Yibeltal Tafere Bayih

Application of Preservation Metadata for Long-Term Accessibility of Digital Objects

Supervisor: Raivo Ruusalepp

Master Thesis
International Master in Digital Library Learning
2010

Declaration

I certify that all material in this dissertation which is not my own work has been identified and that no material is included for which a degree has previously been conferred upon me.

…….Yibeltal Tafere Bayih (Submitted electronically)

Acknowledgments

I am heartily thankful to my helpful supervisor, Raivo Ruusalepp, for his encouragement, guidance and support. The advice that he gave me truly helped for smooth progress of this thesis. He really enabled me to develop better understanding of the subject. His support and co-operation is much indeed appreciated.

I am indebted to Sirje Virkus for reviewing the thesis. Her comments helped me to look things from different dimensions. I also thank her for her unforgettable support during the study for the last two years.

My grateful thanks also go to the individuals who I have interviewed for this thesis at NAE, NLE and NLW. Really, their keen participation made this study true. Thank you so much all!

My special thanks go to all DILL community in Oslo University College, Norway, Parma University, Italy and Tallinn University, Estonia for their support, encouragement and knowledge they have offered and shared in the last two years. I also thank the Erasmus Mundus Programme that sponsored my study.

I would like to thank Getaneh Agegn for his help and sharing of ideas. I also thank my fellow classmates for sharing any ideas and giving a truly memorable friendship experience.

Lastly, I would like to thank my parents for their sacrifice to educate and make me a better citizen. I am truly and deeply indebted to them!

Abstract

In the process of digital preservation metadata management for long-term accessibility of digital objects has been an important discussion point internationally. However, there is a gap in the implementation of preservation metadata standards from theory to practice. Hitherto little research has been conducted to show the application of preservation metadata and therefore new case studies on both implementation and use of metadata standards in preservation strategies is needed. The aim of this thesis is to study the extent of implementing standard preservation metadata in the preservation practice at memory institutions.

This study adopts a qualitative method based upon a pragmatic approach and uses the case study strategy. Metadata experts/specialists in three memory institutions (National Library of Estonia, National Archives of Estonia and National Library of Wales) were interviewed using semi-structured interviews, accompanied by document analysis.

Results of the study show that these memory institutions are recording a wide range of metadata in all categories: descriptive, structural as well as administrative metadata (including the rights, provenance, and technical metadata). They use metadata elements from a variety of metadata standards/schema to suit their practical purposes. However, the level of exploitation of preservation metadata standards differs in scale, data management practices as well as heterogeneity of metadata recorded. Metadata is recorded about different digital objects like books, WebPages, photographs, audio, video, and their files and bitstreams. The level of implementation of metadata for each object type varies between institutions. The application of the PREMIS metadata standard entities varies from institution to institution as it ranges from reviewing/analyzing stage to practical implementation. Significant differences have also been seen between national libraries and archives in mission, process of ingest, influence of their traditional cataloguing practices and types of standards used for the development of their metadata specification. In managing the metadata on the digital preservation processes different problems and challenges have been faced and investigated by these memory institutions and further research should be carried out to study other aspects of metadata implementation.

Keywords: preservation metadata, digital objects, memory institutions, digital preservation, metadata, preservation metadata standards, PREMIS, national library, national archives

Table of Contents

List of Abbreviations

AACR- Anglo-American Cataloguing Rules

AIP- Archival Information Package

AIS- Archival Information System

AMD- Audio Metadata

CCSDS- Consultative Committee for Space Data Systems

CEDARS- CURL Exemplars in Digital ARchiveS

CURL- Consortium of University Research Libraries

DAMS- Digital Asset Management System

DC – Dublin Core

DCMES- Dublin Core Metadata Element Set

DIGAR - Digital Archive of the National Library of Estonia

DROID- Digital Record Object Identification

DSEP- Deposit System for Electronic Publications

EAC- Encoded Archival Context

EAD- Encoded Archival Description

ERPANET- Electronic Resource Preservation and Access Network

ESE - Europeana Semantic Element

IP- Information Package

ISAAR(CPF)-International Standard Archival Authority Record for Corporate Bodies, Persons and Families

ISAD(G) - General International Standard Archival Description

ISO- International Organization for Standardization

JHOVE- JSTOR/Harvard Object Validation Environment

LC-AV-	Library of Congress Audiovisual Metadata
LCSH-	Library of Congress Subject Headings
MARC-	Machine Readable Cataloguing
METS-	Metadata Encoding and Transmission Standard
MIX-	Metadata for Images in XML
MODS-	Metadata Object Description Schema
NAE-	National Archives of Estonia
NASA-	National Aeronautics and Space Administration
NEDLIB-	Networked European Deposit Library
NISO-	The US National Information Standards Organization
NLA-	National Library of Australia
NLE-	National Library of Estonia
NLNZ-	National Library of New Zealand
NLW-	National Library of Wales
OAI-PMH-	Open Archives Initiative-Protocol for Metadata Harvesting
OAIS-	Open Archival Information System
OCLC-	Online Computer Library Center
OPAC-	Online Public Access Catalogue
PLOP-	PDF Linearization, Optimization, Protection
PREMIS-	PREservation Metadata: Implementation Strategies
RLG-	Research Libraries Group
SDB-	Safety Deposit Box
SourceMD-	Source Metadata
TEI-	Text Encoding Initiative

textMD- Technical Metadata for Text

TIFF- Tagged Image File Format

UAM- Universal Archiving Module

UKOLN- UK Office for Library and Information Networking

VMD- Video Metadata

XML- Extensible Markup Language

Terminology

The following terms have been selected from PREMIS data dictionary for preservation metadata version 2.0 (PREMIS 2008) and the ISO reference model of open archival information system (ISO 14721:2003) because of their relevance to this study.

Agent: actor (human, machine, or software) associated with one or more events associated with a digital object.

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.

Digital Object: discrete unit of information in digital form. A digital object can be a representation, file, bitstream, or filestream.

Digital preservation: applies to both born digital and reformatted content. It combines policies, strategies and actions to ensure the accurate rendering of authenticated content over time, regardless of the challenges of media failure and technological change.

Entity: Abstraction for a set of "things" (intellectual, agents, events, object, right) described by the same properties. The PREMIS data model defines five types of entities: intellectual entities, objects, agents, rights, and events.

Event: action that involves at least one digital object and/or agent known to the preservation repository.

Granularity: relative size, scale, level of detail, or depth of penetration that characterizes an object or activity. "Level of granularity" may be used to refer to the level of focus in a hierarchy or to refer to the level of specificity of description.

Intellectual Entity: coherent set of content that is described as a unit, for example, a book, a map, a photograph, a serial. An intellectual entity can include other intellectual entities; for

example, a web site can include a web page, a web page can include a photograph. An intellectual entity may have one or more representations.

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.

Metadata Schema: A formal specification of the semantics and structure of a coherent collection of attributes that can be assigned in the description of a resource, as well as constraints that may apply to such descriptions.

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. 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.

PREMIS (PREservation Metadata: Implementation Strategies) Data dictionary: common data model for organizing/thinking about preservation metadata and guidance for local implementations. It is standard for exchanging information packages between repositories.

Preservation Metadata: information a preservation repository uses to support the digital preservation process.

Rights: assertions of one or more rights or permissions pertaining to a digital object and/or an agent.

Schema: a systematic, orderly combination of elements or terms.

List of Figures

List of Tables

CHAPTER ONE: INTRODUCTION

This introductory chapter outlines the rationale for this research. First, the context in which this research is positioned is given by providing background information that leads to the discussion of the research problem. The statement of the problem describes the preservation metadata standards implementation from theory to practice and how metadata helps to maintain the accessibility of digital objects. The objectives, research questions and the methodology used in the study are then discussed, followed by limitations of the research.

1.1. Background Information

Digital preservation is a set of managed activities necessary to ensure the digital object can be accessed in the future. However, there are challenges like the hardware and software used to store and access digital objects that are continuously upgraded and outdated. Technology obsolescence is generally considered as the furthermost technical threat to ensuring continued access to digital objects (Hockx-Yu, 2006).

To ensure the long-term accessibility of digital objects, metadata is the key factor. Preservation requires special elements to track the roots of a digital object (where it came from and how it has changed over time), to detail its physical characteristics, and to document its behavior in order to emulate it on future technologies. Literature revealed that valuable metadata is the best way of minimizing the risk of digital resources becoming inaccessible and to be most valuable for all and needs to be consistently maintained throughout the process (Alemneh, Hastings and Hartman, 2002; NISO, 2004).

Preservation metadata is a type of metadata that contains information needed to archive and preserve a resource to support the functions of maintaining the fixity, viability, renderability, understandability, and/or authenticity of digital objects in a preservation context. It includes elements of administrative metadata, structural metadata, technical metadata - the subset of administrative metadata that documents detailed format characteristics of files and some rights metadata - the documentation of intellectual property rights, permissions, and restrictions on use. Of course, the scope and depth of the preservation metadata required for a

given digital preservation activity will vary according to numerous factors, such as the intensity of preservation, the length of archival retention, or even the knowledge base of the intended user community (Caplan, 2006).

Universally, there is a growing concern that digital resources will not survive in usable form into the future. This is because most metadata efforts and research are centered on the discovery of resources despite the fact that digital information is fragile and can be corrupted or altered, intentionally or unintentionally. Digital objects may become inaccessible as storage media, hardware and software technologies change. Hence, a number of efforts have been undertaken to perfect the digital preservation methods. Various organizations and agencies internationally have worked on defining metadata schemas for digital preservation like the National Library of Australia (NLA), CEDARS Project and a joint working group of OCLC (Online Computer Library Center) and RLG (Research Libraries Group) to name just a few. Many of these initiatives are based on or compatible with the standard reference model for an open archival information system (OAIS) (ISO 14721:2003) and these high-level preservation metadata initiatives provide much needed information required to manage the long-term preservation of digital resources (Alemneh, Hastings and Hartman, 2002; Lee, 2002).

OCLC and RLG jointly developed a metadata framework called PREMIS (PREservation Metadata: Implementation Strategies) which is outlining types of presentation metadata and developing a set of core elements and strategies for the encoding, storage, and management of preservation metadata within a digital preservation system. Currently, the PREMIS data dictionary influences the world to be an international *de facto* standard for preservation metadata (Caplan, 2006).

1.2. Statement of the Problem

Digital preservation is a relatively new phenomenon and the success of preservation metadata in supporting long-term preservation is largely untried; many specifications for preservation metadata have been published and significant progress has been made towards standardizing a core set of preservation metadata elements. However, “the movement from theory to practice in preservation metadata cannot be traced as a straight line, but rather as a series of

overlapping initiatives straddling research and development, with a substantial dose of cross-fertilization at the boundary"(Lavoie and Gartner, 2005, p.9).

In this regard a lot of efforts have been made to produce conceptual models and concrete metadata dictionaries for implementers of digital preservation services. For example, the set of core elements in the PREMIS data dictionary has now been widely accepted and plays a key role in creating coherence in the digital preservation metadata community. PREMIS provides a foundation to support interoperability across systems and organizations. However, literature revealed that there is a gap in its application into practice and this will have its own future challenge from the very aim of digital preservation like long-term accessibility of digital objects and others issues (Caplan, 2006). So, a number of case studies are expected to report on both implementation and use in carrying out preservation strategies (Caplan, 2006; Dappert and Farquhar, 2009).

Being digital does not necessarily mean being continuously accessible. Access to digital resources through descriptive metadata is only a short-term solution. Preservation metadata plays a significant role in facilitating preservation decisions, detects preservation threats and provides measures for minimizing risks to long-term access (Alemneh, Hastings and Hartman, 2002). On the other hand, issues like the expense associated with creation and maintenance of metadata over time pose practical difficulties.

According to the European research roadmap on access to and preservation of cultural and scientific resources (2007), to keep digital objects usable, meaningful, authentic and reliable requires an understanding of the significant properties that need to survive with the digital object for a long time. Partly, this depends on the chosen file format of the digital object, but most significant properties are determined by the business context in which they were created and used. Various methods are currently being developed to enable the extraction of significant properties of digital objects, but as yet there is little practical experience in this area. The European research roadmap also indicated that "additional fundamental research and practical experiments, covering the many different types of digital objects, are needed to gain a thorough understanding of the underlying issues" (DigitalPreservationEurope, 2007, p.27).

The ERPANET Briefing Paper in 2003 stated that digital preservation strategies (for example, migration, emulation, technology preservation) all depend to some extent on the creation, capture and maintenance of suitable metadata. To preserve digital objects, preserving the right metadata is the key. Hence, due to this and other various roles, metadata is a pressing topic on the research agenda of digital preservation for the coming years (DigitalPreservationEurope, 2007).

Thus, the focus of this study is on practice of preservation metadata at memory institutions that aim to look the extent of implementing theoretical standard in to actual practice especially from the PREMIS standard perspective the *de facto* standard for preservation metadata.

1.3. Aims and Objectives

The aim of this thesis is to study the implementation of standard preservation metadata into practice and the typical difficulties this poses.

The key objectives are:

- To examine the preservation metadata practice in the institutions.
- To identify and analyze the way how international metadata standards have been adopted for the digital preservation process.
- To analyze the way how and reasons for using metadata to support the digital preservation processes.
- To investigate risks that can be anticipated in the current practice of preservation metadata usage in memory institutions.

1.4. Research Questions

The central questions to this study are:

1. How effective are preservation metadata theories into practice?

2. What tools, standards and strategies are adopted for metadata management in practice and why?

3. What is the level of granularity (e.g., representations, files, bitsreams) that preservation metadata is applied in the practice of memory institutions?

4. What type of risks can be anticipated when preservation metadata implemented only partially in practice?

1.5. Methodology

This thesis is using qualitative study methods based upon a pragmatic approach and the chosen research strategy is a case study. The study looks at three memory institutions and their use of metadata in their digital preservation practice. These institutions are the National Library of Estonia and the National Archives of Estonia and the National Library of Wales. Both interviews (face-to-face interview for the first two institutions and interview via email with follow-up for the third institution) and document analysis were used for the data collection exercise. The process of data analysis consists of coding the interviews and organizing codes and the data from documents into themes that correspond with the research objectives and research questions. A more detailed discussion of the methodology can be found in Chapter 3.

1.6. Limitation and Scope of the Research

There are a few limitations that should be outlined in order to have a clearer idea of the scope of this study.

- The number of memory institutions used for the case studies was relatively small. This was mainly due to time and resource constraints of the MA thesis project.

