awareness towards costs in digital preservation

One interviewee indicated that lack of planning and lack proper policies to guide the cost issues in digital preservation were major challenges. He suggested that if proper policies are put in place it will make it easier to determine cost factors in digital preservation.

4.2. RESOURCE ALLOCATION FOR THE MANAGEMENT OF DIGITAL RECORDS

This section establishes the exact areas that organisations allocate funds to and the critical cost factors they prioritised when allocating resources.

4.2.1. AREAS WHERE ORGANISATIONS ALLOCATE FUNDS

According to the survey, the respondents rated the following areas as the ones that management allocates funds in digital preservation.

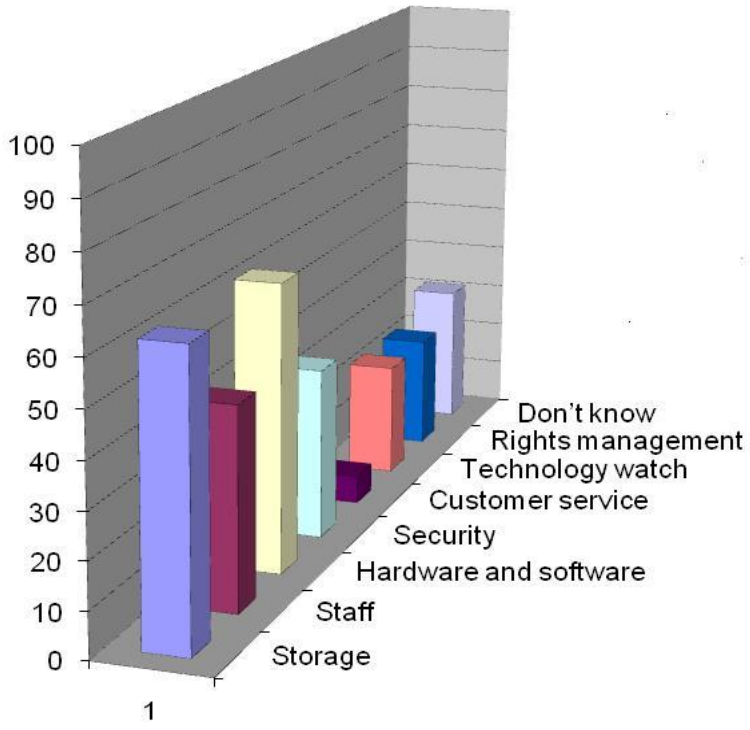


Figure 13: Rating on the response on "which areas does the management allocate resources in digital preservation?"

This shows that the management is more willing to allocate resources on storage and hardware and software in most organisations. Other areas that were highly ranked were staff and security. Customer service is not a key area where most organisations would like to allocate resources in relation to digital preservation.

The interviewees equally confirmed that most organisations allocate a lot of funds on storage and technology. Another external yet related example is that the National Archive of Norway is currently working on acquiring an online digital storage called SAN (storage area network). This means that much of the funds are allocated towards testing of this system and the final implementation.

Other costs that were highlighted included equipment such as disks tape robots and other hardware and software for the system. For example, the National archives allocate funds depending on the project that they have planned and they pointed out that they incurred a lot of costs in developing standards that is the NOARK standards which is a specification of functional requirements for electronic recordkeeping systems used in public administration in Norway. They incurred a lot of costs on the consultancies because they outsourced the services and they paid salaries for five people for three years.

They stated also that once that project was done the costs for standards went down, however they indicated that staff is not a cost because the staff working in the national archive of Norway are employed by the government and so the salaries paid to them are not apart of the funds allocated to them.

4.2.2. THE CRITICALITY OF COST FACTORS IN DIGITAL PRESERVATION

	Irrelevant	Moderately important	Important	Very important	Crucial
Selection costs			√		
Quality assurance costs			√		
Metadata costs				√	
File format conversion costs				√	
Storage costs					√
Security costs				√	
Migration costs				√	
Emulation costs			√		
Error checking costs			√		
Validation costs			√		
Backup costs					√
Disaster preparedness					√
Training costs			√		
Software development costs			√		
Hardware costs			√		
Staff costs			√		
Intellectual property rights agreement costs	√				
Technology watch costs		√			
Outsourcing costs	√				
Interface design		√			
Customer service/support costs			√		

Figure 14: The rating on the criticality of cost factors in digital preservation.

Figure 14 above shows the criticality of the various cost factors as reported in the survey. This is in relation to the areas that organisations would like to be funded in case the resources were limited or areas they thought were critical in their organisation when it came to funding for digital preservation. Figure 14 shows

how the cost factors were rated by most respondents in a scale line of relevant to crucial. Storage, backups and disaster preparedness costs were rated as the most crucial while Intellectual property rights agreement costs and Outsourcing costs were rated as irrelevant.

4.3. USING THE OAIS MODEL FOR DETERMINING DIGITAL PRESERVATION COSTS.

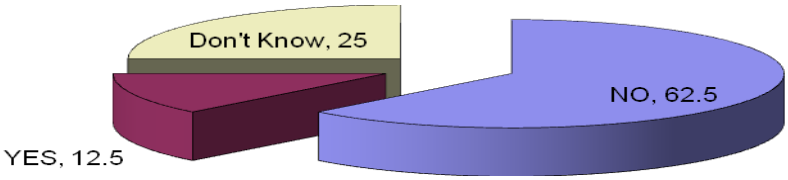


Figure 15:% representation of the online questionnaires on the use of OAIS model in determining costs

From the findings only 12.50% confirmed that they were using OAIS model in determining costs, 62.50% were not using it and 25.00% did not know what OAIS model was. Among the interviewees who knew the OAIS model, one said that the model had both the human and automated functions. People had three roles to play as producers, management and consumer. The other functions should be automated. However, other interviewees confirmed that they were in the process of working out ways in which it could be fully implemented. For example, the National archive acknowledged that they were aware of the OAIS model but were not using it for determining the costs however after discussing with them they were able to acknowledge that it was a good model that could be relevant in identifying cost factors.

4.4. THE RESULTS FOR THE FUNCTIONAL AREAS OF THE OAIS MODEL

This section discusses the various cost factors focusing on the functional areas of the OAIS model which include: Ingest, Archival storage, Data management, Administration, Preservation Planning and Access. It reports the findings of the results as collected through the online questionnaires and the interviews conducted.

4.4.1. INGEST

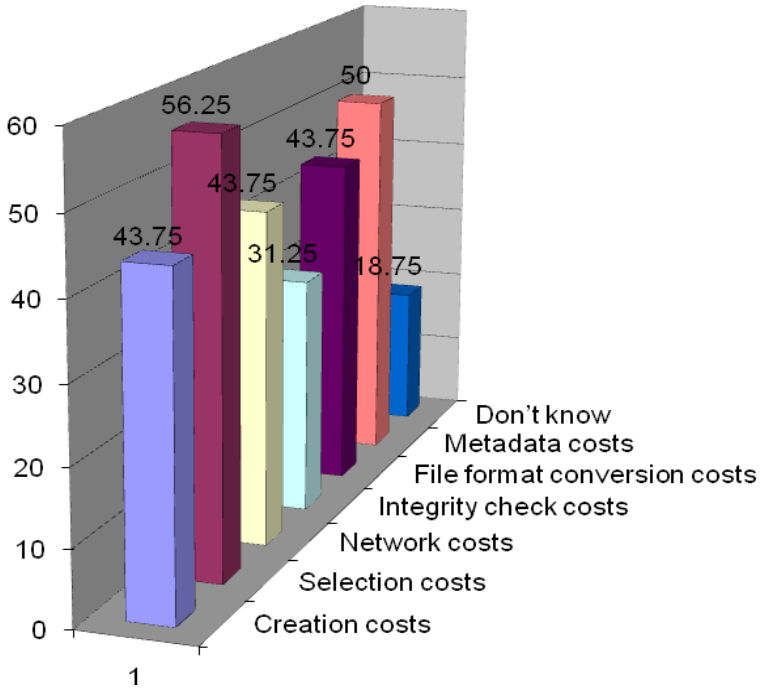


Figure 16: Rating of cost factors in ingest. "What are the cost factors that can be identified in the ingest?"

Figure 16 diagram shows the response of the respondents in the survey in regard to the various cost factors in the ingest function of OAIS model.

The findings of the interviews indicated that the **costs involved in the creation** of digital records is that the staff has to go the organisations to offer advise to the organisation on the best practices in managing records and how to correct mistakes and bad practices during creation of digital records. However, some interviewees stated that the costs normally were incorporated in salaries of the staff as it was part of the job description.

Some interviewees gave recommendation that organisation should develop policies in line with creation of digital objects so as to reduce the costs. The costs which may be incurred will be initial costs of developing the policies or procedures.

Selection costs were also highlighted during the interviews and it received the highest rating in the survey. The interviewees from the National Archives of Norway stated that they have a policy which is a provision concerning electronic records to be transferred to the national archives and that the government ministries and other central offices must adhere to the rules in the provision for them to be approved, so the selection of the digital documents to be archived depends on the policies in the provision.