- Due to geographic distance and potential inconvenience for the respondents, an interview at one of the memory institutions was conducted via email with follow-up questions.

- Only English language implementation documentation was used because of the language barrier.

- The literature review covers only publications in English.

1.7. Significance of the Study

- Metadata is central to digital preservation processes; very few standards for digital preservation metadata exist; the application of these standards into practice is limited due to the complexity of the subject area and existing traditions and practices in institutions involved with preservation; outlining the core reasons why preservation metadata standards fail to be implemented to the full will help memory institutions to plan their metadata and digital archive initiatives.

- This thesis contributes to the research through the case studies that report the implementation of preservation metadata standards in to practice.

- It will also act as a source of reference for those who want to do further research on the same area.

1.8. Outline of the Thesis

The first chapter of this thesis provides a rationale for the study by providing background information which gives context to the work as a whole. The research problem, the objectives and research questions of the study are stated and the perceived limitations further contextualize this study.

Chapter 2 reviews the literature that is pertinent to the topic and that has informed this study. The literature review provides an overview of digital preservation issues; metadata requirements for long-term preservation; preservation metadata standards and their implementation issues are reviewed.

The third chapter outlines the methodology used in this research project. The data collection and analysis methods are discussed.

Chapter 4 comprises the data analysis and main findings. It explores the main themes that correspond to the objectives and research questions of this study.

The final chapter presents conclusions from this research project and offers suggestions for areas of further research.

1.9. Chapter Summary

This introductory chapter has provided background information to this research and discussed the initial stimulus for the study. The research problem has been presented and justifications for continuing this research have been provided. The methodology has been briefly described and limitations as they apply to this study have been addressed. An overview of how this thesis will progress has also been provided. The following chapter reviews the literature as it pertains to this study.

CHAPTER TWO: LITERATURE REVIEW

2.1. Introduction

This chapter reviews various works which are relevant to the research topic of this study. It discusses the relevant concepts needed to find answers to the research problem. It will start with general discussion of issues and challenges in digital preservation then review matters pertaining to the preservation metadata.

Further on, the OAIS reference model is discussed in relation to preservation metadata. The key preservation metadata standards and initiatives, the process of implementation of theory to practice are included in the review so as to provide general understanding of preservation metadata as a whole.

Finally, concepts of interoperability and metadata registries with respect to preservation metadata implementation are discussed.

2.2. Issues and Challenges in Digital Preservation

Digital preservation is defined as "the managed activities necessary for ensuring both the long-term maintenance of a bitstream and continued accessibility of the document contents through time and changing technology" (RLG, 2002, p.3). As Ross (2007) explained, digital preservation is not only about "keeping the bits those streams of 1s and 0s that we use to represent information" but also about "maintaining the semantic meaning of the digital object and its content, about maintaining its provenance and authenticity, about retaining its 'interrelatedness', and securing information about the context of its creation and use" (p.2).

Therefore, in order to keep the digital object for so long, it needs to manage the digital preservation activities soundly.

Lee, Slattery, Lu, Tang and McCrary (2002) point out that:

> Digital preservation involves the retention of both the information object and its meaning. It is therefore necessary that preservation techniques be able to understand and re-create the original form or function of the object to ensure its authenticity and

> accessibility. Preservation of digital information is complex because of the dependency digital information has on its technical environment (pp.93-94).

As Strodl, Becker, Neumayer and Rauber (2007) underline, in the digital library community, digital preservation as the process of keeping digital objects accessible and usable for a certain period of time has turned into one of the most pressing challenges. This is because of the rapid changes and ongoing developments in hardware, software, file formats, information technology infrastructure and computer equipment in general which makes long-term archiving of digital objects a highly complex and diverse matter. In this regard, Lee et al. (2002) also state that digital resources present more complex problems than conventional analogue media as newer digital technologies rapidly appear and older ones are outdated, information that relies on obsolete technologies soon becomes inaccessible.

From this we can understand that the speed of transformation in information technology shows that data can be inaccessible within few years and needs to take action in the process of digital preservation. According to Rosenthal, Robertson, Lipkis, Reich and Morabito (2005), "the goal of a digital preservation system is that the information it contains remains accessible to users over a long period of time." However, the design of such systems is the key problem and there are several reasons for this complexity (p.2).

The first one is the period of time that usually is very long – much longer than the lifetime of individual storage media, hardware and software components and the formats in which the information is encoded, i.e., no media, hardware or software exists in whose longevity designers can place such confidence (Rosenthal et al., 2005). The second one is "digital information is threatened by the speed in which new types of hardware and software replace current versions" (Oltmans and Wijngaarden, 2004, p.23).

The complexity of digital preservation was anticipated by scholars a decade ago. For example, Terry (1997) predicted (as cited in Caplan (2007)), "technological obsolescence, the proliferation of file formats, restrictive intellectual property regimes and the like would see us into an era where much of what we know today, much of what is coded and written electronically, will be lost forever" (p.449). Moreover, the risks of digital volatility both in terms of storage media permanence and of uncontrolled obsolescence of technology reflected

in changes in operating systems, file formats, input and output devices, programming languages and software applications have been recognized as serious threats to the future of exponentially growing digital assets (Bennett, 1997; Cordeiro, 2004; Groenewald and Breytenbach, 2009). This is because new types of hardware, computer applications and file formats supersede each other, making digital information inaccessible in the long-term, i.e., either the formatted bit stream becomes obsolete (media deterioration), or there is no functionality available to decode this bit stream and render the information to the user (Oltmans and Wijngaarden, 2004).

Digital preservation is a crucial issue and calling for measures that go beyond permanent archiving and all stakeholders agree that digital resource preservation encompasses a wide variety of interrelated activities (Alemneh et al., 2002). The main rationale behind digital preservation is to ensure protection of information of enduring value for access by present and future generations and hence it comprises of planning, resource allocation and application of preservation methods and technologies necessary to ensure that digital information of continuing value remain accessible and usable (Das, Sharma and Gurey, 2009).

As Jana et al. (2009) explain, "the strategy of digital preservation is a particularly technical approach to the preservation of digital resources for maintaining and accessing over the long-term even though no one is appropriate for all" (p.22). The fact that made preserving digital resources difficult is that they can only be read by software. This would mean that in order to ensure long-term access to digital resources, we need to preserve all the software, hardware, and operating systems on which the software ran (Alemneh et al., 2002).

Digital information requires detailed metadata perhaps more than any other media to ensure its preservation and accessibility for future generations (OCLC/RLG, 2001) and overall, metadata is the key resource in order to facilitate resource discovery, to organize electronic resources, to facilitate interoperability and legacy resource integration, to provide digital identification and support archiving and preservation of digital objects (NISO, 2004).

2.2.1. Digital Objects

Digital documents are modeled in very different ways. It can be a “sequence of expressions or a sequence of scanned page images, and so on”. Preservation of digital information object does not necessarily involve maintaining all of its digital attributes in addition the management and then preservation of it depend on the model that is applied(Thibodeau, 2002, p.5). According to Thibodeau, for any use in addressing the challenge of digital preservation it is possible to define “a digital object as an information object, of any type of information or any format that is expressed in digital form” (p.5).

Further, he elaborated that “all digital objects are entities with multiple inheritances; that is, the properties of any digital object are inherited from three classes: physical, logical, and a conceptual object, and its properties at each of those levels can be significantly different” (p.6). The digital object levels are described as:

- A physical object- is an inscription of signs on some physical medium, i.e., this level deals with physical files that are identified and managed by some storage system. The physical inscription is independent of the meaning of the inscribed bits.
- A logical object- is an object that is recognized and processed by software. It is a unit recognized by some application software. This recognition is typically based on data type. A set of rules for digitally representing information defines a data type.
- The conceptual object- is the object as it is recognized and understood by a person, such as a book, a contract, a map, or a photograph or in some cases recognized and processed by a computer application capable of executing business transactions. (p.8).

Hence, Thibodeau (2002) illustrated it as follows:

> To preserve a digital object, the relationships between these levels must be known or knowable, i.e., we must be able to identify and retrieve all its digital components. The digital components of an object are the logical and physical objects that are necessary to reconstitute the conceptual object... For example, to retrieve a report stored as a master and several subdocuments, we must know that it is stored in this fashion and we must know the identities of all the logical components. To retrieve a specific order from a sales application, we do not need to know where all or any of the data for that

order are stored in the database; we only need to know how to locate the relevant data, given the logical structure of the database (pp.11-12).

To successfully apply different preservation strategies and manage the digital object, components of the levels of the digital object should be well studied and identified.

2.2.2. Digital Preservation Strategies

There are many digital preservation strategies developed to preserve digital objects and keep them accessible in the long run. Migration and emulation are the most prominent ones (Strodl et al., 2007, p.2).

- Migration: it is the method of repeated conversion of files or objects. A file is converted to either a more current version of its own file format, or to another, which is easier to handle and access.

- Emulation denotes the duplication of the functionality of systems, be it software, hardware parts, or legacy computer systems as a whole, needed to display, access, or edit a certain document. Emulating a certain version of a software system needed to access a file in an outdated version or format and it is the most frequently method in the digital preservation context.

Both strategies have their own requirements, problems, different solutions to the problem and their applicability is also highly challenging and context dependent. Strodl and his colleagues also added that “preservation strategies and specific software tools for emulation or migration must always be chosen according to requirements of individual institutions” (Strodl et al., 2007, p.2).These strategies rely on the preservation of both the original bitstream as well as detailed metadata which will enable it to be interpreted in the future (Hunter and Choudhury, 2003). Most digital preservation strategies depend to some extent upon the capture, creation and maintenance of appropriate metadata, i.e., different kinds of metadata will be required to support different digital preservation strategies or digital information types (Day, 2003b).

Metadata must enable access to the intellectual content of the object (whether by migration or emulation), find the object, manage the object, and allow other versions of the object to be

produced. Besides, for maintaining a history of digital object, metadata is a key part of digital preservation (Jana et al., 2009; Lee et al., 2002).

2.3. Metadata for Long-Term Preservation

According to the National Information Standards Organization [NISO] (2004), metadata is defined as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (p.1).

The American Library Association committee on cataloging: description and access presented the formal working definition of metadata in 2004 as it is a structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities (ALA, 2000).

Unless the content of the digital object is described with descriptive, structural and technical and administrative metadata and preservation applications must not be accompanied by metadata, a digital object does not have any meaning to a human being (Groenewald and Breytenbach, 2009). Thus, metadata is critical and plays an important role in digital preservation but complex. Appropriate preservation and metadata management are vital if digital objects are to stand any chance of surviving over time with their intellectual integrity uncompromised (Jana et al., 2009; Lee et al., 2002). The following section will discuss about preservation metadata and related issues.

2.3.1. Preservation Metadata

Preservation metadata is the information infrastructure that supports the processes associated with digital preservation and facilitates the long-term retention of digital information (OCLC/RLG, 2002; NLNZ, 2003). Based on OCLC/RLG and NLNZ reports, preservation metadata has a lot of uses and objectives such as it will be used to store information supporting preservation decisions and actions, document preservation processes, such as migrations, transformations and emulations, record the effects of preservation processes, ensure the authenticity of preservation masters over time and enable objects for which the

institution has assumed preservation responsibility to be identified (OCLC/RLG, 2002; NLNZ, 2003).

Furthermore, according to NLNZ (2003), preservation metadata addresses two functional objectives. These are "providing the institution with sufficient knowledge to take appropriate actions in order to maintain a digital object's bit stream over the long-term and ensuring the content of an archived object can be rendered and interpreted, in spite of future changes in storage and access technologies" (p.3).

According to OCLC/RLG working group on preservation metadata report on 2002 "the importance of preservation metadata has been underscored by the efforts of a number of organizations to develop metadata of this type in support of their own digital preservation activities". Though the community who practiced digital preservation has got immense benefit from this work, still they luck coordination and unable to reach metadata framework for digital preservation that represented a consensus of leading experts and practitioners (OCLC/RLG, 2002, p.1).

As a result, ensuring the long-term preservation of information in digital form will be one of the greatest challenges in the twenty-first century (Day, 2003a).

Day (2003b) stated that even though the generation and maintenance of preservation metadata remains a prerequisite of ensuring the successful preservation of digital objects, it will be assumed to be expensive and "the difficulty of ensuring digital preservation without metadata may mean that it is ultimately a cheaper and more effective option than the alternatives" (p.4).

However, "preservation needs to be addressed throughout the life cycle of digital material in order to be effective" and capturing and managing the appropriate technical and preservation metadata is vital component of digital preservation in the early stages of the lifecycle to guarantee the digital files are not changed in any way (Woodyard, 2004, p.17).

Lavoie and Gartner (2005) point out that though preservation metadata is still a fairly new issue, it has moved quite rapidly from theory to practice. On the other hand, they also

highlight that the movement from theory to practice in preservation metadata cannot be traced as a straight line. According to them, this is due to partly, efforts to develop solid foundations for digital preservation techniques and practices are paralleled by an immediate need to implement capacity to secure the long-term retention of digital materials and currently perceived to be at risk. Hence, according to their recommendation, "it is useful to establish two endpoints for the development of preservation metadata the OAIS Information Model at one end, and the PREMIS working group at the other end with a number of important initiatives taking place in between"(p.9).

Thus, release of the framework prompted interest in moving it towards a more implementable status. As a result, a number of institutions and projects have released preservation metadata element sets over the past few years by reflecting a wide range of assumptions, purposes, and approaches (Lavoie and Gartner, 2005).

2.3.2. The Need for Preservation Metadata

Properly used metadata will help to identify the name of the resource, who created it, who reformatted it, and other descriptive information and provide unique identification and links to organizations, files, or databases which have more extensive descriptive metadata about this resource (this is particularly important in the event that the digital file and its metadata become separated) and also facilitate the long-term access of the digital resources by explaining the technical environment needed to view the work, including applications and version numbers needed, decompression schemes, other files that need to be linked to it, among others. Including preservation, metadata has an important role in digital resource management regardless of which preservation strategy, emulation-based or migration-based, are adopted, the long-term preservation of digital information will involve the creation and maintenance of metadata (Calanag, Sugimoto and Tabata, 2001; Besser, 2000).

Particularly, analyses of the goals of long-term digital preservation have led to a solid understanding of the types of metadata that is needed, i.e., preservation metadata which is "the essential information to ensure long-term accessibility of digital resources" (Dappert and Farquhar, 2009, p.1).

Anderson, Delve, Pinchbeck and Alemu (2009) also stress that “indeed, without appropriate metadata any attempt to ensure the longevity and authenticity of digital objects cannot succeed” (p.16).

Preservation metadata, according to Cordeiro (2004), can“include a wide range of elements for a variety of management purposes and show various levels of detail” (p.11). As discussed by Lavoie and Gartner (2005), preservation metadata is important because:

- **Digital objects are technology-dependent.**

The contents of digital objects cannot be accessed directly by users unlike print books or oil paintings. “Instead, a complex technological environment, consisting of software, hardware, and in some cases network technology, sits between the user and the object’s contents. Rendering and using digital objects requires the availability of this environment, or at least some technically equivalent substitute”. That is why, simply preserve a digital object is not enough. Therefore, it is important especially to carefully document the technological environment of an archived digital object to ensure it remains usable for current and future generations since the constant pace of technological change inevitably makes today’s technologies obsolete (Lavoie and Gartner, 2005, p.6).

- **Digital objects are mutable.**

Lavoie and Gartner (2005) indicate that “digital objects can be easily altered, either by accident or design, with potentially significant consequences for an object’s look, feel, and functionality”. Beyond this, many forms of digital storage media have relatively short lifespan. It raises the specter of “bit rot” i.e., the gradual degradation of stored bits leading to partial or even complete information loss. Lavoie and Gartner underline that “even the act of preservation itself can alter the form or function of a digital object , for example, when an object is migrated from one format to another in order to keep pace with changing technologies”. Due to these and other rationales, it is vital to accompany an archived digital object by metadata documenting its provenance and authenticity in particular, its

salient characteristics at the time of creation, how those characteristics have been altered over time, by whom, and for what purpose (p.6).

- **Digital objects are bound by intellectual property rights.**

For the most part, digital preservation actions are pre-emptive in nature, i.e., seeking to avoid damage rather than to repair it. As Lavoie and Gartner (2005) underlined "once a digital object is corrupted, or the means to access it lost, its contents may be lost forever". Taking these and other things in to consideration, digital preservation must often take place early in the information life cycle and while the material is still under copyright. Thus, it is important to document the intellectual property rights associated with an archived digital object in order that long-term preservation actions can be coordinated with any rights restrictions binding on the object (Lavoie and Gartner, 2005, p.6).

Thus, to sum up, preservation metadata is indispensable since it enables a digital object to be self-documenting over time, and positioned long-term preservation and access, even as ownership, custody, technology, legal restrictions, and even user communities are relentlessly changing (Lavoie and Gartner, 2005).

As indicated by Besser (2000); Alemneh et al. (2002), preservation metadata is an approach to provide sufficient technical information about digital resources. This supports the two primary strategies for preservation of digital resources. The first one is migration, i.e., transfer of digital resources from one generation to a subsequent generation and the second one is emulation, i.e., developing techniques for imitating obsolete systems on future generations of computer.

Effective long-term preservation of a digital object requires further metadata specific beyond description, technical and administrative metadata management. "The type of information that needs to be recorded includes details of provenance, ownership, fixity, an event log to record actions performed on it and any technical and rights information that is necessary to deliver it to the end user" (Gartner, 2008, p.10).

2.3.3. Types of Preservation Metadata

According to Carignan et al. (2006), preservation metadata "overlaps with technical and administrative metadata, detailing important information about the digital file, including any changes in the file over time and management history". For Carignan and his colleagues preservation metadata is useful for digital objects long-term retention and use but does not support discovery or use of digital files. They emphasized that "the object meant to be preserved by preservation metadata is the preservation master digital object itself" (p.8).

Day (2005) also states that:

> It is understood that preservation metadata is to be all of the various types of data that allows the re-creation and interpretation of the structure and content of digital data over time. He continued and indicated that such metadata needs to support an extremely wide range of different functions, including discovery, the technical rendering of objects, the recording of contexts and provenance, to the documentation of repository actions and policies (p.19).

Therefore, according to Day (2005), conceptually, preservation metadata covers the popular division of metadata into descriptive, structural and administrative categories.

Anderson et al. (2009) also indicate in their report on the state-of-the-art in metadata standards and approaches in Europe, "preservation metadata covers administrative, technical and structural metadata, highlighting the somewhat fluid nature of definitions in this field that make it difficult to consistently draw clear boundaries around different kinds of metadata"(p.17).

The NISO framework of guidance for building good digital collections in 2004 also described preservation metadata in such a way that it is a subset of administrative metadata aimed specifically at supporting the long-term retention of digital objects (p.27).

According to PREMIS data dictionary in 2008, preservation metadata spans a number of categories typically used to differentiate types of metadata like administrative (including rights and permissions), technical and structural. Particular attention was paid to the documentation of digital provenance (the history of an object) and to the documentation of

relationships, especially relationships among different objects within the preservation repository (OCLC/RLG, 2008).

Maxymuk (2005) states that:

> There are five overlapping types of metadata that allow an institution to manage, preserve and provide access to digital resources. Descriptive metadata for resource discovery, i.e., it describes the content of the digital object or collection for instance title, author, and subject data. Administrative metadata details management information like location, access control, and copyright. Technical metadata outlines file characteristics such as file format, scanning specifications, file size, software used, quality, and extent. Structural metadata controls the relationship of the parts of a compound complex objects, like the pages and chapters in an e-book or the audio and text in a PowerPoint presentation. Lastly, preservation metadata is used to document the preservation process used to create the digital object or collection (p.147).

Lynch (1999) also indicate that metadata should accompany and make reference to digital objects, providing associated descriptive, structural, administrative, rights management, and other kinds of information.

Preservation metadata represents a repository's best guess as to what information will be necessary in order to make it possible to use a digital object in the future, given the likelihood of changes in technology, format obsolescence, and other risks. The use may differ depending on the nature of the item, the user community, the institution and preservation strategies/techniques (different strategies may demand different pieces of information be recorded) (Caplan, 2006).

Thus, no universal preservation metadata element set and no expectation that there will or should ever be one because of the above mentioned reasons. “Even PREMIS attempts only to be a core set of things that most working preservation repositories are likely to need to know in order to support digital preservation” (Caplan, 2006, p.12).

As discussed above, still it is difficult to draw a clear boundary around what types of information fall within the scope of preservation metadata. However, with a lot of arguments and discussions, consensus seems to have settled around five major areas relevant to preservation metadata and stated as follows.

- Provenance includes “custodial history of the digital object, potentially stretching back to the time of the object’s creation, and moving forward through successive changes in physical custody and/or ownership”.

- Authenticity include “information sufficient to validate the archived digital object is in fact what it purports to be, and has not been altered, either intentionally or unintentionally, in an undocumented way”.

- Preservation activity includes “the actions taken over time to preserve the digital object, and record any consequences of these actions that impact the look, feel, or functionality of the object”.

- Technical environment includes “hardware, operating system, and software applications, needed to render and use the digital object in the state in which it is currently stored in the repository”.

- Rights management includes “any binding intellectual property rights that limit the repository’s powers to take action to preserve the digital object and to disseminate the object to current and future users” (Lavoie and Gartner, 2005, p.5).

To sum up when preserving digital information for long-term, different metadata are important. Descriptive, technical and structural metadata are essential for the description of different digital objects. Preservation metadata is necessary to describe the provenance, fixity, context and rights.

The next section will discuss about OAIS reference model. This is because the OAIS information model provides an abstract framework for thinking about preservation metadata and particularly relevant to describe the metadata requirements for long-term preservation, or in other words, it has direct relevance to the issue of preservation metadata (OCLC/RLG, 2002).

2.4. The OAIS Reference Model

The OAIS reference model initiative was started by the Consultative Committee for Space Data Systems (CCSDS) of the National Aeronautics and Space Administration (NASA) and became an ISO standard in 2003. The OAIS Reference Model is a conceptual framework for a generic archival system which is dedicated to a dual role of preserving and maintaining access to digital information over the long term, defining both a functional model and an information model for preservation activities. It describes the environment in which an archive resides, the functional components of the archive itself, and the information infrastructure supporting the archive's processes (Lavoie, 2004; Caplan, 2006; OCLC/RLG, 2002).

An OAIS is "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 has two major components. The functional model (it has six components: ingest archival storage, data management, preservation planning, access and administration) and information model (CCSDS, 2009, p.1-1).

"The information model broadly describes the metadata requirements associated with retaining a digital object over the long-term. This information model is particularly valuable because it was developed in conjunction with a functional model of a digital archiving system" (Calanag, Tabata and Suginoto, 2004, p.60).

The OAIS reference model has proven to be significantly influential in answering the most fundamental questions concerning preservation metadata, particularly on its scope like what types of information are included in this class of metadata and how is it distinguished from, or overlap with, other classes of metadata. Thus, "it introduces the concept of an Archival Information Package (AIP), which is the digital object being preserved along with its associated metadata" (Calanag et al., 2004).

As Lavoie and Gartner (2005) described:

> The OAIS reference model provides a high-level overview of the types of information needed to support digital preservation, including representation information, preservation description information (which can be broken down into reference,

context, provenance, and fixity information), packaging information, and descriptive information. These information types can be interpreted as the general categories of metadata needed to support the long-term preservation and use of digital materials, and have served as the starting point for a number of preservation metadata initiatives (p.2).

It is represented diagrammatically in Figure 2.1.

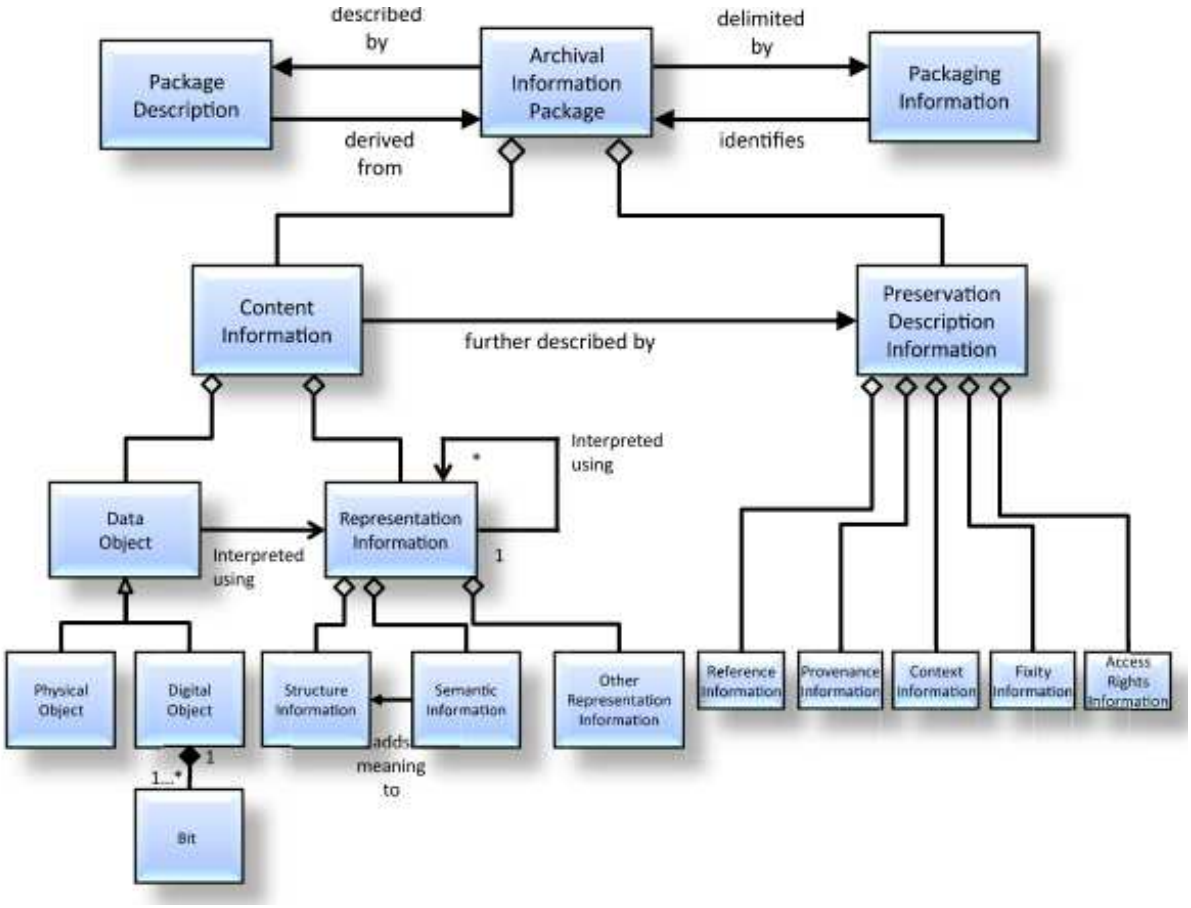

Figure.2.1. Archival Information Package (CCSDS 650.0-P-1.1, 2009, p.4-38)

"The information model defines a number of different Information Objects that cover the various types of information required for long-term preservation". These information objects are described as follow (Day, 2003b, pp.2-3).

- Content Information - the information that requires preservation.

- Preservation Description Information - any information that will allow the understanding of the Content Information over an indefinite period of time.

- Packaging Information – the information that binds all components into a specific medium.

- Descriptive Information - information that helps users to locate and access information of potential interest.

2.4.1. Why We Need Standards?

As it is stated by Knight (2005), the lack of internationally agreed standard on preservation metadata is a key inhibitor to full implementation of a preservation metadata strategy and makes difficult for any organization to commit the resources required to move from the conceptual development to a practical implementation. Even previously, the necessity for common approach to metadata has been noted and acknowledged in the library community for as long as inter-institutional co-operation has been practiced. Particularly, in the 1960s it was recognized “when the MARC standard and AACR cataloguing rules were created to standardize practices into a form which would make full use of the then emerging computing technologies” (Gartner, 2008, p.5).

As recommended by Oltmans and Wijngaarden (2004), though implementation of digital archives has benefited from standardization efforts, e.g., the OAIS reference model and international projects like NEDLIB, the development of permanent access technology is still in its infancy. “Information technology companies have only recently become aware of the problem of relatively short-term accessibility of digital objects”. Currently, some projects have started to develop procedures, tools and methods for the future accessibility of digital objects. However, “these initiatives have been small-scale and scattered. Intensive international co-operation and joined R&D effort is needed” (p.23).

2.5. Preservation Metadata Standards and Initiatives

In the past ten years, there have been a number of initiatives aimed at developing preservation metadata standards. These initiatives are originating from national and research libraries, the archives and records domain, digitization and other projects.