The national archive does thorough selection which involves the costs for the staff that do the selection. The archival storage media approved according to the provision include CD-R, DVD-R or DVD+R, Hard disk with USB2.0-connector, and memory stick with USB2.0-connector.

Network costs: this is applicable especially where producers transmit the SIPs online. However, from the interviews, the costs of transmitting the SIPs are normally incurred by the producers this is especially in the archives where the producers should submit an archival version for long term preservation However, the interviewees recommended that there should be a good network system but this depends on the type of organisation whether it is commercial or non commercial. In the case where it is a commercial entity an interviewee said that the commercial organisation offering the long term preservation services should provide good network for transmission of the digital records online however this costs are later transferred to the designated community. The interviewees also recommended that the network system should ensure security of the SIPs that are transmitted through. The suggestion was that the SIPs can be transmitted in encrypted formats or as agreed upon by the people responsible.

The costs incurred in quality checks were considered to be enormous as they involved a lot of testing which was both done by the staff and also system checks to ensure that they met all the requirement, this means that the more the testing the higher the costs incurred in terms of time and the systems in place to do the checks. However, one interviewee indicated that long term preservation has not been considered a cost since a lot of testing is done to the records to ensure that they can be kept as long as they are needed and thus higher costs are incurred in quality checking of the digital records to be kept for long term preservation.

File format conversion costs: this was ranked high because the costs involved in handling the various file formats is relatively high and therefore they try to convert the records into standard formats that can be preserved for a long time. They pointed out that the costs incurred include the manpower and the hardware and software for the conversions However, the interviewees indicated that the digital objects especially the ones received at the archive have to follow set standards and that they only allow limited number of file formats which they are able to support and preserve.

4.4.2. DATA MANAGEMENT

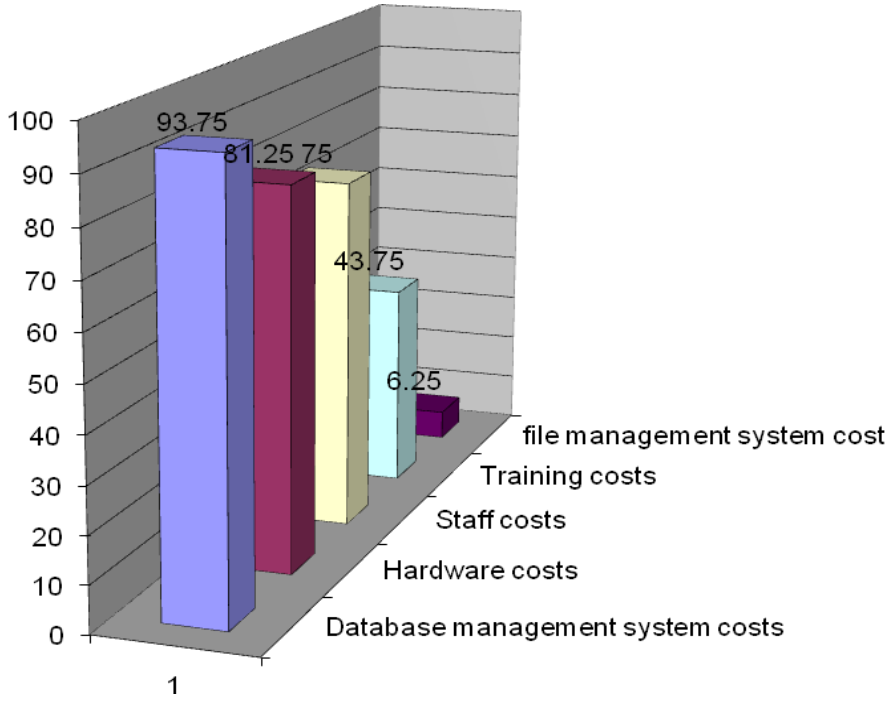


Figure 17: Rating on cost factors. "What are the cost factors that can be identified in the Data management?"

The main cost factor in data management that was pointed out was those incurred in acquiring a **Database Management Syste**m (DBMS). However, the interviewees clearly pointed out that the kind of system adopted varied from one organisation to the other and this depended on so many factors like the functions, stability and the type of organisation. This has effect on the cost for example organisations like libraries and archives will require DBMS systems that are more costly than small organisations which are not stable which may incur DBMS costs.

The hardware that supports the data management was highlighted as cost factors in digital preservation. This included the real costs for the hardware and the costs for the support, maintenance and the operation of the systems.

Staff is key in the operation of the data management system. The costs that were highlighted included staff with skills on data management for example a systems administrator, database managers and Database administrators. However, most of the interviewees indicated that this function was done by IT staff in most of the organisations and that almost all the organisations have IT departments that handles and manages the data for the organisation. This meant that as indicated earlier most organizations assumed that Digital preservation is a function of IT in their respective organisations. This is a cost factor that organisation needs to take into consideration when planning for long term preservation of digital objects. According to a study that was conducted in the National Archives of Sweden, they indicated that:

When the cost are divided into technology, staff and premises it turns out that the cost of labour accounts for 39% of the total. This will increase in the years to come, as salaries will go up and as more staff will be needed to manage the system as it grows. If one attempts the impossible, maybe even the ridiculous, and tries to make forecasts beyond a period of 10 years, the only thing which is certain is that salaries will escalate as well as the general cost index (Palm, 2006).

This clearly indicates that the staff costs will increase with time and proper plans have to be put in place in terms of planning for the long term costs of salaries for staff especially when it involves staff with special skills in particular areas like Database management etc.

Training costs: funds used to train the staff to improve their skills on specific areas in data management. One reason that the interviewees pointed out for the need of training was the continuous change in technology. This meant that, for the staff to be aware of the changes in technologies and to be at par with the current trends in terms of technology there was a need for them to receive continuous training which means that organisation are obliged to plan and set aside funds for the continuous training and upgrading of staff skills.

4.4.3. ARCHIVAL STORAGE

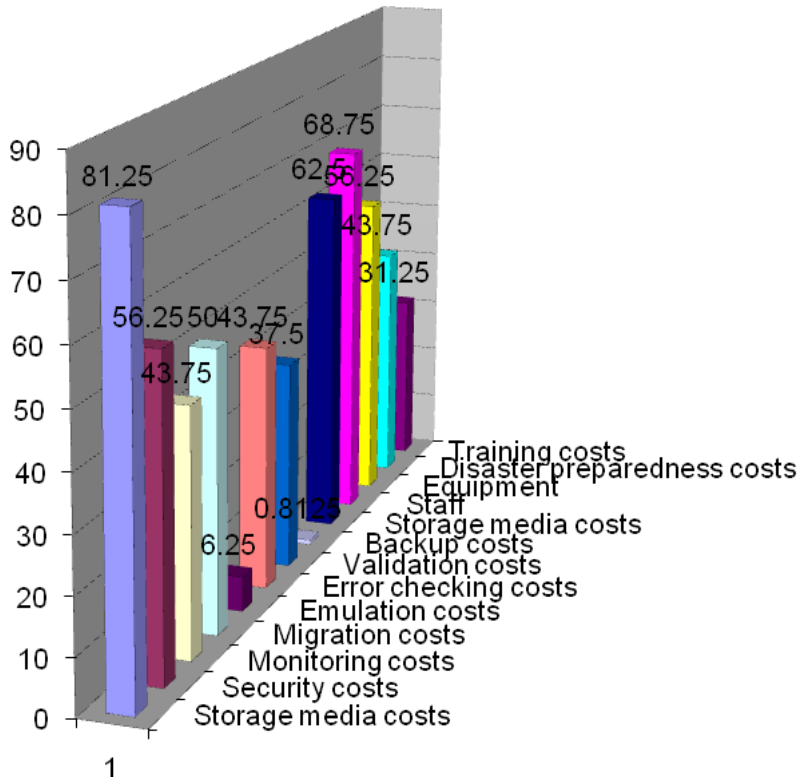


Figure 18: Rating on cost factor. "What are the cost factors that can be identified in the archival storage?"

Storage media costs were ranked the highest in terms of costs incurred in the archival storage. However, the interviewees pointed out that the storage media costs varied depending on organisations maturity, sustainability and technology. For example, some organisations are technologically advanced and are using advanced systems in archival storage e.g. the SAN (Storage Area Network) system which the National Library of Norway is using. These are advanced systems that support all the functionalities of the archival storage like backups, security, validation, and error checking. The costs of acquiring these systems are higher compared to those that are small in size and may be using CDs or Tapes as storage media.

However, all interviewees acknowledged that the storage media was a cost factor that had to be taken into consideration in long term preservation of digital objects. One interviewee pointed that the costs for storage media can be reduced in future by using tapes with larger storage capacity which is still debatable depending on several other factors. However, other cost factors like hardware, software, support, maintenance should be taken into consideration alongside the storage media costs

Staff was a cost factor that was highlighted in the archival storage. According to one interviewee, this should include staff with expertise of running the systems and those involved in preparing the AIPs and taking backups. The staff should have skills in handling the AIPs to ensure that they are well kept, especially where it is done manually, the staff should be involved in checking the AIPs for errors and other quality checks. This however can be an automated function but still requires staff to monitor and carry out any operations related to the archival storage function.