Some of the initiatives have essentially been more closely structured on the OAIS model's definition of an AIP, e.g., the specifications developed by CEDARS and NEDLIB projects while others have been pragmatic responses to the immediate resource management needs of the institution, e.g., the NLA and the NLNZ (Anderson et al., 2009).

These initiatives work on standardizing preservation metadata specification to solve the problem related to preservation and accessibility of digital materials. As a result, they came up with different metadata specifications and played a great role for the development of digital preservation field particularly in the area of preservation metadata. Thus, in the following section different preservation metadata standards and initiatives are discussed.

- **The Research Libraries Group (RLG)**

The RLG's metadata set was aimed at facilitating the preservation of and access to digital images which makes it of limited use for other types of digital objects in this preliminary attempt. The RLG elements illustrate the relationship of preservation metadata to the three broad categories of metadata defined as descriptive, administrative, and structural. Even though preservation metadata can potentially straddle all three metadata types, its focus lies with the latter two. It was not implemented widely but it helped reinforce the discussions to work on preservation metadata not just for images but in general for digital objects (OCLC/RLG, 2001).

- **PANDORA Logical Data Model**

National Library of Australia was one of the first institutions to actually build a digital archive with the establishment of the PANDORA archive of web-accessible materials in 1996. The NLA metadata element set focuses on "information we need out of the system to manage

preservation". Other metadata requirements, such as resource discovery, are not considered. The element set explicitly addresses the metadata needs of different levels of descriptive granularity, assessing the relevancy of particular elements at three different levels: collection, object, and sub-object (file). However, the assumption is maintained that the object is the primary focus of description. No assumptions are made about the specific nature of the processes used to implement preservation (e.g., migration or emulation) - the element set is technology-neutral (OCLC/RLG, 2001, p.17).

- **The National Library of New Zealand**

The National Library of New Zealand developed a metadata schema to support the digital preservation activity of the NLNZ. This metadata schema was seen as significant in virtue of having been one of the first preservation metadata schemas that was actually implemented. This metadata schema includes information about hardware and software environments and also includes information about rights and provenance. The schema recognized the possibility of future changes and revisions to comply with other international standards (NLNZ, 2003).

- **Networked European Deposit Library (NEDLIB) Metadata Elements**

NEDLIB was a collaborative project of European national libraries led by the National Library of the Netherlands. This project defined a functional model based on the OAIS reference model. The functional model is called Deposit System for Electronic Publications (DSEP). The DSEP data model includes the original bit stream of digital publications, metadata, software, and packaging information. It stores and manages metadata separately from the digital object (bitstream). This is because while the bitstream does not change, the metadata for it may be changed frequently (Day, 2001). NEDLIB's metadata specification was explicitly based on OAIS and focused specifically on the metadata needed to address problems of technical obsolescence Unlike CEDARS (OCLC/RLG, 2001).

- **CEDARS Preservation Metadata Elements**

A CEDARS project was a collaborative effort involving UKOLN (The UK Office for Library and Information Networking) and CURL (Consortium of University Research Libraries). The

CEDARS metadata specification explicitly attempted to translate the abstract OAIS model into more practical metadata specifications. It defined preservation metadata sets broadly as "the information required to support meaningful access to the archived digital content and includes descriptive, administrative, technical and legal information". The metadata element set is also intended "to enable the long-term preservation of digital resources and applicable to a broad class of digital objects, in expectation that the typical digital library collection will contain a diverse range of formats". In addition, the specification is wished-for to be independent of the level of granularity at which metadata is assigned (OCLC/RLG, 2001, p.17).

2.6. OAIS to PREMIS - Preservation Metadata from Theory to Practice

The OAIS Model is the common framework guiding a significant proportion of recent international research on digital preservation. OAIS provides a framework to unify the concepts and terminology in the community. Its information model as stated in section 2.4 defines categories for preservation metadata (Dappert and Farquhar, 2009).

Both the earlier framework and the PREMIS data dictionary build on the OAIS reference model. The OAIS information model provides a conceptual foundation in the form of taxonomy of information objects and packages for archived objects, and the structure of their associated metadata. The framework can be viewed as an elaboration of the OAIS information model, explicated through the mapping of preservation metadata to that conceptual structure (CCSDS, 2002). The PREMIS data dictionary can be viewed as a translation of the framework into a set of implementable semantic units. However, it should be noted that the data dictionary and OAIS occasionally differ in terminology usage. This is because of the fact that PREMIS semantic units require more specificity than the OAIS definitions provided and which is expected when moving from a conceptual framework to an implementation (OCLC/RLG, 2008).

2.6.1. PREMIS (PReservation Metadata Implementation Strategies)

Later, an attempt by CEDARS and NEDLIB projects together with the NLA specification to define a preservation metadata schema were taken forward by an international working group called OCLC/RLG and produced a metadata framework to support the preservation of digital objects that uses the OAIS information model as part of its basic structure (OCLC/RLG, 2002).

The PREMIS data dictionary consolidates several earlier efforts to produce conceptual models and concrete metadata dictionaries for implementers of digital preservation services. It define a core set of implementable, broadly applicable preservation metadata elements, supported by a data dictionary and identify and evaluate alternative strategies for encoding, storing, managing, and exchanging preservation metadata(OCLC/RLG, 2008). PREMIS defines five kinds of entities: intellectual entities, objects, agents, events and rights (Caplan, 2009). The following PREMIS data model as shown in Figure 2.2 below shows the relationships of those entities.

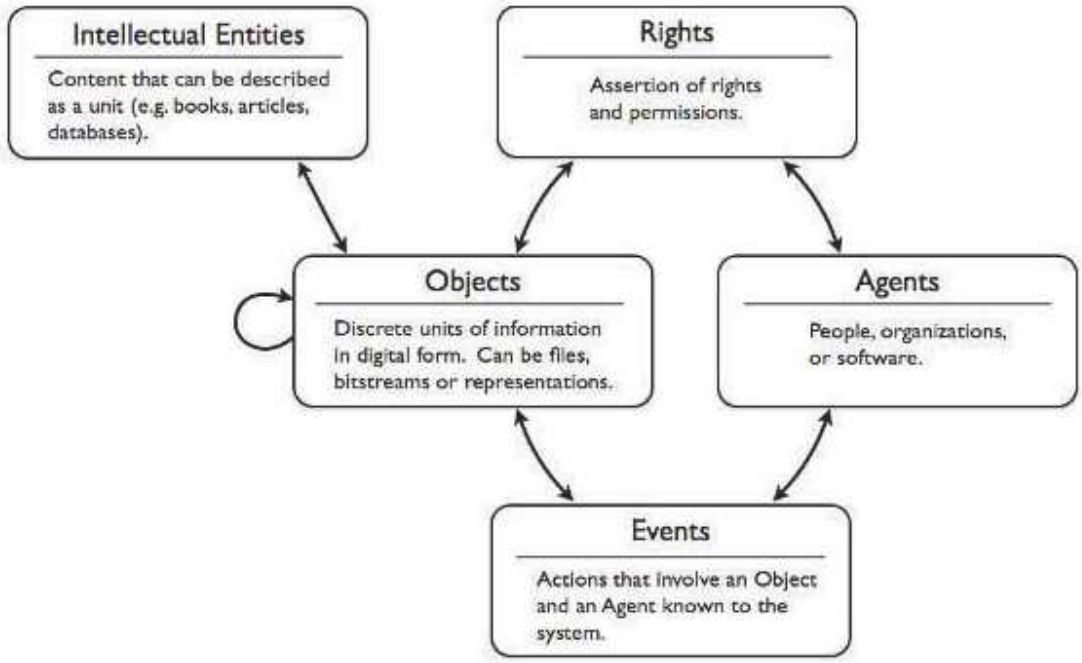

Figure 2.2. PREMIS data model (Caplan, 2009, p.8)

The PREMIS data model shows the relationships of entities. The following section describes the five entities of PREMIS (Guenther, 2009).

Intellectual entity: Set of content that is considered a single intellectual unit for purposes of management and description (e.g., a book, a photograph, a map, a database). It is not fully described in PREMIS data dictionary, but can be linked to in metadata describing digital representation.

Objects: Discrete unit of information in digital form. Objects are what repository actually preserves. According to PREMIS there three types of objects.

- **Representation**: set of files, including structural metadata, that, taken together, constitutes a complete rendering of an intellectual entity.
- **File**: named and ordered sequence of bytes that is known by an operating system.
- **Bitstream**: data within a file with properties relevant for preservation purposes (but needs additional structure or reformatting to be stand-alone file) (Guenther, 2009).

Events: "an action that involves or impacts at least one object or agent associated with or known by the preservation repository". It helps to document digital provenance and can track history of object through the chain of events that occur during the objects lifecycle (OCLC/RLG, 2008, p.130).

Agents: Person, organization, or software program/system associated with an event or a right (permission statement).

Rights: An agreement with a rights holder that grants permission for the repository to undertake an action(s) associated with an object(s) in the repository.

The data model is a useful framework for distinguishing applicability of semantic units across different types of entities and different types of objects. It gives organizational convenience for development and use unlike traditional "flat" metadata management structures.

The PREMIS data dictionary is a comprehensive, practical resource for implementing preservation metadata specification in digital archiving systems produced from an international, cross-domain consensus-building process, and has since become the *de facto* standard for metadata used in support of the digital preservation process (OCLC/RLG, 2008). Now it is in its second version which was released in 2008 and has been widely accepted and plays a key role in creating coherence in the digital preservation metadata community and provides a foundation to support interoperability across systems and organizations (Dappert and Farquhar, 2009). As a result, undoubtedly, PREMIS is the best established schema to deal with preservation metadata (Gartner, 2008).

The group's aims are to develop a comprehensive preservation metadata framework applicable to a broad range of digital preservation activities, and to examine issues surrounding the practical use and implementation of metadata to support digital preservation processes (http://www.oclc.org/research/pmwg/).

From the PREMIS perspective, preservation metadata includes different categories of metadata including rights metadata and provenance metadata, not only limited to technical metadata. It also recommends the use of controlled vocabularies for preservation metadata values and having a central registry of environments metadata which can be shared by different users (OCLC/RLG, 2005).

The PREMIS data dictionary has a set of elements from which a number of separate XML schemas have been derived. These include the object itself (including identifiers, checksums, information on its creation and its relationships to other objects), events (such as its creation and how and when it has been processed), agents associated with its preservation (people, organizations and software), and rights associated with it (Gartner, 2008).

Few studies are conducted how preservation metadata are practiced in various institutions. A study by Woodyard-Robinson (2007) on how institutions implemented PREMIS indicated that "institutions use PREMIS in different ways. None implemented PREMIS 'as is' but instead they used different mechanisms. Some digital preservation software tools such as DROID/ PRONOM, JHOVE, NLNZ metadata extraction tools but only very few automatic

metadata extraction tools from the objects themselves was reported according to her. Regarding representation of the PREMIS metadata, most of the institutions implemented PREMIS using either a relational database management system or XML and some repositories want to keep environment metadata on an external repository. According to her report, still too few implementations of PREMIS having reached sufficient maturity to support firm conclusions on exemplary implementation practices. However, according to Anderson et al. (2009) "compared to other preservation metadata schemas, PREMIS is in the happy position of being widely implemented by libraries and archives" (p.37).

The reasons why I am studying the use of PREMIS in this thesis is because it is the best standard we have and is the most widely applicable across all sorts of institutions, digital preservation contexts and system implementations which is oriented towards practical implementation. It has close links with the OAIS standard. The PREMIS data dictionary supplies a critical piece of the digital preservation infrastructure, and is a building block with which effective, sustainable digital preservation strategies can be implemented. It is the first comprehensive technical specification for preservation metadata produced from an international, cross-domain, consensus-building process (Anderson, Hallahan, Kays and Whitworth, 2009).

2.7. Preservation Metadata and Interoperability

Interoperability gives digital repository systems the ability to exchange metadata information and use the exchanged information. Thus, in the information community, interoperability, i.e., capturing and reusing of metadata, is one of the most important principles in metadata implementation.

According to NISO's (2004) explanation, interoperability is:

> The ability of multiple systems with different hardware and software platforms, data structures and interfaces to exchange data with minimal loss of content and functionality. Using defined metadata schemes, shared transfer protocols, and crosswalks between schemes, resources across the network can be searched more seamlessly. Describing a resource with metadata allows it to be understood by both humans and machines in ways that promote interoperability (p.2).

Managing the growing number of standards currently being developed and implemented, the transfer of metadata or information packages containing metadata to other repositories and services and the capture and reuse of existing metadata are all aspects of interoperability that will need to be addressed by digital repositories (Day, 2003b).

However, metadata standards and formats that have been developed to support the management and preservation of digital objects raise several questions about interoperability (Day, 2003a). Some of them are the following.

- Will repositories be able to cope with the wide range of standards and formats that exist?
- Will they be able to transfer metadata or information packages containing metadata to other repositories?
- Will they be able to make use of the 'recombinant potential' of existing metadata? (p.4).

As Day (2003a) suggests that “the precise way in which future intra-repository co-operation will work remains to be worked out in detail and it seems likely that repositories will need to exchange information packages or metadata with other repositories”. On Day’s suggestion developing standard exchange-formats, possibly based on existing standards like METS might be a solution or in other contexts, “the exchange of information packages between repositories may become dependent on the sophisticated conversion facilities that could be supported by registries, e.g. of file formats or metadata” (p.8).

2.7.1. Metadata Registries

Registries are starting point for successful data sharing; they offer an authoritative place to find resources for exchanging or reusing data for institutions. They provide metadata elements maintained by an organization or community of interest. The objects referenced in a registry can include entire standards or specifications, components of the standards or specifications, XML schemas or schema components, software components, data elements, database structures, or related documentation ('Metadata Rules', 2003). Thus, the aim of preservation

metadata standards and initiatives like PREMIS is to have a standardized schema that can be acceptable for all parties in the field of digital preservation and recommends for having a central registry of metadata which can be shared by different users and therefore metadata registries could have tremendous value for its effectiveness.

For the management of metadata, registries, which are central locations where metadata definitions are stored and maintained, are an important tool in providing information on the definition, origin, source, and location of data (Anderson et al., 2009; NISO, 2004). "The metadata registry provides an integrating resource for legacy data, acts as a lookup tool for designers of new databases, and documents each data element. Registration can apply at many levels, including schemes, usage profiles, metadata elements, and code lists for element values" (NISO, 2004, p.11).