Migration and emulation are inevitable costs that digital objects have to undergo to allow for future access. This is done through constant process of refreshment and migration and this requires long term financial commitment. The interviewees pointed out that in most organisations in Norway it takes 3- 5 years before migration is done. This requires proper financial planning and budgeting to ensure that the digital objects are migrated successful through the various stages of their lifecycle. One interviewee pointed out that the greatest cost that they incurred was the continuous testing of the systems which basically involves manpower costs in this case experts and a lot of time. However, it was pointed out that most organisations have not adopted emulation because of the uncertainty of the future though its one of the probable approach that from the literature shows that it can be cheaper than migration since the initial costs of emulation are high but subsequent costs tend to decrease.

The interviewees indicated that **backups** were cost factors in digital preservation and that they varied depending on the level of technology of the organisation for example some organisations as mentioned earlier adopted systems that were complete and supported backups through mirrors and virtualization by automatic up scaling, this means that the costs in backups were part of the system. However, for smaller

organisations which have to do backups manually, it involves taking into consideration the costs involved in undertaking the backups. Other costs associated with backups as pointed out by the interviewees included **offsite storage costs w**hich some said they incurred the costs as they had to deposit some of the backups with other organisation that offer such services.

Disaster preparedness is a cost that an organisation should not ignore, they should conduct training for the staff on how to manage disasters in case they occur. This will help the organisation in managing risk in the future. One interviewee indicated that most organisations do not prepare for disasters, however proper measures should be put in place including disaster preparedness policy. Also, the equipments, hardware, software and other physical requirements should be put in place for example fire extinguishers etc. These are costs that should be taken into consideration to ensure that in case of a disaster the digital objects can be secure.

Monitoring, error checking, security and validation costs can be viewed from two angles, one way is if the system is automated then this functionally are automatically integrated within the system, however if the functions have to be done manually this will involve staff and tools that support the functions . These are meant to ensure that the AIPs are correct and maintain the integrity of the AIPs in their custody. This can incur a lot of costs especially where they are done manually, so it is advisable to automate the functions as a way of reducing costs

4.4.4. ADMINISTRATION COSTS

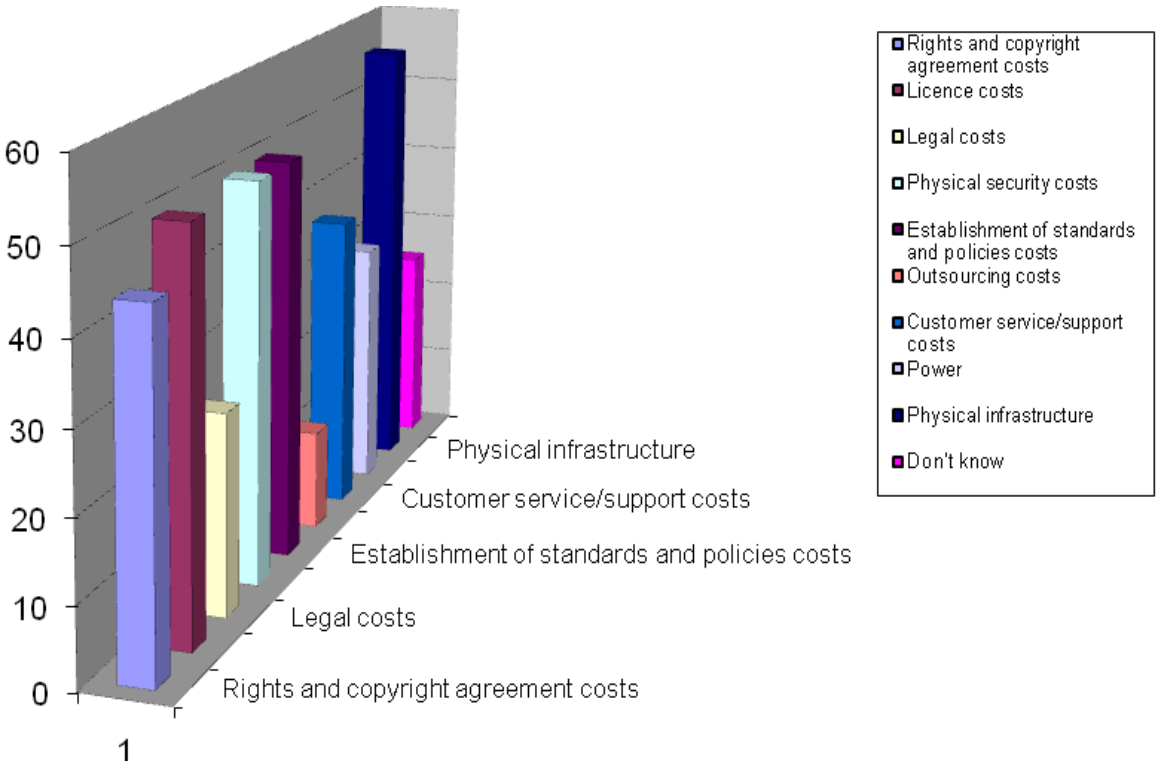


Figure 19: Rating of cost factors. "What are the cost factors that can be identified in the administration?"

Rights and copyright agreement costs were considered to be on high levels this was between the producers and the bodies handling the digital records. The negotiation here according to one interviewee was on the metadata that should be included on the digital records and that this depended on the clients. If they are commercial based then costs will be incurred however non commercial there were no costs incurred for example the National Archive of Norway never incurred any costs in negotiating for rights of submission or custody. However, in the National Library one interviewee pointed that that they incurred copyright costs in giving access to some digital objects, he gave an example that they needed to pay a certain amount of money per page for some eBooks.

Physical access control involves the security issues by providing a mechanism that restricts or allows physical access to the elements of the archives. The costs that were pointed included building, strong locks security access system, developing access cards, hiring an agency to monitor the security of the archive. One interviewee pointed out that a security agency can be hired which monitors the archive and triggers an alarm

in case of any unauthorized access. The costs vary depending on the level of security which an organisation is willing to adopt.

Standards and best practices can be a cost saving issue, however the initial cost of developing them which may include costs of hiring experts who are paid to develop these standards may be high but subsequently it goes down. For example one interviewee indicated that the National archive of Norway incurred a lot of costs in developing standards and this costs included consultancy costs and salary for five people for a period of three years.(ERPA,2003) points out that the use of standards help to reduce costs and that format standards help in reducing the maintenance and ease migration procedures, the same applies to standardization in metadata enables interoperability which enables easy and efficient information exchange interaction .

Customer service support: administration function is the hub of the other functions of the OAIS model and thus customer service is very central in the management of the archive. The costs that were pointed out by the interviewees included staff costs, customer relationship management system to manage the details of the customers, training costs for the staff and any hardware or software that would support in the administration of the archive. Customer service from the perspective of the OAIS model is a function that manages the consumer accounts and it involves collecting billing information and sending bills and receiving payments. This perspective is based on commercial archives however for non commercial archives the functions of customer service may tend to differ as consumers or customers generally do not pay for the service of accessing the archives.

Physical infrastructure costs include the power supply, cooling system, physical housing, security, fire protection and cabling. One of the interviewee indicated that they received a grant from the government and used 50% to pay for rent and this included the physical infrastructure of the whole building. It was a major cost factor although they did not consider it purely on preservation but it was a cost especially taking into consideration the special rooms that were conditioned to provide the required conditions for long term preservation of materials in their custody. They also include costs of maintaining the building, repairs and ensuring that the physical infrastructure is in good conditions.