Day (2003a) argued that "metadata registries may be a useful way of helping to manage this diverse metadata within a digital preservation system, and to preserve aspects of its context and original functionality. Registries could also contain authoritative mappings between different standards, thereby helping to facilitate the exchange of metadata or information packages between repositories and end users" (p.6).

This may be important because nowadays a wide range of metadata standards have been developed that have relevance to digital preservation.

To continue to work towards greater convergence and interoperability, the preservation community has faced challenges as mentioned in 2.7. However, when it comes to metadata, there is a considerable common involvement in content creation and networked service delivery as well as a widespread desire to reduce or avoid completely any duplication of effort which has given support to the development of metadata registries (Anderson et al., 2009; NISO, 2004).

In general, according to Day (2003a), a metadata registry component of a digital preservation system would have the following basic functions.

- It would act as an authoritative source of information.

It contains information about the metadata terms and vocabularies used within the repository. “Wherever possible, metadata would be kept in its original format and the registry would provide information on how it should be interpreted and gives information on its context. The repository can add (or import) information on new metadata schemas when they become available” (p.5).

- It helps to support the ingest process.

It can be used “to support the ingest process by providing mappings that could be used to help populate the metadata used by the repository itself” (p.5).

- It supports the export of metadata.

The mappings maintained within the registry “support the export of metadata or information packages from the repository” (p.5).

2.8. Chapter Summary

This chapter reviewed the literature relevant to this study. It discussed the issues and challenges of digital preservation and looked at topics related to preservation metadata such as digital objects in connection with metadata and preservation problems as well as metadata in preservation for long-term accessibility of digital objects. In this discussion, it is observed that ensuring the long-term preservation of information in digital form is one of the greatest challenges of the information society. This is because new types of hardware, computer software applications and file formats supersede each other and make digital information inaccessible in the long-term. It is also indicated that more than any other media, digital objects requires detailed metadata to ensure its preservation and accessibility for future generations.

The literature on the OAIS reference model, preservation metadata standards and initiatives together provide a valuable conceptual framework, general understanding of preservation metadata as a whole as well as support the transformation of theory to practice (e.g., from OAIS to PREMIS) by describing and explaining the preservation metadata development

processes for implementation. This discussion indicated that the OAIS reference model has proven to be significantly influential in answering the most fundamental questions concerning preservation metadata. It has been a starting point for preservation metadata standard development that has resulted in the PREMIS standard. However, there are still too few implementations of PREMIS having reached sufficient maturity to support firm conclusions on exemplary implementation practices.

Literature on interoperability and metadata registries were also discussed. Since the number of standards currently being developed and implemented are growing, the management and transfer of metadata or information packages containing metadata to other repositories and services as well as the capture and reuse of existing metadata are all aspects of interoperability and metadata registries that need to be addressed by digital repositories.

CHAPTER THREE: METHODOLOGY

This chapter discusses the methodology used in the study. It specifically explains the research approach and the research strategy. It also explains the data collection techniques, sampling strategy, data analysis and ethical considerations of the study.

3.1. Research Approach

3.1.1. Qualitative Approach

This thesis used a qualitative approach to study the extent of implementing standard preservation metadata into practice at memory institutions. The choice of one method to employ over the other is dependent upon the nature of the research problem definition together with the kind of information that is needed. The qualitative approach was the preferred solution for this study because the nature of the research questions required that the topic should be explored in detail for which descriptive and detailed data needed to be collected.

Qualitative approach was suitable for this study as, according to Patton (2001), qualitative research uses “a naturalistic approach that seeks to understand phenomena in context-specific settings, such as real world setting where the researcher does not attempt to manipulate the phenomenon of interest” (p.39).

As Denzin and Lincoln (1994) explained that qualitative researchers study things in their natural settings, attempting to make sense of or interpret phenomena in terms of the meanings people bring to them. Creswell (1994) also underlined that in qualitative research, the researcher builds a complex, holistic picture, analyzes words, reports detailed views of informants, and conducts the study in a natural setting.

This research was interested to describe and explain on actions in local practices of preservation metadata. Thus, the philosophical stance for this study is a pragmatic approach which is used “to determine the meaning of words, concepts, statements, ideas and beliefs. It implies that we should consider what effects which might conceivably have practical

bearings. Then our conception of these effects is the whole of our conception of the object" (Peirce (1878) as cited in Johnson and Onwuegbuzie, 2004, p.17). Hence, the pragmatic approach helps to practice contributions and active participation in testing and exploring new ways of working.

Qualitative research design is iterative rather than linear, i.e., data collection and research questions are adjusted according to what is learned. In other words, qualitative study is typically more flexible than quantitative study. It allows greater spontaneity and adaptation of the interaction between the researcher and the study participant (Mack, Woodsong, Macqeen, Guest, and Namey, 2005).

Thus, in the context of this research, the researcher used such approach to move back and forth between design and implementation to ensure correspondence among research question formulation, literature, data collection strategies, sampling strategy and analysis. In addition, it helped the researcher as a verification strategy, i.e., to verify facts or fill gaps that had been created along the research process and to identify when to continue or modify the research process in order to achieve reliability and validity. In favor of this idea, Srivastava and Hopwood (2009) argue that the visiting and re-visiting of the facts helps to verify and also gain a new insight and helps to refine the focus of the research. They extended their argument by stating that an iterative process or qualitative data analysis should be considered as a reflexive process, not as a repetitive task because it is the key to sparking insight and developing meaning.

Hence, it was necessary to use the qualitative method for studying preservation metadata, which is rich in semantics and to make sure that all the meanings of elements get accounted for. It was also because of the research questions that were framed as open-ended questions that can support discovery of new information and the language barrier (the respondents were Estonian, study in English). Thus, it was better to approach them face-to-face for better understanding of the practice of preservation metadata and to explain the questions as needed to gain better ideas on the facts of the phenomenon and to get more in-depth qualitative information.

3.2. Research Strategy

3.2.1. Case Study

In this research, a case study was employed as a research strategy. This research strategy is generally preferred when answering "how" and "why" questions about a particular topic (Yin, 2009). Accordingly, this method will enable us to understand the complex real activities as well as to investigate an area of interest in depth and therefore is particularly appropriate. As described by Patton (1987), case studies become particularly useful where one needs to understand some particular problem or situation in great-depth, and where one can identify cases rich in information.

According to Noor (2008), case study is preferred when the questions are targeted to a limited number of events or conditions and their inter-relationships. In favor of this and in explaining what a case is, Yin (1989) suggests that the term refers to an event, an entity, an individual or even a unit of analysis. It is an empirical inquiry that investigates a contemporary phenomenon within its real life context using multiple sources of evidence.

Hence, case study was suited for studying this research problem because no thorough analysis exists yet in the literature and I needed to collect my own data because the problem is very practical and need to conduct almost a "field study" to understand the issues involved in implementing the theoretical metadata standards.

Principally, Anderson (1993) describes case studies as being concerned with how and why things happen, allowing the investigation of contextual realities and the differences between what was planned and what actually occurred. He also added that case study is chosen as a strategy because it is not intended as a study of the entire organization rather it is intended to focus on a particular issue, feature or unit of analysis in order to understand and examine the processes and activities in organizations.

Accordingly, the unit of analysis for this case study was "*preservation metadata*". In this case study, preservation metadata was assumed as a contemporary phenomenon that had been initiated and opened for discussion by and within digital preservation community especially in

libraries and archives considering for long-term accessibility of digital collections. Therefore, case study, as a research strategy, was best suited to examine such interventions of memory institutions in implementing standard preservation metadata into practice in their digital preservation process considering their context, i.e., goals and settings. This was supported by Yin (2009, p.18) who defined the case study research strategy as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context; especially when the boundaries between phenomenon and context are not clearly evident".

3.3. Data Collection Technique

The primary data for the analysis were collected through interviews. Secondary data were obtained through document analysis by gathering information from the institutions' websites, documentation about their preservation metadata and other relevant documents commended by the interviewees.

Interviewing is one of the most common methods for collecting data in qualitative research. It allows participants to provide rich, contextual descriptions of events. Interview as a data collection technique is also one of the most significant sources of obtaining case study information (Yin, 2009). Glesne and Peshkin (1992) also state that data collection methods like interviews - are dominant in the naturalist paradigm.

According to Gray (2004), if the objective of the research is largely exploratory, the aim of using interviews as a means of gathering in-depth information was to probe for more information and attain highly personalized data. This allowed the researcher to probe for more detailed responses where the respondent was asked to clarify what they had said.

A semi-structured interview technique was chosen to collect data from metadata experts/specialists about the implementation of standard preservation metadata in their respective institutions. Semi-structured interview as a data collection technique for this study was chosen because they are non-standardized and are often used in qualitative analysis (Griffee, 2005) and it also offered sufficient flexibility to approach different respondents differently while still covering the same areas of data collection (Noor, 2008).

The interview questions were compiled in such a way that the researcher identified different themes (for example, what preservation metadata standards, preservation strategies, metadata categories, tools used, about problems and challenges, etc) based on the research problem and questions while reviewing different literature for the study For the most part, the PREMIS data dictionary and works related to it was used. After the questions were designed, they were reviewed with the supervisor. Based on the inputs from the review the questions were redesigned. Questions are available in appendix A.

According to Griffee (2005), a semi-structured interview means that questions are predetermined, but the interviewer is free to ask for clarification, can change the order of the questions can give explanations or leave out questions that may appear redundant. So, the main job is to get the interviewee to talk freely and openly while making sure you get the in-depth information on what you are researching.

Semi-structured interview is the most adequate tool to capture how a person thinks of a particular domain. Its combination of faith in what the subject says with the skepticism about what she/he is saying, about the underlying meaning, induces the interviewer to go on questioning the subject in order to confirm the hypothesis about his/her beliefs (Honey, 1987).

This research also used documentary evidence to supplement as well as to compensate for information gathered from interviews. Additionally, documents provide guidelines in assisting the researcher with his inquiry during interview.

Thus, the researcher conducted interviews (face-to-face interviews for the two institutions and an interview via email with follow-up for the third institution). The researcher travelled on April 14, 2010 to Tartu, the second biggest city in Estonia, in order to conduct the interview at the National Archives of Estonia and the interview took around 2 hours and 30 minutes. In the case of National Library of Estonia the interview was also conducted face-to-face on April 22, 2010 and it took nearly 1 hour and 15 minutes. These interviews were all recorded on Olympus Digital Voice Recorder and loaded to the computer for the sake of expediency for transcription. A written note was also taken to complement the recordings. In the case of the third institution because of geographic distance and time of inconvenience to the respondents,

the interview was conducted via email with follow-up from April 5 to 26, 2010 and the researcher was satisfied with the data collected with this technique too. However, a face-to-face interview was found much more informative than the e-mail, and that seeing metadata in action at the memory institution was an important aspect, not just reading what someone tells me they have in place.

3.4. Sampling Strategy

Understanding what purpose research will serve should be a decisive factor in selecting a qualitative sample. Qualitative researchers perform sampling with a purpose (Byrne, 2001) and 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 (Gray, 2004).

In practice, qualitative sampling usually requires a flexible and pragmatic approach since qualitative research is an iterative process as stated in section 3.1.1, i.e., it is permissible to change the recruitment strategy, as long as the proper approvals are obtained (Marshall, 1996).

Purposeful sampling is the most common sampling technique that the researcher actively selects the most productive sample for qualitative study to answer the research question and it is used generally in case study research. This can involve developing a framework of the variables that might influence an individual's contribution and will be based on the researcher's practical knowledge of the research area, the available literature and evidence from the study itself (Marshall, 1996). Thus, purposive sampling is used in this research as a sampling strategy.

Therefore, institutions that practice digital preservation, the National Library of Estonia, the National Archives of Estonia and the National Library of Wales, were taken to see to what extent the theoretical metadata standards were implemented. There were several reasons for selecting these institutions into the sample. First, they already have digital collections and are practicing digital preservation; they also have a legal obligation to preserve digital materials. The experience of managing digital collections of the memory institution was taken into

consideration. These institutions have practical experience with implementation of digital preservation and metadata management and therefore it is good to study the preservation metadata implementation with them. The study also applied to contrast the preservation metadata practice of the library vs archive and the selection was deliberate to study if any differences exist and what they might be. The third institution was just used for verification purposes and deliberately chosen from a different country to act as a comparison for the two from Estonia. The researcher also contacted several other institutions in the region without result, but it was the National Library of Wales that volunteered to cooperate with this study. Second, the choice for the first two institutions was influenced by the geographic proximity (the digital archive of the National Archives of Estonia is located in Tartu, the second biggest city in Estonia; the National Library of Estonia is located in Tallinn, the capital of Estonia).

The names of respondents were initially determined in each memory institution with their job responsibilities, position and involvement in the subject studied, i.e., preservation metadata. However, respondents were selected from each memory institution on the basis of the researcher's individual judgment, where permitted, and in consultation with the head of digital preservation unit of each memory institution. The selection was done on the ground that the respondents could provide the necessary information needed for the research (Noor, 2008). A total of six metadata experts/specialists were selected for the interviews: three from the National Archives of Estonia (the interview was held in a group), one from the National Library of Estonia and two from the National Library of Wales (the interview via email was done on both persons separately). The choice was based on the experts' job responsibility and position they have in the digital preservation unit and the availability of metadata experts/specialists in each memory institution. Among the kind of job and position they hold are the deputy director of the digital preservation unit, metadata expert/ specialist, software designer, project manager and database administrator. Based on this and other given information the researcher focused on the metadata experts/specialists and the deputy director of the digital preservation unit who has connection with the metadata management for the interview.

3.5. Data Analysis

The data analysis of this research has followed the principle of qualitative data analysis process. Data analysis in qualitative research can be defined as consisting of three concurrent flows of action: data reduction, data display, and conclusions and verification. These flows are present in parallel during and after the collection of data (Miles and Huberman, 1994). Hence, qualitative data analysis process is not linear rather it is iterative and progressive. Their relationships and data collection efforts are depicted in figure 3.1.