4.4.5. PRESERVATION PLANNING

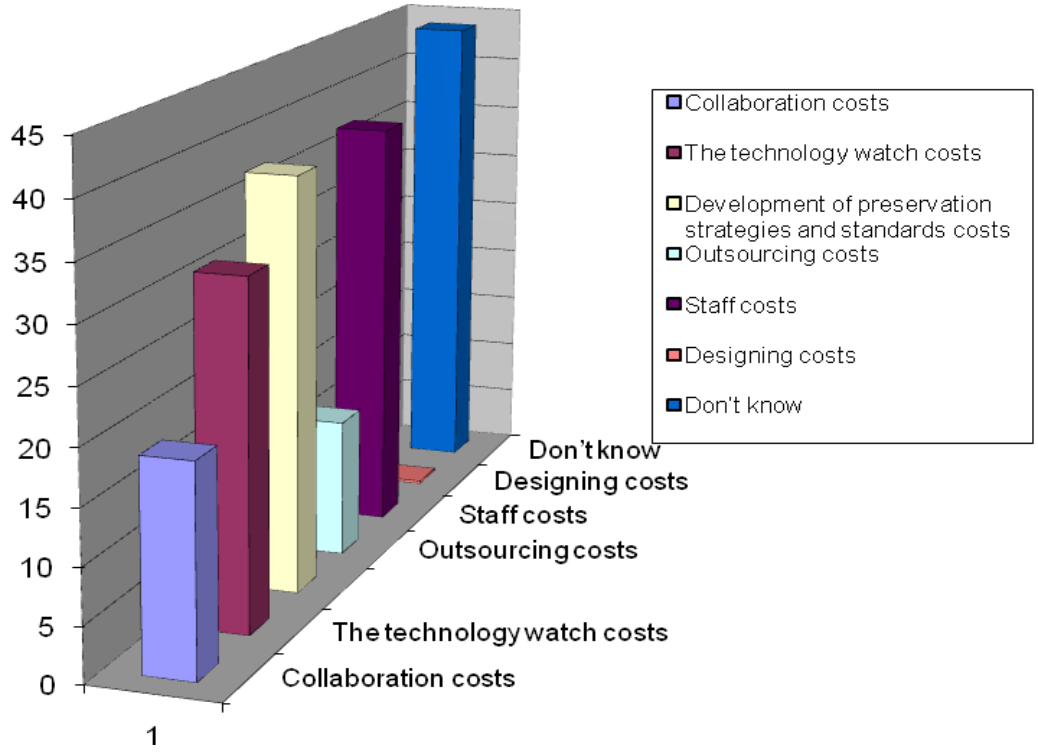


Figure 20: Rating on cost factors in preservation planning. "What are the cost factors that can be identified in preservation planning?"

Collaboration costs: These costs included those involved in organising surveys, workshops and other formal review forums. They indicated that specific costs vary depending on the interest of the people involved but the aim is to track the changes and need of the designated community. These costs are shared among the people or organisations involved.

Technology watch involves monitoring the technology and the designated community who in this case vary from one organisation to another. The costs that were highlighted were staff costs involved in the technology watch and the monitoring of online tools. One interviewee pointed out that they were free and they never incurred any cost in using the online monitoring tools.

Developing preservation strategies and standards: the main cost highlighted was the salaries for the experts who develop the preservation strategies. The interviews indicated that most organisations employ experts to develop the preservation strategies and the standards. The costs can be high but it is not a recurring cost and therefore once the strategies and standards are in place, then the costs goes down. This might also be done by outsourcing the services for example if an organisation find it expensive to higher experts or the staff don't have skills and knowledge on developing preservation strategies and standards they

can outsource the service . This however varies, at times it can be cheaper, and the administration has to weigh the option in order to make the best decision.

Staff should be involved in the process of developing strategies for preservation and standards as they are the ones who will implement them. The interviewees indicated that normally the cost of staff is part of the salaries, however if the staff are involved in areas that are outside their job description then extra costs will be incurred on them as they are part of the team involved in preservation planning. The costs of designing packages and migration plans are also linked to the expertise and the staff involved.

4.4.6. ACCESS

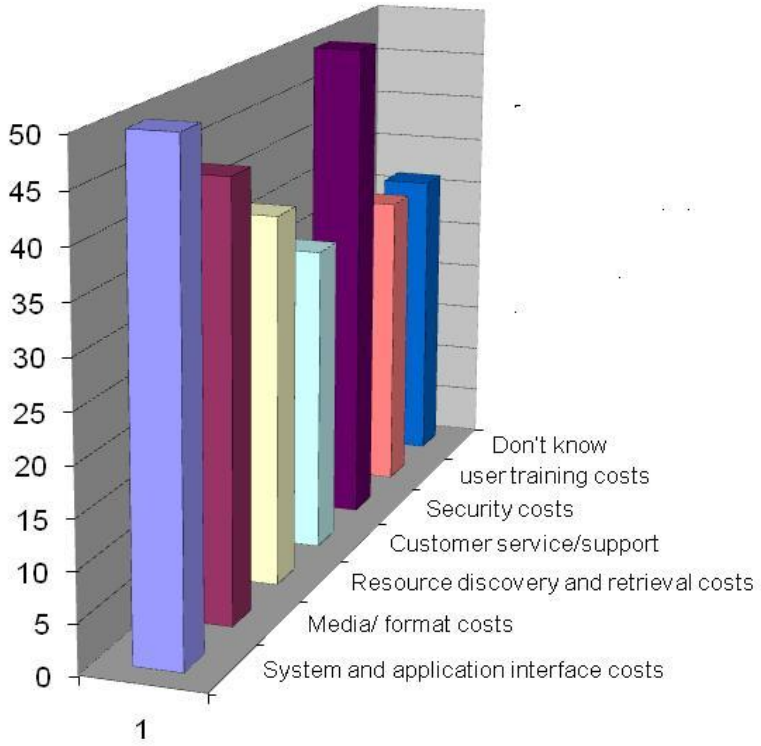


Figure 21: Rating on cost factors in access. "What are the cost factors that can be identified in access?"

System and application interface: the cost of designing interfaces that allow the designated community to have access to the contents of the archive especially in case where access is provided online was highlighted as one cost factor. This will require expertise which is a major cost as they are involved in designing interfaces that support user needs. One interviewee pointed out that the experts involved in the designing of the interfaces should have relevant skills and knowledge e.g. knowledge on information architecture and designers to be able to design interfaces that will enable the designated community to access the DIPs. The costs will vary depending on the level of expertise that an organisation acquires.

Media / format costs, the consumer request may be an order which may require the archival storage to prepare a formal dissemination information package (DIP) for online or off line delivery. This will require that the DIP is delivered in a certain format for example if it is off line delivery it could be in form of CD or DVD. The interviewees stated that organizations should be aware of the media costs involved in providing access to DIPs.

Resource discovery and retrieval costs: this may include search engines that allow for users to locate the resources within the archive. The catalogues and indexes require to be developed and the costs include staff with special skills to develop them. In advanced cases they may be automated systems that allows for online searching or one which has got inbuilt functions that support resource discovery and access.

Security costs: this involves limiting access to the users. This can be done by providing a system that controls access by users for example providing users with usernames and passwords to control access to the content of the archive. The various digital objects may have different levels of access other archives have laws that restrict access of certain digital objects for a certain period of time in such cases users are not allowed access even some staff members may not have access to those digital objects. The security costs will include systems that support such securities, this were pointed out by the interviewees that they vary from one organisation to another depending on the level of security they are willing to adopt and the sensitivity of the digital records in their custody.

User training: this can be done to ensure that users can be able to access the contents of the archive on their own. The costs of training the users may be high at initial stages but may save on costs. This is because staff would not be required to support users and this will reduce on the overall costs of supporting users. One interviewee indicated that specific costs would likely be the cost of staff involved in training users, the documentation and the time taken to train the users.

5. 0 SIMPLE COST FACTOR MODEL

This chapter presents the matrix for the cost factors from the analysed data with an aim of coming up with a simple cost model that reflects on data collected from both the interviews and the questionnaires. The simple model is basically mapped into the functions of the OAIS reference model. This result in a simple formula for tentatively calculating the costs of digital preservation based on the analysed data over a given period of time.

5.1. MATRIX FOR THE COST FACTORS IN DIGITAL PRESERVATION

This matrix presents the cost factors identified from decomposing the OAIS model and the relative cost ratings received from the interviews and the questionnaires. This gives a clear and simplified picture of the significance that each cost factor was accorded. This matrix is very significant to organisations who would wish to identify the specific costs that apply to their individual situations.

OAIS FUNCTIONAL ENTITIES	OAIS FUNCTIONS	COST FACTORS	CONSIDERATIONS	RELATIVE COST RATING BASED ON THE RESULTS FROM THE QUESTIONNAIRE
Ingest	Receive submission	Creation costs	Decisions made when a digital object is being created have a significant effect on the cost and less influence means mostly higher costs. The adoption of best practices during creation can help simplify the task of preserving the digital objects and this will lead to reduction in costs. The costs involved are • The costs of promoting good practices • The costs of correcting mistakes and bad practices during creation	important
		Selection costs	Selection of the digital objects during	important

			submission enables assessment of future costs associated with each new submission. The cleanup costs are normally higher if selection is not done during submission of the digital objects.	
		Networks costs	A good network system that allows successful submission of the SIPs. The cost of the network services should include access, authentication, confidentiality, integrity controls and management of communication between senders and receivers of information in a network.	Moderately important
		Evaluation costs	The cost level involved in evaluation will depend on the size and complexity of the digital resource and how well documented it is. Automation of evaluation process will decrease the costs involved.	Important
	Quality assurance	Integrity check costs	The data integrity service ensures that data is not altered or destroyed in an authorized manner and it applies to data in permanent data stores and to data in communications messages. The costs are incurred in acquiring a system that can monitor and check the integrity of the digital objects i.e. the use of Cyclic Redundancy Checks ,checksums or system logs	Important