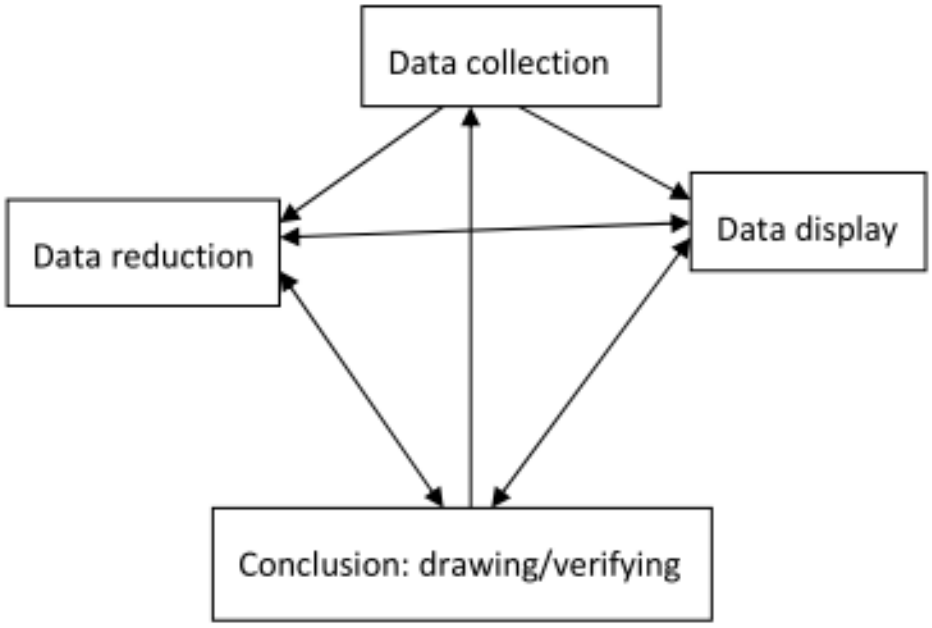

Figure 3.1. Components of data analysis (Huberman and Miles, 1994)

For this study, the data collected (i.e. interview transcripts, written notes - notes taken at the time of the interview to complement the transcripts and data from documents) was reduced and then organized and displayed so that conclusions (i.e. regularities, patterns, differences/similarities, explanations, propositions) could be drawn from the data. The following sections describe the process of data analysis with the use of the above presented data analysis model.

Data Reduction

The face-to-face interviews of this research were recorded on Olympus Digital Voice Recorder and loaded to the computer for the convenience of listening for transcription. A written note was also taken to complement the recordings. These data that appeared in written notes and transcriptions had been selected, simplified, abstracted and transformed in a way that helped to sharpen, sort, focus, discard, and organize the data in a way that allowed for conclusions to be drawn and verified. This was more suitable to do within-case analysis since this study as a single case study relied on it.

Data Display

The reduced data was taken and organized by themes and compressed so that conclusions could be more easily drawn. The data was extended in a piece of text and tables that provided a new way of arranging and thinking about the more textually embedded data that allowed me to extrapolate from the data enough to begin to discern systematic patterns and interrelationships.

Conclusion Drawing and Verification

At this stage, I was stepping back to consider what the analyzed data mean and to assess their implications for the questions at hand. For verification purpose, I revisited the entire collection of data from interviews and data from documents as many times as possible to cross-check or verify these emergent ideas. In addition, as a single case study, the data analysis of this study also relied on within-case analysis.

3.6. Credibility Strategy Employed in the Research

Establishing the credibility of research findings can be achieved through various strategies. Shenton (2004) stated that qualitative methodology applies iterative questioning, frequent debriefing sessions and tactics to help ensure honesty in informants as a means of establishing credibility on the result of research. In this study, iterative questioning had done in data collection dialogues and also frequent debriefing sessions had carried out between the

researcher and superior (for example, the interview questions were discussed between the researcher and the supervisor frequently). Moreover, the researcher used tactics to help ensure honesty in informants when they were contributing data (for example, participants encouraged to be frank). Therefore, adoption of these techniques along with the data sources (metadata experts with supplementary documents) was an option to ensure credibility.

3.7. Ethical Considerations

The purpose of the interview was explained to the interviewees. Interviewees approved that their responses can be used in the context of this research. The anonymity of the interviewees was maintained and hence the information acquired from interviews at NLE, NAE and NLW was used with proper care. The researcher took and discussed responses into the appropriate context. When the results of this study were reported, it was represented accurately what was got from the interviews and documents.

3.8. Chapter Summary

This chapter has provided a detailed discussion of the methodology used in this research. It began with a discussion of the rationale for the chosen methodology. A qualitative study with a pragmatic approach is probably the appropriate way to answer the research questions of this study. A case study research strategy was chosen. This study used semi-structured interview and document analysis as data collection method. Semi-structured interviews were the appropriate method to obtain rich and in-depth information from the respondents and document analysis helped to get supplementary information. The sampling strategy, the method of data analysis and ethical issues were then discussed. The following chapter presents the analysis and findings of the research.

CHAPTER FOUR: ANALYSIS AND FINDINGS

4.1. Introduction

This chapter presents and discusses what was found from the study. It starts by presenting some background information about memory institutions and their digital archives to make the discussion easy to follow. It continues presenting the digital archiving process like how materials were ingested, preservation strategies adopted, software and tools in use, etc in each of these memory institutions.

Furthermore, it discusses the practice of preservation metadata such as categories of metadata used and its management, how metadata is obtained and stored, and interoperability issue in each memory institution. It also presents the level of application of PREMIS entities and metadata standards/schema within the memory institutions.

The comparison of preservation metadata practice of national libraries and archives is also discussed. Problems and challenges faced in the process of digital archiving are investigated. Finally, it concludes with discussion.

4.2. Background Information on the Memory Institutions in the Study Sample

4.2.1. The National Library of Estonia

In recent years Estonia has been confidently establishing itself as a strong IT power. It was the first country in the world to run governmental e-Elections. Its taxpaying system is largely internet-based so that over 90% of Estonians submit their tax declarations online. Also, there is e-Banking, where 98% of all transaction is carried out online. As for the public sector, Estonia has e-Schools and e-Police. All of these are strong indicator of the nation's commitment to deploying IT technologies for optimizing and enhancing processes in every aspect of life. One of the beneficiaries of Estonia’s e-movement is the National Library of Estonia. The NLE was established in 1918 and offers thousands of published materials for

public access. The Library launched electronic information services in 1993. Since then, the amount of e-services on its website has increased manifold, so the primary task for NLE during past few years has been to create and increase online access to the services to meet growing requirements of the public. For instance, in 2006 NLE joined the EU co-funded Books2eBooks project which involves developing a new online service of ordering digital books. Under this project, an e-book is produced on request and is delivered to customer in PDF with a full-text search enabled (NLE, 2008).

According to the library website, the NLE concentrates on developing information environment inside and outside its organization, focusing on every social structure with information needs. The Library, through its complex activities, plays an important role in the Estonian cultural life. Besides that the library has developed better knowledge of its role in information society and adapted its activities to the needs of the changing society. Currently, the library's main goal is to develop user-oriented library and providing open access to its collection, targeting the services, widening access to the collections, implementation of services based on new information technology, and improving service quality are considered equally important in developing library services (http://www.nlib.ee/584).

In general, its main objective is to help along the development of Estonian republic and each person by cooperating with other information and library organizations in collecting, preserving and making accessible information resources (http://www.nlib.ee/17606).

A tremendous project was launched by NLE in 2006. The Library began to register and store Estonian publications in the digital archive called DIGAR. Printed materials–newspapers, magazines and books, accumulated during 80+ years since the library opened, make up the largest part of its assets. Due to wear and tear, the library stopped lending newspapers to public. Because of their irreplaceability, it was also important to retain the materials in their original form. Microfilm came in helpful for a while, but became obsolete due to its inefficiency and inconvenience: no text search, no indexing, and gets scratched (NLE, 2008).

4.2.1.1. DIGAR (Digital Archive of National Library of Estonia)

Digital archive DIGAR–electronic collection of national documents was developed under the EU project *ReUse*. It contains online publications, pre-print files of periodicals, websites, books scanned by the National Library, and Estonian newspapers published during 1800-1944. The archive has 170 depositors. In 2007, the Library began to systematically collect, archive and describe books, journals and maps from the Library's own collections, as well as Estonian-language publications with non-Estonian domain. Additional value to the scanned texts is provided by cross-text search via optical character recognition (OCR). The creation of the digital archive constitutes the implementation of the whole integrated process: collecting, processing, preserving and making accessible of the national publications. This is one of the main tasks of the National Library of Estonia (NLE, 2007).

DIGAR operating agreement is not mandatory. A contract is awarded, if the publishers have the desire to determine their edition of temporary restrictions on use and the easiest way to comply with an agreement with the e-contract form in DIGAR website. DIGAR contains publications of Estonian state agencies, local government, and scientific, and educational publications. If the publisher does not harm the copyright and related rights in the interests of the owner, the archive only permits the depositor to manage its digital files in accordance with the contract. The contract is awarded to the National Library of Estonia (archive manager), and the surrendering of the original file (depositor). The agreement offers the archive manager for long-term archiving service to ensure the authenticity and integrity of the archived version and allow access to the archived edition (http://digar.nlib.ee/). Accordingly, in the later section, this study is going to discuss the metadata implementation in DIGAR.

4.2.2. The National Archives of Estonia

The National Archives of Estonia (NAE) is the centre of archival administration in Estonia. The main task of the National Archives is to ensure preservation and usability of society's written memory, documented cultural heritage for today and future generations. On the other hand, the National Archives guarantees the protection of citizens' basic rights and duties and the transparency of the democratic state through the holding and preservation of authentic

documents. It is a government agency in the domain of the State Chancellery, which includes Estonian Historical Archives, Estonian State Archives, Estonian Film Archives and six regional Archives: Harju, Lääne,Lääne-Viru, Saare, Tartu and Valga. All Estonian public archives, except Tallinn City Archives and Narva City Archives, belong to the National Archives system. The National Archives deals with the archives economic and development activities with the support of Administrative Bureau and Development Bureau. The digital archive management issues are in the domain of Digital Preservation Bureau on national and international levels (http://www.ra.ee/en/about_us/&i=1).

The establishment of the National Archives system started already at the beginning of the Republic of Estonia. The current Historical Archives was established in 1920 in Tartu as the holding place for historically significant institutions documents and the State Archives in Tallinn as the keeper of documents of active institutions. The National Archives collect and preserve archival records that document Estonian history, culture, statehood and society, regardless of the time or place of their creation, or the nature of the medium. Its vision is to ensure the durability and the use of information reflecting Estonian society, in the present and the future (http://www.ra.ee/en/about_us/&i=1).

4.2.2.1. Digital Archiving in NAE

As it is stated above, NAE is a government agency and a system of state owned public archives, including the Film Archives. It is the leading authority in Estonia in the field of long-term preservation, both for born-digital and digitized content. NAE is officially acknowledged as the Estonian competence centre for mass digitization from microfilms. It is actively participating in the process of issuing guidelines regulating the use of governmental datasets, document management systems, use of digital media, metadata generation and appraisal processes. NAE participated in the EU co-funded QVIZ (Query and context based visualization of time-spatial cultural dynamics) and PROTAGE (PReservation Organizations using Tools in AGent Environments) projects. To reach the goals of PROTAGE, the NAE is collaborating with the NLE. The NLE and NAE have been closely collaborating in developing the Estonian national strategy for digital cultural heritage and are taking

collaborative efforts to standardize and simplify the creation, description and use of Estonian cultural heritage (PROTAGE Project, 2010).

The task of digital preservation as a branch of archive management in NAE is to ensure the permanent preservation of digital data despite the changes in society and technology. According to the NAE, its digital archives focus mainly on the following specific issues (http://www.ra.ee/en/digital-preservation/&i=6).

- How to ensure the usability of digital files when the software used to create these is no longer available and usable?
- How to ensure the usability of a data carrier in a situation where the necessary hardware for reading the data is no longer available and usable?
- How to ensure that the data on a data medium is not lost following the physical destruction of the data medium?
- How to ensure the understanding of the content in case of major changes in society and ways of thinking?

As principle, the National Archives digital archives are working based on the following.

- The preservation of data is ensured via double preservation in various locations and thorough back-up and recovery procedures.
- The reproduction of data is ensured with the help of migration policy: when the file format software (or hardware) is not supported anymore, the file is migrated into a new format, for which a "regular user" has the necessary hardware and software in their computer.
- The creation and management of thorough descriptions (or metadata) helps to find and understand the data (http://www.ra.ee/en/digital-preservation/&i=6).

Thus, this study is particularly concentrating and discussing on the later sections about the aspect of practicing the preservation metadata for long-term preservation in the digital archiving process of NAE.

4.2.3. The National Library of Wales

The National Library of Wales (NLW) is the national legal deposit library of Wales, located in Aberystwyth. In 1873, a committee was set up to collect Welsh material and houses it at the University College in Aberystwyth. Leading Welsh people and Members of Parliament worked hard to establish a National Library and a National Museum. Aberystwyth was selected as the location of the Library partly because a collection was already available in the College. Both the Library and Museum were established by Royal Charter on the same day, 19 March 1907. In 1996 a large new storage building was opened, and in recent years many changes have been made to the front part of the building to make it more open and welcoming. A new Royal Charter was granted in 2006 (http://www.llgc.org.uk/index.php?id=6).

The NLW contains a mountain of knowledge about Wales and the world–millions of books on every subject, thousands of manuscripts and archives, maps, pictures and photographs, films and music, and electronic information. It recognized that electronic developments present a major challenge to the traditional role of libraries and those significant changes are needed in order to deal with new information technology and digital documents. In March 2003 NLW adopted its first digital preservation policy and strategy (http://www.llgc.org.uk/index.php?id=6).

4.2.3.1. NLW Digital Archive

The NLW digital archive is the computer hardware and software that stores digital data for the long term, and it is this store is used for digital preservation. The digital archive has been developed since 2003 and at present it stores primarily digital image files created by NLW's digitization programme. It also accommodates audio recordings as well as other categories of digital materials. The data archive makes the NLW of Quantum Amass technology to store large amounts of data in an easily retrievable and automatically upgradeable form. The aim of digital preservation activities in NLW is to preserve and maintain Welsh and relevant non-Welsh library and archive materials to ensure they are available for current and future use. It is expected that most materials selected for retention by NLW will be preserved in their

original formats, whether they are digital or non-digital materials. NLW also seeks to help others preserve Welsh information resources for which they accept responsibility (NLW, 2008).

4.3. Digital Preservation Process at the Memory Institutions

As we learnt from above the selected memory institutions are practicing digital preservation in their digital archive. It is obvious that preservation metadata is implemented in the process of digital preservation and hence discussing the process of digital preservation at memory institutions has momentous value for having comprehensive understanding of preservation metadata implementation. Thus, this section is going to discuss how memory institutions were obtaining materials to their digital archive, what preservation strategies they were adopting, what kind of software and tools were used to support the tasks of preservation of digital materials and their similarities, difference and implications.