	Generate AIP	File format conversion costs	The growing complexity of digital objects has led to increase in file formats. The challenge is how to handle the various formats. Use of open formats which have been well documented, have undergone thorough testing and are non-proprietary and usable on different hardware and software platforms minimises the frequency of migration and reduces the risk and costs in their preservation	Very important
		Documentation costs	Organisations should put more resources into promoting good practice to depositors as this will better the documentation and hence lower the costs associated with reading, amending and managing the documentation. However, the costs will increase if the documentation is done manually. Therefore automation is required to save on costs.	Important
	Generate descriptive information	Metadata costs	The more the descriptive information the higher the cost and this should be linked with the value of the records. The costs can be reduced by limiting the file formats, adopting tools for metadata extraction and reuse of metadata from other sources. The costs include staff and the time taken in generating the metadata.	Very Important

Archival storage	Receive data	Storage media costs	The cost of media storage will include the costs of acquiring the storage media. Appreciable costs can be incurred in the purchase of storage media and that organisations have to make sure an adequate amount is budgeted for effective storage and backup media.	Crucial
	Manage storage hierarchy	Security costs	The security costs that ensures appropriate level of protection for the AIP i.e. on-line, off-line or near-line storage.	Very important
		Monitoring costs	The costs involved in monitoring error logs to ensure that the AIPs are not corrupted during transfers. The monitoring systems can be automated to reduce costs or if done manually the costs are likely to go up.	Important
	Replace media	Migration costs – Refreshment, Replication, Repackaging and/or Transformation.	Each type will have a different costing and it may be necessary to employ multiple preservation methods in parallel. The reason may be the heterogeneous collection of different object types or the necessity to enhance the probability of success in preservation.	Very important
		Emulation costs	Emulation requires more initial investments, which makes it inappropriate for short-term preservation. For a proper cost comparison, the costs of each strategy should be specified in relation to the term for preservation.	Relatively important

	Error checking	Error checking costs	Costs incurred in putting in place a system that monitors the errors and provide notification of potential errors.	Important
		Validation costs	The cost of validation of the completeness of the object on delivery will include the time taken to obtain any necessary documentation, and the time spent checking the object received against the related documentation. This will largely be a technical role but one which could probably be carried out at a junior level once policies and strategies are in place. In time this is also something that can be built into the routines of the archive administration and might be done increasingly by automated means.	Important
	Disaster recovery	Backup costs	The costs of backups which should be stored offsite. The costs of testing regularly the backups to ensure that the data is still readable and has not been altered. Costs incurred in the purchase of additional licenses or obtain special permission from the software vendor in order to back up applications.	Crucial
		Disaster preparedness costs	This are costs incurred in ensuring that there is a plan in place in case of any disaster and will require the development of a disaster preparedness program.	Crucial

		Training costs for the staff on disaster preparedness	The costs involved in training the staff on disaster preparedness.	Very important
Data management	Administer Database	DBMS costs (a software for managing database)	The decision on the DBMS to acquire depends on the organisation and the decision of the administration and costs involved.	Very important
		Hardware costs for the management of the data	Costs vary and the decision depends on the organisation's decision. This can depends on advise from vendors	Very important
		Staff costs – systems administrator / database manager	They require specific skills. Well trained, skilful and experienced people are expensive but will reduce overall costs.	Crucial
		Training costs for the other staff on the use of the DBMS.	Training needs to be up-to-date and adequate depending on the job-function. This requires funding.	Important
Administration	Negotiate submission agreement	Rights and copyright agreement costs	This represents ongoing costs, it may cost more at earlier stages until processes and workflows are in place to deal with it on a day today basis.	Irrelevant
		Licence costs	Licence fees charged for permission to preserve the digital object, and to allow continued access to it.	Relatively important
		Legal costs	Costs involved in negotiating for the submission of the digital objects.	Irrelevant
	Physical access control	Physical security costs (doors, locks and physical infrastructure	This service prevents unauthorized access to information system resources and also prevents the use of a resource in an	Very important

		that ensures servers are secure)	unauthorized way. This service may be applied to various aspects of access to a resource (e.g., access to communications to the resource, the reading, writing, or deletion of an information/data resource, the execution of a processing resource) or to all accesses to a resource.	
		Staff(guards)	The costs of the staff in charge of the security of the digital objects.	
	Establishment of standards and policies	The costs of establishing standards and policies i.e. time , staff, legal costs	Standards help reduce costs. Policies enable consistent and cost effective management	Important
		Outsourcing costs (sourcing the external services to help in developing the policies)	The costs may be higher initially but it will help reduce the costs	Irrelevant
	Customer service	The Customer Relationship Management System costs (CRM)	The costs involved in acquiring a CRM system for managing customer accounts.	Important
		Staff costs	The costs involved in hiring staff to manage the system	Crucial
		Training costs and skill upgrading costs	The costs of training the staff and upgrading their skills.	
Preservation planning	The monitor designated community	Monitoring costs through surveys, community	The aim is to get information on emerging standards and requirement alerts to develop preservation strategies and standard functions	Important

		workshops.	Co-operation with other organisations may lower the costs through synergies and economies of scale.	
	Monitor technology	The technology watch costs	Implementing an effective technology watch or IS strategy to avoid potential loss of access to the digital holding as a result of obsolescence and tracking emerging digital technologies	Relatively important
	Develop preservation strategies and standards	The costs of developing the preservation strategies and standards	The actual costs of developing the standards and the preservation strategies.	Important
		Outsourcing costs	In case the staffs are lack competency. This can be costly at initial stage.	Irrelevant
		Staff costs	The staff involved in the developing of the strategies and standards.	Crucial
	Development of packaging designs and migration plans	Designing costs	It will involve experts in the field who will design the packaging and migration plans	Important
Access	Coordinate access activities	System and application interface	The costs designing system and user interface that allow users to have access to the digital objects.	Very important
		Media format costs	The media costs involved in providing access of the DIPs to the designated community e.g. CDs and DVDs etc.	Relatively important
		Customer support /service	The costs include staff costs, customer relationship management system to manage	Very important

		the details of the customers, training costs for the staff and any hardware or software that would support in the administration of the archive.	
	Findings aids costs	The costs associated with developing and making available online search resource discovery tools i.e. online catalogues.	Very important
	User training	The costs involved in training the users to access the digital objects. These include the staff costs and the time required.	Important

5.2. THE SPECIFIC COST FACTORS FROM THE RESEARCH FINDINGS

The tentative formula presented below shows in a simplified way how organisations can calculate the costs of digital preservation over a given period of time. The assumption is that different organisations have different costs factors which apply to their organisations. The formula enables them to select the cost factors that directly have influence in their organisations and then calculating the exact costs based on the exact values at that particular point in time. This therefore serves as an awareness tool for calculating the digital preservation costs.

Total Preservation costs =∑(

Ingest costs

- Creation costs
- Selection costs
- Network costs
- Integrity check costs
- File format conversion costs
- Metadata costs

Archival storage costs

- Storage media costs
- Security costs
- Monitoring costs
- Migration costs
- Emulation costs
- Error checking costs
- Validation costs
- Backup costs
- Staff costs
- Equipment costs
- Disaster preparedness costs
- Training costs

Data management costs

- Database Management system costs
- Hardware costs
- Staff costs
- Training costs

Administration costs

- Rights and copyright agreement costs
- Licence costs
- Legal costs
- Physical security costs
- Establishment of standards and policies costs
- Outsourcing costs
- Customer support/ service costs
- Power costs
- Physical infrastructure costs

Preservation planning costs

- Collaboration costs
- Technology watch costs
- Development of preservation strategies and standards costs
- Outsourcing costs
- Staff costs

Access costs

- System and application interface costs
- Media and format costs
- Resource discovery and retrieval costs
- Customer service/suppo rt
- Security costs
- User training costs

6.0. CONCLUSION AND FURTHER RESEARCH

6.1. CONCLUSION

This chapter is a summary of the key findings drawn from the online questionnaires and the interviews. It focuses on the main issues learnt from the study, further details can be studied from the questionnaire findings. This has been done by restating each of the research objective posed in chapter one and trying to answer them in a summarised form. This is followed by a table that tries to represent the cost factors in economical terms and lastly a recommendation for further research.

This study examines the cost factors in preserving digital records in organisations using the Open Archival Information System (OAIS) model. The model helps breaks down the preservation process into smaller sub-processes which could then be assessed more easily with regard to their required resources and the associated costs factors. Identifying the correct cost factors will allow organisations to budget better for the preservation of digital records. The study adopted a case study with interviews, online questionnaires and documentation being used to collect data. The study population was the representative of groups of professionals in Norway which include the EDOK (part of the Norwegian computer society), Verdiskaperne (information management network) and the LongRec project in Norway.

Below is a review of the research objectives and the summary of the key findings:

1. **To establish how organisations identify cost factors in digital preservation**

The study revealed an interesting aspect regarding cost consciousness and cost maturity in digital preservation. The great majority 87.5% indicated that they did not have a cost model or they did not know if there exists one. Only 12.5% of the respondents mentioned that they had a cost model. In my opinion it is crucial that organisations apply a cost model as it helps planning for the future and it can act as a guide in funding the continuity and sustainability of digital preservation projects.