The respondent from the NLE mentioned that materials are obtained to the digital repository in three ways. These are:

- In-house digitized materials are ingested to the repository.
- Through harvesting from the web, and
- Publishers cooperate with the library to ingest their pre-print material to the digital repository.

In the last case, each provider gets a user account and is given access to FTP server and web interface. Files have to be uploaded by using a FTP client. Each publisher has its own FTP-directory. Web-interface for content providers (publishers) helps to organize uploaded files. Thus, providers add the minimal metadata to the digital object like descriptive metadata (title, author(s), and publication year), administrative metadata (copyright statement and access restriction) and linking uploaded files to object and assigning linked files properties (comment, sorting order, access restriction) and then send digital object for processing.

The NAE accepts materials from agencies. They can be private companies and/or persons that fulfill the criteria of the NAE. Most of ingest and pre-ingest actions have to be done in the

agencies. The archival requirements for datasets mostly concentrate on the metadata and archival formats. A system called Universal Archiving Module (UAM) is developed in the National Archives. Those agencies installed UAM to generate semi-automatically archival metadata out of the record management metadata and allow restructuring of the records and validation of metadata and file formats. This supports coordination and approval procedures between data producers and the National Archives. Usually, the procedure begins with the agency asking for permission to transfer archival records to the National Archives. After that a timeframe has to be agreed on and the agency can start to prepare the transfer. The agency has to provide most of the information for archival description and the records transferred have to be provided according to the archival guidelines (archival materials have to be used, metadata has to be provided according to certain rules, etc.). The appraisal group/archival inspectors of the National archives will decide based on the functions and higher level general descriptions of the digital object. In the National Archives the archival descriptions are put into the Archival Information System (AIS), the metadata is validated and the data is sent to off-line preservation.

In the NLW, legal deposit, digital visual and audio collections are created, received or recorded off-air as part of the collections of the National Screen and Sound Archive of Wales, digital surrogates (including preservation master digital copies) of analogue material in NLW collections resulting from digitization programmes, electronic publications received under voluntary legal deposit and published on physical carriers such as diskettes and CD-ROMs, archiving of websites, archival collections which comprise of electronic elements (e.g. files on physical carriers) are donated to the library.

The process of ingesting materials to the digital archive of the national archives is unlike to the national libraries. The National Archives use a system called UAM for agencies to automate the ingesting process and the material should get the approval of the appraisal group called archival inspectors. However, the national libraries obtain materials to their digital repository mainly from results of their digitization program, publishers and voluntary legal deposits. Thus, the national archive have much tighter control over setting requirements and conditions for the quality of material it ingests than national libraries and this is perhaps

because of their mission difference. At this process, the digital archives make agreement with the owner of the material and the digital archives have got a chance to obtain some descriptive metadata (for example, author, title, data of publication, etc) and administrative metadata like rights and permission issues etc which are one of the major components of preservation metadata.

4.3.1. Preservation Strategies at Memory Institutions

Two institutions, NAE and NLW, are using migration as a preservation strategy while the NLE makes only conversion of file to PDF for text and TIFF for image documents. Currently, the NLE do not archive audiovisual materials. The NLE is researching and watching the situation and making two meetings (October and April) every year to discuss the issues.

According to the NLW, the library is continuing to use and develop appropriate preservation strategies for differing formats of digital material. The formats of the digital objects are assessed to decide upon the most effective preservation strategy. Decisions are based upon the intellectual content, physical medium and the perceived use of objects. Refreshing, migrating and emulation are still seen as appropriate strategies for digital preservation, depending upon the circumstances. However, the NLW currently devise migration pathways for different formats of material, depending upon the evaluation of their significant properties and continue with refreshing of data to ensure verification of data.

NAE believes migration is the simplest and widely accepted strategy and it is the best strategy for it for the time being. NAE migrates the digital objects to PDF for text, TIFF and PNG for images, BTR for audio and PG for video and CSP for databases. In the NAE, the actual computer files are embedded in the XML in transfer package, i.e., there is one transfer package for one record which includes different binary files. The XML files of the preserved files of Archival Information Packages (AIPs) are found separately. In the case of migrated files the metadata is not embedded because of its own drawback. The NAE is also looking for emulation and others strategies and analyzing their merits and demerits. Whenever the need arise, the NAE is ready to use them.

Respondents explained that they are making or will make the best decisions they can with an eye toward the future even though the future is uncertain. All institutions are researching for the better strategy based on their goal and ready to use them when it is necessary if the strategy is going to be feasible for their system. One institution, NLE, indicated that none of the converted formats (PDF, TIFF) are approaching obsolescence, so it is not urgent but they are doing technology watch and it is a hot discussion issue in the institution.

4.3.2. Software and Tools in Use at Memory Institutions

A variety of software and tools are utilized in these three institutions for different tasks in order to assist in the preservation of its digital object, e.g., for capturing preservation metadata, format identification, validation, and characterization of digital objects. Those software and tools are depicted in the following table.

Table 4.1 Software and tools in use at memory institutions

Software and Tools	NLE	NAE	NLW	Purpose
DAMS			✓	To enable material to be ingested into the library's digital archive, managed throughout its lifecycle and accessed by the public.
DROID		✓	✓	To identify the precise format of all stored digital objects, and to link that identification to a central registry of technical information about that format and its dependencies.
Fedora version 2.0.	✓			For creation, management and preservation of digital documents.
HTTrack	✓			For harvesting from the web.
JHOVE	✓	✓	✓	For format identification, validation, and characterization of digital objects.
Linux 'file' command			✓	To determine the type of data contained in a computer file.

md5sum			✓	To verify the integrity of files (i.e., to verify a file has not changed as a result of file transfer, disk error, etc.). The MD5 hash (or checksum) functions as a compact digital fingerprint of a file because almost any change to a file will cause its MD5 hash to also change.
MediaInfo			✓	For supplying technical and tag information about a video or audio file
Oracle	✓	✓		To organize, store and retrieve data.
POLP	✓			For linearizing, optimizing, repairing, analyzing, encrypting and decrypting PDF documents and to extract technical metadata.
PRONOM		✓	✓	Provides impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value.
Sybase		✓		To organize data and make it available to many users in a network.
Tessella SDB system		✓		Allow to store and preserve digital objects.
UAM	✓			For preparation and transfer of digital documents extracted from electronic records managements systems.
web databases used LAMP architecture (Linux/Apache/MYSQL/PHP)	✓			For deploying web applications.

For more information on these tools website references are provided at the end of the literature references.

JHOVE is used in all the three institutions. PRONOM and DROID are used in two institutions, NAE and NLW. In general, a variety of software and tools used in each institution as it is stated. One institution, NLW, indicated that the reason for selecting tools it uses is because of their widely acceptance and use within the digital preservation field. NAE indicated its preference for example for using the Tessella safety deposit box (SDB) system to manage its digital archiving is because it is developed basically to a similar approach to PREMIS.

One institution, NAE, use an external tool called Universal Archiving Module created by the National Archives of Estonia for agencies for the preparation and transfer of digital documents extracted from electronic records managements systems. Use of UAM requires the ability of an institution's electronic record management system to export documents and their metadata in XML format.

NLW is currently investigating options for a facility for people and establishments outside of the library to be able to submit resources to the Library's repository using online submission tool. It also began on developing the CDAS system (CD Accessioning System) in order to deal with a growing number of archive collections arrive in the library on physical media carriers such as CDs and DVDs.

To sum up, as it is depicted in table 4.1., there are software and tools that have got a chance to be utilized by three of the institutions (e.g., JHOVE), by two institutions (e.g., PRONOM and DROID). The other tools are used by one institution. This is perhaps because of the requirements and the mission of the institutions and the functionality of the tool. However, if it was by the mission of the institution at least the two national libraries should get used wide similar software and tools. Rather, this may have an implication on the acceptance of the software and tools by the memory institutions. Thus, this is a good signal for those of who are producing software tools for digital preservation field. In the study it was also observed that the National Archives of Estonia have developed a valuable system called UAM to support the preparation and transfer of digital documents to the digital archive unlike the national libraries.

4.4. Preservation Metadata Practice in Memory Institutions

4.4.1. Categories of Metadata and its Management

The three institutions are recording a wide range of metadata in their digital archives. Generally, they handle and categorize metadata into three broad groups: descriptive, structural and administrative metadata. It is also observed that these institutions understand that administrative metadata include provenance information, rights information, technical metadata and some information necessary for the long-term preservation of digital objects. However, the volume/scope/length of information recorded varies from institution to institution. For example, NAE indicated that for the moment the rights and permissions information is not a serious problem. Therefore, it is not concentrating on rights that much, unlike the national libraries (NLE and NLW). However, NAE believes that it will be an issue in the future and it explained that whenever the need arises, it can handle it in the future since its system is extendable.

The interviewed institutions indicated that preservation metadata is found within other categories of metadata. They consider different international and national standards to record the information about digital objects.

For example, NLE mapped to DC and ESE (Europeana Semantic Element- the new format used by the Europeana portal) for its descriptive metadata. ESE is Europeana "Schema" for the prototype based on the Dublin Core Metadata Elements Set (DCMES) (ISO). NLE believes that it has not adopted a wide range of standards for other categories of metadata instead it is looking them as a reference for the development of its own specification. NLE current system records information like:

- filename,
- fileSize,
- UploadDate,
- MimeType,
- ChechsumSHA1,

for PDf file:

- CreationDate,
- Optimized,
- Author,
- Title,

for TIFF file:

- ImageSize,
- ImageWidth,
- Imagelength,
- BitsPerSample,

- ChecksumMD5,
- ScannerType,
- ScanTwain,
- OCRsoft and
- hasVersion,
- Tagged,
- Encrypted,
- Producer,
- ModDate,
- Page_size,
- PDF_version and
- Pages
- compresion,
- orientation,
- xresolusion,
- resolutionUnit,
- software, and
- datetime

Other metadata elements like the relationships of the objects when and which objects are related and how they are related are also recorded. These metadata elements are stored in the form of an xml file in the database. The original file sent by the publisher, archive file and the user file are stored in the database and their relation is indicated. The metadata and the object are stored separately.

In the NLW metadata is obtained internally via catalogue records, Electronic Programme Guide (for off-air recordings), externally provided metadata through OAI-PMH and it is stored in METS documents within the Digital Asset Management System. NLW mapped its descriptive metadata to MODS and Dublin Core. According to the NLW, the amount of information recorded is dependent on the end purpose. In NLW administrative metadata is developed in in-house workflows. It uses PREMIS, TEI, textMD and MIX to handle it. Administrative metadata for NLW includes information on how the digital document was scanned, its storage format, when it is created, etc (often called technical metadata), copyright and licensing information, and information necessary for the long-term preservation of the digital objects (preservation metadata). In the NLW a lot of the technical metadata is extracted using software applications such as JHove and MediaInfo. This information is then contained within the METS documents. Structural metadata is handled through structural map within METS document. The sample METS document of NLW is found in appendix B.

Unlike the National libraries, NLE and NLW, NAE use the Estonian adoption of ISAD(G) and ISAAR(CPF) standards for resource discovery. It is not using EAD (encoded archival description) or EAC (encoded archival context). It indicated the reasons for the choice of

these standards are, first, it is the only standards available. Second, everyone is using it and NAE is also using it. In the data management module, the SDB System is adopted. For NAE, technical and some administrative metadata are dependent on the output of different tools. NAE is defining workflows and services and analyzing elements generated by tools whether it can be mapped to its requirement or not. NAE is also pragmatically working to generate technical metadata using tools automatically and also indicated that administrative information should be gathered or summarized from the events and reporting module of data management system.

For NAE, structural metadata is hidden in two places: in the manifestation file and their relations and ISAD(G) in the higher level. NAE in general is dealing with the metadata element identification on a more holistic approach not clearly separating different categories of metadata for different types of entity. Metadata including some information like files and manifestations are stored as flat XML files in databases. Security backup copy of the descriptions stored separately. For searching and management purpose NAE used different databases to build a good query easily and facilitate searching in their system for updating and other management purposes.

However, in two institutions, NLE and NAE, it is observed that they didn't clearly define and demarcate some metadata elements for which category it is and still they are working on it.

As it is discussed in the literature review in section 2.3.3., preservation metadata comprises of different categories of metadata and institutions must record and handle them properly for achieving long-term accessibility of digital objects. However, as it is stated above these memory institutions recoded and handled various metadata elements differently in their digital archiving process and the scope of information they recorded is quite different. For example, in terms of preservation metadata standard adoption and scope/depth of recorded metadata information, NLE is far behind as compared to NAE and NLW.NAE is doing a good job in defining metadata information for different types of digital objects. NLW adopted PREMIS and other standards and it recorded a wide range of metadata information as compared to NLE and NAE.

4.4.2. Interoperability

Managing the transfer of metadata or information packages containing metadata to other repositories is a crucial task for digital repositories because it helps them to co-operate for exchange of information packages or metadata. As it is discussed in section 2.7., the aspect of interoperability need to be addressed by digital repositories since capturing and reusing of metadata is one of the main tasks of digital preservation. Thus, the following paragraphs describe the interoperability practice of memory institutions.

METS (Metadata Encoding and Transmission Standard) is the most commonly used scheme in these institutions. They use it to import and export metadata.

NLE can transfer metadata only or metadata and the object together with the partners like Europeana. Import as well as export is practiced in the NLE with its project partners and other libraries. It uses METS for this purpose. Mostly its partners ask only for the metadata and want to access the material from the NLE database. This depends on the need of the institution and both techniques are possible according to the NLE.

The NLW preservation repository is capable of exchanging metadata or information packages containing metadata through OAI-PMH and metadata contained within Dublin Core section of METS documents.

The NAE has planned to handle the interoperability issue with the SDB system functionality for import and export that uses the METS schema.

4.4.3. Application of PREMIS Entities

One of the main principles behind PREMIS is that it needs to be very clear about what it is going to be described. PREMIS defines five kinds of things called entities: intellectual entities, objects, agents, events and rights. Thus, in the following section, it is tried to see the application of PREMIS entities within these institutions from the general level since some institutions are either not using PREMIS or on the way to use it.

To make it clear for the discussion, NLE is not using PREMIS but is still analyzing it and other standards for the development of its own metadata schema specification. Currently, NLE is writing down a specification for its own digital repository software based on its requirements and wants to implement it in the coming year.