The study revealed that although the majority of the represented organisations had included digital preservation in their overall organisation's objectives and budgeting but this was hidden away in the general IT budget. There should be a clear demarcation between the budgetary allocation for the digital preservation and other general IT activities. This will not only allow proper planning in the organisations but it will increase the visibility and the importance of this tasks.

The study revealed clearly that the lack of awareness on cost issues in digital preservation is a major challenge. Organisations should put more emphasis on informing their employees about digital preservation costs especially those responsible for the long term preservation.

2. **To find out how organisations allocate resources towards management of digital records**

The study revealed that organisations were more willing to allocate resources on storage and technology than on human resource. The respondents did not consider the staff as a major cost factor as they were just regarded as part of the general fund allocation not specifically to digital preservation.

Another interesting observation from the study was that organisations considered the IPR (Intellectual Property Rights) and outsourcing costs as irrelevant in digital preservation. However, the issue of copyright laws has recently received a lot of attention in the media and organisations should be more careful to ensure that they meet the legal requirements. Infringement of copyright laws can potentially imply large cost to an organisation. Outsourcing can be a considerable cost factor in cases where organisations are required to acquire experts to carry out certain rather infrequent tasks such as designing preservation plans, migration etc. where the regular staff lacks competency to carry out the tasks.

3. **To establish how OAIS Reference model can be used when determining costs for digital preservation.**

One of the interesting findings of the study was that only 12.5% of the online respondents were aware that the OAIS model can be used for determining costs in digital preservation. The OAIS model is a relevant model for determining cost factors not just a theoretical framework. Therefore, organisation can adopt it as a standard in identifying cost issues in digital preservation. This was evident from the study as most respondents gave a positive opinion of its relevancy in identifying cost factors in digital preservation.

In summary the following were the key findings of the study:

- Most organisations did not have a cost model for digital preservation.
- Few organisations were using the OAIS model as framework in determining cost factors in digital preservation.
- There were different ratings on the cost factors decomposed from the OAIS model.

From the study it became apparent that the following issues will have a significant impact on the cost of digital preservation:

- Automation is a significant factor in controlling the costs of large-scale digital preservation. Manual processing is one of the largest cost items for digital preservation. Consequently, automated processes and automated evaluation (tests) are significant factors in reducing the costs of digital preservation. This is also supported by the findings from Beagrie et al. (2008) who stated that the goal of any long term digital storage must be to automate as much as possible.

- The cost factors may differ between institutions and countries, but the model for the estimation of costs has a wider relevance and can be used to make similar calculations in other situation *(Palm, 2006)*. The application of this cost factor model depends on the nature of the organisation, sustainability, maturity and the level of technology.

- Preservation costs will increase in future due to increase in volume of information generated therefore calculating costs needs to take place in a realistic context and should take into consideration the designated community.

6.2 FURTHER RESEARCH

The study has provided a wide but focused insight into the topic of cost factors in digital preservation. It would be interesting to take a specific case study (which can be a specific organisation) and carry out a detailed analysis in relation to the cost factors in digital preservation, as this thesis focused on representatives of professionals in Norway and not on a specific organisation.

In the thesis an outline of a tentative formula for a cost factor model is suggested. A further study would shed light on how to exactly calculate costs for specific cost factors. An example of a formula for calculating metadata costs, migration costs, emulations costs and standard costs was given in the literature review. This might be done on other cost factors making the formula more relevant to different situations.

This thesis only focused on using OAIS model as a cost framework. Research can be done to include other cost models too and thereby contributing in understanding the issues of cost in digital preservation.

Finally, during the study it was realized that the discussed OAIS cost factors could be translated into more recognizable economical terms with help of a cost metrics (see figure 22 below). By applying this metrics to the various cost factors (i.e. OAIS functions and their required resources) it would be possible to provide cost statements more easily understandable and usable by the financial department in an organisation. A color coding (red, orange, and green) or some grades (1, 2, and 3) in the cells of the cost metrics would give easy and appealing information about the relative size of the involved costs. Pursuing this idea is however left for further research.

Cost	**Extend**: • Na= not applicable • low, • medium, • high	**Frequency of occurrence** • never, • rare= every 3-5 yrs, • regularly= 1-2 /yr, • frequently = weekly or daily)	**Shared** • Y=with other IT/human resources, • N= purely long-term storage	**Other characteristics?**
Direct cost in $ (investment / outsourcing)				
Time				
Hardware /Software tools				
Computation power				
Man power				
Expertise				
Energy				

Figure 22: An example of a cost metrics that may be used to translate the various cost factors into more "economic" terms.

REFERENCES

1. ALA (American Library Association) (2007). *Definitions of digital preservation.* Chicago: American Library Association. Available at: http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf
2. Ashley, K. (2000). *Digital Archive Costs: Facts and Fallacies*. Paper presented at the DLM Forum '99. Retrieved on 29th April 2009 from http://ec.europa.eu/archives/ISPO/dlm/fulltext/full_ashl_en.htm
3. Beagrie, N., Chruszcz, J. & Lavoie, B. (2008). Keeping Research Data Safe: A Cost model and guidance for UK Universities. Joint Information Systems Committee. Retrieved on 27th June 2009 from http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf
4. Beedham,H. Missen, J. Palmer, M &Ruusalepp, R (2004). *Assessment of UKDA and TNA compliance with OAIS and METS standards.* Accessed on 6th June 2009 from http://www.jisc.ac.uk/uploaded_documents/oaismets.pdf
5. Bennett, J. C. (1997). *A framework of data types and formats, and issues affecting the long-term preservation of digital material*. JISC/NPO studies on the preservation of electronic materials, British Library Research and Innovation Report, **50**
6. Best, K., (2007). *Reducing the costs of standards activities*. Kava Corporation. Accessed on 9th June 2009 http://www.kavi.com/solutions/whitepapers/Kavi_ReducingCost.pdf
7. Brown, A. (2003).*Digital preservation guidance. Selecting storage media for long term preservation.* Accessed on 1st April 2009 http://www.nationalarchives.gov.uk/documents/selecting_storage_media.pdf
8. Business Dictionary (2009). Business Dictionary. Accessed on 24th June 2009 at http://www.businessdictionary.com/definition/cost.html
9. Calanag, M.L., Sugimoto, S., & Tabata, K. (2001). A metadata approach to digital preservation. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications.*
10. CCSDS (Consultative Committee for Space Data Systems) (2002). *Reference Model for an Open Archival Information System (OAIS).* Blue Book, Issue 1. Washington, DC (US): CCSDS Secretariat, January 2002. Technical report. CCSDS 650.0-B-1. Recommendation for Space Data System Standards. Available at: http://public.ccsds.org/publications/archive/650x0b1.pdf.
11. Cedars project (2002). *Cedars Guide to Digital Preservation Strategies*. Retrieved on 15th march 2009 from http://www.leeds.ac.uk/cedars/guideto/dpstrategies/.

12. Charmaz, K., (1995). Grounded theory in Smith, J.A., Harre, R., &Langenhove, L.V. (eds) *Rethinking methods in psychology*. London: sage.
13. Chen, S., (2001). *University of Missouri-Columbia the paradox of digital preservation*. Accessed on 3rd April 2009 http://www.mediamatters.net/docs/resources/Digital%20Preservation/ParadoxOfDigitalPreservation.pdf
14. Cornell University Library (2003) .*Digital preservation management: Implementing short term strategies for long term problems.* Accessed on 10th April 2009 from *http://www.icpsr.umich.edu/dpm/dpm-eng/tutorialprint.pdf*
15. Crystal, A., & Land, P. (2003). *Metadata and search* .Global corporate circle DCMI 2003 workshop. Retrieved on 31st may, 2009 from http://dublincore.org/groups/corporate/Seattle/
16. Cunningham, A. (2000). *Dynamic descriptions: recent developments in standards for archival description and metadata*. Canadian Journal of Information and Library
17. Currall, J., Johnson, C. & McKinney, P. (2006). *Digital preservation as an albatross*. Retrieved on 6th march 2009 from https://dspace.gla.ac.uk/bitstream/1905/535/1/espida_jcpmcj_albatross.pdf
18. Day, M. (2003). *Integrating metadata scheme registries*. http://www.siderean.com/dc2003/101_paper38.pdf
19. Doane, M. (2003) *Metadata, Search and Meaningful ROI.* Global Corporate Circle DCMI 2003 Workshop [PowerPoint slides]. Retrieved on 15th may 2009 from. http://dublincore.org/groups/corporate/Seattle/Circles-Workshop-Papers/DC2003-Doane.ppt
20. Dollar Consultancy (2002). *Digital Quality Assurance Tools: Technical Evaluation and Recommendations*. Smithsonian Institution Archives
21. Eakin, L., Friedlander, A., Schonfeld, R. & Choudhur, S (nd). *A Selective Literature Review on Digital Preservation Sustainability***.** Accessed on 23rd April, 2009 from http://brtf.sdsc.edu/biblio/Cost_Literature_Review.pdf
22. ERPANET (2003). *Cost orientation Tool*. Accessed on 6th June 2009 from http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf
23. Gantz, J F, Chute, C, Manfrediz A, minton, S, reinsel,D. Schilichting w. & Toncheva, A. (2008). *The diverse exploding Digital universe: An updated forecast of worldwide information growth through 2011.IDC whitepapers*. Retrieved on 4th march 2009 from http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf
24. Granger, S. (2000). Emulation As a Digital preservation Strategy. DLib Magazine volume 6. No 10. Retrieved on 20th may 2009 from http://www.dlib.org/dlib/october00/granger/10granger.html

25. Gray, E.D. (2004). *Doing Research in the Real World*. Sage publication. London. Thousand Oaks. New Delhi.
26. Hancock, D.R and Flowers, C.P. (2001). *Comparing Social desirability responding on World Wide Web and paper administered Surveys, Educational technology Research and development.*
27. Hendley, T. (1998). *Comparison of Methods and Costs of Digital Pres*ervation. Retrieved on 20th from http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html
28. IBM system storage (2008). *Manage exponential information growth and costs with archiving and retention solutions from IBM*. Retrieved on 6th march 2009 from ftp://ftp.software.ibm.com/common/ssi/pm/br/n/tsb03002usen/TSB03002USEN.PDF
29. Investor dictionary (2009) .Investor Dictionary. Accessed on 24th June 2009 from http://www.investordictionary.com/definition/cost.aspx
30. Jones, M. & Neil B. *Preservation Management of Digital Materials: A Handbook.* Online: Digital Preservation Coalition, 2002. Retrieved on 24th may 2009 from URL:http://www.dpconline.org/graphics/handbook/.
31. Kol, N.& Oltmans, E. (2005). *Comparison between emulation and migration*. Retrieved on 24th may 2009 from http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file959.html#article0
32. Lavoie, B. (2004). *Thirteen ways of looking at..... digital preservation.* DLIB magazine July/ August, vol 10 number 78. Retrieved on 21st April 2009 from *http://www.dlib.org/dlib/july04/lavoie/07lavoie.html*
33. Lavoie, B. F. (2003). *Incentives to Preserve Digital Material: Roles, Scenarios and Economic Decision-Making.* Retrieved On 4th April 2009 from http://www.oclc.org/research/projects/digipres/incentives-dp.pdf
34. Lavoie, B., (2004). The Open Archival Information System Reference Model: Introductory Guide, DPC Technology Watch Series Report. Retrieved on 3rd may 2009 from http://www.dpconline.org/docs/lavoie_OAIS.pdf
35. Lougee, W.P. (2002). *Diffuse Libraries: Emergent Roles for the Research Library in the Digital Age.*
36. Mestl, T., Nymoen, K. Boyum, J. &Gaustad, L.(2007). *Survivability of digital records. Technical Report*. Det Norske Veritas(DNV). Norway.
37. Miles M. B & Huberman, A.M (1994) qualitative data analysis, 2nd edition. Thousands oaks, CA:sage
38. National Archives of Scotland (NAS)(2005). *Records Management*. Retrieved on 6th march 2009 from http://www.nas.gov.uk/recordKeeping/recordsManagement.asp#whatIsRec

39. Palm, J. (2006). The Digital Black Hole. National Archives, Stockholm, Sweden. Retrieved on 27th June 2009 from http://www.tape-online.net/docs/Palm_Black_Hole.pdf
40. Pickard A. (2007). *Research Methods in Information*. Facet Publishing, London.
41. Ruiz, A., S.(2003). *Cost Definition*. Retrieved on 24th June 2009 from http://www.loscostos.info/english/definition.html
42. Russell, K., & Weinberger, E. (2002*). Cost elements of digital preservation.* Retrieved on 15th march 2009 from http://www.leeds.ac.uk/cedars/documents/CIW01r.html
43. Ruusalepp, R. (2003). *AHDS Archive Ingest Procedures Framework AHDS Preservation Procedures Manual*. Estonian Business Archives, Ltd. Accessed in http://ahds.ac.uk/preservation/ingest-procedures-review.pdf
44. Testbed Digitale Bewaring. (2005). *Costs of Digital Preservation*. The Hague, Netherlands: Nationaal Archief of the Netherlands. Retrieved on 5th may 2009 from http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf
45. Warren**,** H.S. (2003). *Hacker's Delight*. Addison Wesley Publishing Co., Boston. Retrieved on 31st march 2009 from http://www.hackersdelight.org/crc.pdf
46. Woodyard-Robinson, D. (2006). *Costs and business modelling*. Accessed on 5th June 2009 http://www.dpconline.org/graphics/inststrat/costs.html
47. Yin, R. (2009). *Case Study Research*. Fourth edition. Thousand Oaks : Sage publication
48. Yin, R.K (1994). Case study Research. Design and methods. 2nd edition. Thousand Oaks : Sage publication

APPENDIX 1. ONLINE QUESTIONNAIRE

COST FACTORS IN DIGITAL PRESERVATION

Doreen Kerubo Mageto
Oslo University College
dorinkerubo@yahoo.com
Mob. No. +4799873972

Dear respondent,
I am a student of International Master in Digital Library Learning(http://dill.hio.no/) at Oslo University College undertaking a research on DETERMINING COST FACTORS USING OPEN ARCHIVAL INFORMATION SYSYEM (OAIS) MODEL IN PRESERVING DIGITAL RECORDS IN ORGANISATIONS. My supervisors are Thomas Sødring of Oslo university college (Thomas.Sodring@jbi.hio.no) and Inger-mette of DNV, Norway (Inger.Mette.Gustavsen@dnv.com)

The aim of this research is to use the OAIS functional model in identifying the cost factors that organisations need to be aware in preserving digital records. The cost factors will enable organisations to budget for the preservation of digital records. This study forms a major component of the requirement for the fulfilment of the masters in DILL. The results of the research shall also be a contribution to the LongRec (Long-Term Records management) project headed by the Det Norsk Veritas(DNV) in collaboration with a number of case partners commercialization partners and research partners. The results shall be published in the project's public web site (http://www.longrec.com/). At the end of the research you will receive a draft with the initial findings based on the answers of all the participants, The outcome of the survey shall enable us to come up with a formula for estimating costs of digital preservation over a period of time which shall be part of the draft.

In order to elicit the required information, you are kindly requested to respond honestly and objectively to all the items in the questionnaire to the best of your knowledge. This information will be used for academic purposes only and the information provided will be treated with the confidentiality it deserves

Thanks in advance
Yours faithfully
Doreen Mageto

1) 1. Is digital preservation included in the organisational overall strategic objectives?	
YES	

NO	

2) 2. Is the organisation using any cost model in determining cost factors in digital preservation	
YES	
NO	
IF YES, (Please Specify):	

3) 1. In practice, what cost factors do you take into consideration when preserving digital objects?	
Staff costs	
Security costs	
Technology costs	
Storage costs	
Legal costs	
Don’t know	
Other (Please Specify):	

4) 4. What challenges do you face when determining cost factors in digital preservation?	
Lack of funding	
Lack of awareness	

Lack of staff resources	
Lack of evidence of benefit	
Don’t know	
Others (Please Specify):	

5) 5. What areas in the overall management of digital records does the organisation allocate funds?	
Storage	
Staff	
Hardware and software	
Security	
Customer service	
Technology watch	
Rights management	
Don’t know	
Others (Please Specify):	

6) 6. Is the cost of digital preservation included in the budget of the organisation?	
Yes	
No	

7) 7. Rate the following cost factors in digital preservation in order of criticality to your institution, with: irrelevant, moderately important, important, very important and crucial.						
	Irrelevant	Moderately important	Important	Very important	Crucial	Other (Please Specify):
Selection costs						
Quality assurance costs						
Metadata costs						
File format conversion costs						
Storage costs						
Security costs						
Migration costs						
Emulation costs						
Error checking costs						
Validation costs						
Backup costs						
Disaster preparedness						
Training costs						
Software development costs						
Hardware costs						
Staff costs						

Intellectual property rights agreement costs						
Technology watch costs						
Outsourcing costs						
Interface design						
Customer service/support costs						

8) Has your organisation allocated ongoing, budgeted resources for the long-term care of digital materials?	
YES	
NO	

9) Does your organisation use the OAIS model in determining cost factors in digital preservation?	
YES	
NO	
IF YES (Please give details):	

10) What cost factors are taken into consideration during ingestion/ acquisition (tick the ones that apply to your organisation).	
Creation costs	
Selection costs	

Network costs	
Integrity check costs	
File format conversion costs	
Metadata costs	
Don’t know	
Other (Please Specify):	

11) What cost factors are taken into consideration in data management in your organisation?	
Database management system costs	
Hardware costs	
Staff costs	
Training costs	
Other (Please Specify):	

12) What costs are incurred at the archival storage stage in your organisation? (Tick the ones that apply to your organisation).	
Storage media costs	
Security costs	
Monitoring costs	

Migration costs	
Emulation costs	
Error checking costs	
Validation costs	
Backup costs	
Storage media costs	
Staff	
Equipment	
Disaster preparedness costs	
Training costs	
Other (Please Specify):	

13) What administration cost factors do you take into consideration in your organisation?	
Rights and copyright agreement costs	
Licence costs	
Legal costs	
Physical security costs	
Establishment of standards and policies costs	
Outsourcing costs	
Customer service/support costs	

Power	
Physical infrastructure	
Don’t know	
Other (Please Specify):	

14) What cost factors are taken into consideration during preservation planning in your organisation? (Tick the ones that apply to your organisation.	
Collaboration costs	
The technology watch costs	
Development of preservation strategies and standards costs	
Outsourcing costs	
Staff costs	
Designing costs	
Don’t know	
Other (Please Specify):	

15) What costs factors does your organisation take into consideration when providing access to the digital records?	
System and application interface costs	
Media/ format costs	

Resource discovery and retrieval costs	
Customer service/support	
Security costs	
user training costs	
Don’t know	
Other (Please Specify):	

16) Do you have any other relevant information that can be of help in this research?	
YES	
NO	
Other (Please Specify):	

17) Do you know of any valuable sources that can add value into the research topic?	
YES	
NO	
Other (Please Specify):	

18) To help us achieve maximum results from this survey we would like to know if you can be be willing to be interviwed .

YES	
NO	
IF YES (please provide email):	

19) is there any information that you would like to share to contribute to this research ?

APPENDIX 2. INTERVIEW GUIDE

Doreen Kerubo Mageto
Oslo University College
dorinkerubo@yahoo.com
Mob. No. 99873972

Dear respondent,

I am a student of International Master in Digital Library Learning at Oslo University College undertaking a research on DETERMINING COST FACTORS USING OPEN ARCHIVAL INFORMATION SYSYEM (OAIS) MODEL IN PRESERVING DIGITAL RECORDS IN ORGANISATIONS. My supervisors are Thomas Sødring of Oslo university college (Thomas.Sodring@jbi.hio.no) and Inger-mette of DNV, Norway (Inger.Mette.Gustavsen@dnv.com)

The aim of this research is to use the OAIS functional model in identifying the cost factors that organisations need to be aware in preserving digital records. The cost factors will enable organisation to budget for the preservation of digital records. This study forms a major component of the requirement for the fulfilment of the masters in DILL.

In order to elicit the required information, you are kindly requested to respond honestly and objectively to all the items in the interview guide to the best of your knowledge. This information will be used for academic purposes only and the information provided will be treated with the confidentiality it deserves

Thanks in advance

Yours faithfully

Doreen Mageto

INSTRUCTIONS

Kindly answer the following questions as honestly and objectively as possible. Confidentiality about the information you give will be maintained.

SECTION A

Aim: To establish how the organisation determines the cost factors for digital preservation

1. What in your opinion are the cost factors one should consider in digital preservation?
2. In practice, what cost factors do you take into consideration when preserving digital objects?
3. How does your organisation identify the cost factors for digital preservation?
4. What challenges does your organisation face when dealing with cost issues in Digital preservation?
5. Is digital preservation included in the organisational overall strategic objectives?
    a. If yes, please explain.

SECTION C

Aim: To find out how the organisation allocates resources towards the management of digital records?

1. What areas in the overall management of digital records does the organisation allocate resources?
2. What are the factors that the management take into consideration when allocating funds for digital preservation?
3. How does the management of the organisation decide on the funds to be allocated for the management of the digital records?
4. Is the cost of digital preservation included in the budget of the organisation? Explain?
5. What priority areas in digital preservation would you like to be financed when resources are limited?

SECTION D

Aim: To establish how OAIS model can be used when determining costs for digital preservation?

1. Does your organisation use the OAIS model in determining cost factors in digital preservation? If yes. Explain.

2. We would like to receive your opinion on the following OAIS functional model in relation to cost in digital preservation.

    a. Ingest
        i. What are the cost factors involved when accepting digital objects (SIPs) from originators / producers?

        ii. What costs are incurred in checking for the quality of the digital objects during submission?

        iii. What cost factors do you incur when handling file formats of digital objects?

        iv. What are the costs incurred in preparing metadata for the digital objects.

    b. Archival storage
        i. What are the specific storage costs that you incur in preserving digital objects?

        ii. What cost factors do you take into consideration when replacing media (e.g. migration, emulation etc)?

        iii. What cost factors do you consider when undertaking backups for the digital objects?

        iv. What are the security costs involved in preservation of digital objects?

    c. Data management

        i. What are the cost factors that you consider in data management?

        ii. To what extent are the following issues of concern when determining cost factors in data management
            1. Staff costs

            2. Database management system

            3. Training costs

    d. Administration
        i. What are the cost factors that you consider when negotiating for submission agreements?

ii. What costs do you incur when ensuring the security of the digital objects in your custody?

iii. What cost factors are involved when establishing standards and policies in digital preservation?

iv. What costs are involved in auditing submission (both the Submission information package and archival information packages)?

v. What costs are involved in customer service in relation to digital preservation?

e. Preservation planning
   i. What costs do you incur when collaborating with other organisations on preservation of digital objects?

   ii. What costs factors should be taken into considerations when monitoring changes in technology?

   iii. What costs are incurred in developing package designs and migration plans?

f. Access
   i. When providing access to customers what cost factors does the organisation take into consideration?
   ii. What costs are incurred in customer service/support when giving access to digital objects?
   iii. What security issue should organisations be aware of when giving access to digital objects?
   iv. What other factors should be taken into consideration in providing access of digital objects to designated community?

THANKS FOR YOUR COOPERATION

APPENDIX 3. LONGREC PROJECT

LongRec (Long term Records management) is a three year research project (2007-2009) partly funded by the Norwegian Research Council. The project constitutes the Norwegian team of the InterPARES 3 project. LongRec addresses several research challenges, each of which is assigned a short name for simplicity: records transition survival (READ), long-term usage (FIND), preservation of semantic value (UNDERSTAND), preservation of evidential value (TRUST) as well as legal, social, and cultural framework (COMPLIANCE). Each research challenge is addressed by:

- General studies compiling state of the art and best practice of the area.
- Research on selected sub-topics, performed by the research partners and by one PhD student for each research challenge.
- One or more case studies with the LongRec case partners.
- Studies on opportunities for products and services with the commercialization partner

The primary goal of this joint-industry project is the Persistent, Reliable and Trustworthy Long-Term Archival of Digital Documents, with Emphasis on Availability and Use of Documents. The particular problems addressed by the LongRec project typically emerge when document lifetime exceeds 20 years, and LongRec imposes no upper limit on the lifetime. The main objectives for the LongRec partners are

a) To enable transition to digital original documents and digital work processes for information that must be available and in use over decades, and

b) To explore the potential for commercial products/services in this area.

Case studies addressed by LongRec include documentation for physical objects that are in use for decades (ships, oilrigs, power plants, and others) and documents in public registries. Results may be generalised to other cases, e.g. health information. LongRec goes beyond the digital preservation area addressed by libraries and (public) archives in that documents also need to be used (retrieved, updated, verified) subject to constrains related to ownership and authorisations. All parts of a document's environment (technology, processes, organisations, roles/people, and ownership) must be expected to undergo several changes during the lifetime of the document. LongRec goes beyond state of the art in records management by addressing long-term aspects, and preservation not only of availability and readability, but also of semantic value (meaning, context) and evidential value (trustworthiness).

Benefit for customers

LongRec provides solutions to overcome limitations regarding the long-term perspective of work processes, both within a company and across company borders, digital processes which are increasingly relying on digital documents. By establishing a scientific knowledge base for handling the long-term perspective, it becomes viable to build products and services for the long-term perspective. By offering trusted, long-term storage services the service provider takes on a great responsibility to provide services for an unknown future. In order to make this a viable business it is essential that LongRec has to be built on sound, scientifically established standards and methods that are also supported by technology, demonstrating conformity with regulatory or legal constraints.

Benefit for society

The established theory, mechanisms, and technology will enable companies to trust long-term (several decades) storage of digital original documents, and be able to use and update the documents throughout their lifetime, thus strengthening and promoting the understanding of management of long-term digital documents in the educational system, to the academic research community, and to the society at large.

The case partners

The case partners of LongRec project include :DNV(Det Norsk Veritas), BBS, fast search and transfer, CSAM International, Nasjonalbiblioteket, StatoilHydro, Riksarkivet, Utenriksdepartementet , Brønnøysundregistrene , NTNU , NR , InterPares3 and Katholieke universitet Leuven.

Having looked at the case, the following section will focus in presenting the data collected based on the research question and the areas that were researched under each research question. The first research question focused on finding out how organisation determine costs in digital preservation and the areas investigated included finding out if they had a cost model, if digital preservation part of the organisation's overall objectives , the general cost factors in digital preservation and finally the challenges organisations face in determining cost factors in digital preservation.