NAE is looking at PREMIS along with other standards for the development of its own metadata schema. It has looked at the concepts of PREMIS and found some of the concepts useful, e.g., the differentiation between different types of entities even though it is possible to argue over using four or five separate entities. NAE added that in a theoretical level it seems reasonable to use PREMIS for preservation metadata and decided last year to use the SDB system. The SDB system is using a similar approach to PREMIS to separate preservation metadata into different entities. So, it was possible to notice that NAE has good notion about PREMIS metadata standard. However, NAE does not yet know in detail to what extent PREMIS could be implemented in its new system. The task of looking at PREMIS and defining the detailed metadata elements at the lower data level is not yet completed. NAE does not currently map its metadata element specifications to PREMIS.

NLW uses PREMIS as information for preserving its digital objects. PREMIS was chosen because it is an international standard and widely used within the field of digital preservation. The library attempts to include all mandatory elements required by PREMIS and adheres to data constraints. The library also attempts to provide as much information as is necessary and often completes elements that are obligatory and not mandatory.

A variation of understanding and progress of preservation metadata practice and use of PREMIS is observed within these institutions. NLE is not currently using it. NAE is looking at it alongside with other metadata standards and NLW is using it. Hence, the use of PREMIS data dictionary or equivalent metadata elements in the memory institutions for intellectual entities, object entities, event entities, agent entities and rights entities are discussed in the following.

Even if two out of three institutions do not fully support PREMIS, I have studied the metadata they recorded and have made connections between PREMIS and their metadata because

PREMIS can be used for systems in development as a basis for metadata definition and/or for existing repositories as a checklist for evaluation.

4.4.3.1. Intellectual Entity

PREMIS defines an intellectual entity as "a set of content that is considered a single intellectual unit for purposes of management and description: for example, a particular book, map, photograph, or database". PREMIS does not actually define any metadata pertaining to intellectual entities because there are plenty of descriptive metadata standards to choose from. Rather, PREMIS does say “an object in a preservation system should be associated with the intellectual entity it represents by including an *identifier* of the intellectual entity in the metadata for the object” (Caplan, 2009, p.9).

Thus, as stated above, NLE is in a reviewing/researching stage and not using PREMIS for the moment. NAE is in the progress of implementing its system. So, in both cases, the implementation of the semantic unit pertaining to intellectual entities, for example, linkingIntellectualEntityIdentifier or even equivalent metadata element for it, is not observed and not clearly indicated how to use these semantic units in the future.

In the case of NLW, its METS documents often contain a sourceMD section which points to the corresponding catalogue record for the item in the OPAC. NLW has its own vocabulary for certain aspects of such a type.

4.4.3.2. Objects

“Most of PREMIS is devoted to describing digital objects. Objects are what are actually stored and managed in the preservation repository. PREMIS defines three different kinds of objects (representation, file, bitstream) and requires implementers to make a distinction between them”. In the PREMIS data dictionary the information that can be recorded for object entities include (Caplan, 2009. p.9):

- a unique identifier for the object (type and value)

- fixity information such as a checksum (message digest) and the algorithm used to derive it
- the size of the object
- the format of the object, which can be specified directly or by linking to a format registry
- the original name of the object
- information about its creation
- information about inhibitors
- information about its significant properties
- information about its environment
- where and on what medium it is stored
- digital signature information
- relationships with other objects and other types of entities

Thus, in describing object entities, the types of objects that memory institutions manage are varying. They record a range of metadata elements for all object entities. Three of the institutions record metadata about representations, files and bitstreams. Representations and files implemented more. However, files and bitstreams are taken as the same in most cases and bitstreams are implemented less commonly. One institution, NAE, indicated that it is preparing different metadata descriptions for each type of object entity.

NLE records few information about object entity (books, website, photographs, files) as compared to the PREMIS data dictionary e.g., filename, fileSize, MimeType, ChechsumSHA1, ChecksumMD5, and lacks some semantic units may be because of not considering PREMIS as an information for preserving digital objects.

In the case of NAE in its conceptual data model, it represents quite a lot of metadata elements for different types of object entities (books, website, photographs, audio, video, files). NAE believes that it may vary when it will be implemented practically.

The NLW records wide range of information about object entity. It uses some semantic units of PREMIS and represent in XML using the METS structure (e.g., ObjectIdentifier (Type and

Value), objectCategory, originalName, fixity, format, formatRegistry, significantProperties, relationship, relatedObjectIdentification, linkingObjectIdentifier). You can refer to the NLW METS template document in appendix B for more information.

Thus, from this it is possible to understand that the NLW recorded a wide range of information as compared to the NLE and NAE.

4.4.3.3. Agents

People, organizations and other entities can act as agents. In the PREMIS data dictionary the information that can be recorded about agents includes (Caplan, 2009):

- a unique identifier for the agent (type and value)
- the agent's name
- designation of the type of agent (person, organization, software)

Thus, all of the three institutions use some form of agent entity however the level of detail of recorded information is quite different. In NLE, agents are not handled directly but tried to handle in the other way round. In NAE, agents are modeled and handled clearly by specify like software, people, producer, agency, etc. NLW adopts the PREMIS sections within the METS documents. This METS document contains information regarding for example what software and tools were used to perform certain events, whether the event was successful or not and also further information regarding these are contained within PREMIS agents. You can refer for detail information to the NLW METS template document in appendix B.

Even though the level of metadata information recording and the way of handling of agents are different, all three memory institution handled the agents' entity to some extent.

4.4.3.4. Events

The event entity aggregates metadata about actions. It is up to the repository which actions to record as events (OCLC/RLG, 2008). In the PREMIS data dictionary the information that can be recorded about events includes:

- a unique identifier for the event (type and value)
- The type of event (creation, ingestion, migration, etc.)
- the date and time the event occurred
- a detailed description of the event
- a coded outcome of the event
- a more detailed description of the outcome
- agents involved in the event and their roles
- objects involved in the event and their roles

“Each preservation repository must make its own decisions about which events to record as a permanent part of an object's history. PREMIS recommends that actions that change an object should always be recorded” (Caplan, 2009, p.10).

Hence, event entity is handled in the institutions to a varying degree. According to the NAE, in the PREMIS events are described as general thing however it is different in practice. For example , the SDB system that the NAE is adopting describes events clearly in separate events like validation event, identification event, property abstraction event, embedded byte stream discovery event, component measurement event, component discovery event, etc with different descriptions and also for migration and emulation different events with different descriptions. NAE used IP logic rather than manifestation and file set logic and also want to include digitization events (software, hardware and profile elements). The provenance information is coming from events on one side and different relationships of different manifestations or AIP. Different events have a keyword attached to them.

The NLE strongly believes that event entities information must be recorded and handled in a detail way. However, currently, it is observed that the NLE handled little information about event entity. The NLE needs to work more on this issue.

NLW uses PREMIS for this entity as well. As we know PREMIS event provides details of what process the original has been subjected to and the result of that event. PREMIS sections within the METS documents contain information regarding what software and tools were used to perform certain events, etc. NLW uses its own vocabulary for certain aspects such as

values in PREMIS events which are currently being updated. Information regarding the archive level file, e.g., file format, version and validation with link to technical registry (PRONOM), the relationship of the derivatives created to the archival file and the date and time at which these were created along with cyclical redundancy checks, software and hardware used to create the derivatives are all handled. You can refer for more information to the NLW METS template document in appendix B.

It is observed that event entity information has got strong emphasis in all memory institutions and they worked hard to handle it. However, still the depth of the information they recorded is varying. In this regard, the NAE and the NLW were doing well than the NLE.

4.4.3.5. Rights

The rights entity aggregates "information about rights and permissions that are directly relevant to preserving objects in the repository. Each PREMIS rights statement asserts two things: acts that the repository has a right to perform and the basis for claiming that right". In the PREMIS data dictionary the information that can be recorded in a rights statement includes (Caplan, 2009, pp.11-12):

- a unique identifier for the rights statement (type and value)
- whether the basis for claiming the right is copyright, license or statute
- more detailed information about the copyright status, license terms, or statute, as applicable
- the action(s) that the rights statement allows
- any restrictions on the action(s)
- the term of grant, or time period in which the statement applies
- the object(s) to which the statement applies
- agents involved in the rights statement and their roles

Therefore, like any other entities, these memory institutions handled rights in their own way. The NLE is recording little information about right entity such as rights of the owner,

restrictions on use, comments and explanations but a bit hard to correspond to the rights entity in the PREMIS data dictionary.

Rights for preservation do not concern NAE for the moment and does not handle rights entity in detail.

NLW handles the rights through using PREMIS and METSRights. The rights the repository has over the object and also what others can do with the object. NLW adopts semantic units pertaining to rights from PREMIS for example rightsStatement, rightsBasis, copyrightInformation, rightGranted. (Refer appendix B).

Thus, it is noticeable that the way and level of handling the rights entity is quite different in these memory institutions. NAE is almost not recorded the rights information. NLE is recorded little information and NLW is recorded quite good information.

4.5. Metadata Standards and Schema in Use at Memory Institutions

A range of metadata is required in order to successfully manage and preserve digital objects. These institutions use variety of standards and/or schema to record different metadata elements. Those standards and/or schema are depicted in the following table.

Table 4.2 Metadata standards and schema in use at memory institutions

Standard/ schema	NLE	NAE	NLW	purpose
Dublin Core	✓		✓	For representation of the bibliographic / descriptive metadata elements of in the libraries
EAD			✓	For encoding of finding aids (collection-level description)
Europeana Semantic Element (ESE)	✓			For recording the descriptive metadata elements
ISAAR (CPF)		✓		For recording descriptive metadata elements in the archive

Estonian adoption of ISAD(G) and ISAAR(CPF)		✓		For representation of descriptive metadata elements
ISAD (G)		✓		For recording descriptive metadata elements in the archive
Library of Congress Audiovisual Metadata (LC-AV) -Audio Metadata (AMD)			✓	To represent technical metadata specific to audio files
-Video Metadata (VMD)			✓	To represent technical metadata for digital video object e.g. bit rate, compression codec.
MARC21	✓		✓	For representation and communication of bibliographic and related information in machine-readable form and it is mapped to DC / ESE/MODS/ EAD in the libraries
METS	✓	✓	✓	For encoding descriptive, administrative, and structural metadata and expressed using the XML schema language.
METSRights			✓	For Rights Declaration.
MODS			✓	For representation of bibliographic elements.
PREMIS		✓	✓	For the management of preservation metadata of digital objects.
MIX			✓	To manage technical data elements of digital image collections which is expressed in XML schema language
TEI			✓	For representation of texts in digital form.
textMD			✓	For detailing technical metadata for text-based digital objects

For more information on these standards website references are provided at the end of the literature references.

As it is depicted in the table, institutions use or adopt different metadata standards and schema for management of different categories of metadata. METS is in use by all three institutions. One institution, NAE, explained that it is reviewing existing metadata standards/schemas for the development of its own set of elements. The NAE do not want to use exclusively any one of the above schemes, but plans to adapt recommendations and elements to fit its requirements for particular materials and actions to be managed. It expects to use different tools based on the existing schemas for technical metadata.

The NLE explained that it is looking and analyzing different standards and schema for implementation. It is in the process of determining its own metadata specification and developing its own requirements.

As it is indicated in the table, the use of metadata standards and/or schema in these three institutions has a huge difference. In this regard, the NLW is in a happy position as compared to NAE and NLE. NLW indicated that all local metadata is mapped to elements from within recognized standards and it attempts to adhere to these standards as much as possible. NLW adopts a wide range of standards/schema to manage the metadata elements of digital objects.

4.6. National Libraries vs National Archives

Differences between national libraries and national archives in terms of materials accepted and the ingest process are significant and reflect the differences in mission. The main difference is in the primary type of material collected – publications vs public records, and in the way the collection happens (national archives have a much tighter control over setting requirements and conditions for the quality of material it ingests, compared with National Libraries that have to accept pretty much everything the publishers give them).

In the national Libraries, their traditional catalogue or OPAC has connection to their metadata for the preservation of the digital documents however this is not seen in the NAE. For example, in NLW, often the traditional metadata description standards such as the bibliographic records created in the catalogue according to MARC21, AACR2, LCSH are mapped to descriptive metadata such as MODS, Dublin Core or EAD e.g. MARC21 is

mapped to other metadata schemas for use in management of digital objects usually using crosswalks (e.g. MARC to MODS) or in-house stylesheets. Metadata is obtained internally via catalogue records, Electronic Programme Guide (for off-air recordings), externally provided metadata through OAI-PMH. Its METS documents often contain a sourceMD section which points to the corresponding catalogue record for the item in the OPAC.

In the case of NLE, it is harvesting the descriptive metadata elements from its OPAC which is based on Z39.50 to the digital archive especially for digitized materials and mapped MARC21 to DC and/or ESE.

For the NAE, the practical implementation is basically in detail as a question of technical compliance and it is along the way to implement its own schema. It doesn't use any international standards but for reference purpose. It uses Estonian metadata record management adoption of ISAD(G) and ISAAR(CPF) for object level records description like who created, signer, and for example different description for different types of file formats. However, NAE is not 100% compliance with it. Some national file tuning or changes are done and the NAE is developing its own schemas. It uses the Tessella SDB data model for its data management module and planned to include other types of metadata elements and others from the output of tools.

To sum up, there are similarities and differences in the kind of metadata standard and/or schema they implemented. For example, the study revealed that METS and PREMIS are the only standards that are commonly used by both the national archives and national libraries. The national archives are attracted to the archival standards for collection level description and the national libraries to other standards as discussed in section 4.5. This perhaps has an implication on the development of metadata standards and gives a clue that the requirements of the digital archive of the national archives and national libraries need to be further explored. National libraries are harvesting some metadata elements from their OPAC and mapping those metadata elements to their system but this is not revealed in the case of national archives.

4.7. Problems and Challenges

The studied memory institutions have faced different problems and challenges in managing the metadata on the digital preservation process for ongoing accessibility of digital objects.

In the process of digital preservation particularly metadata management is a challenging task for all institutions. For example, the decision what preservation standards, tools and media would be used, what preservation strategies and techniques for addressing the threats would be taken and how to automate preservation actions are some of the challenging tasks they faced.

NAE and NLE are designing their own metadata schema based on their best knowledge and analyze the risks and they believe that some additional problems may arise in practice. This is because of different challenges and uncertainties like funding for digital preservation, the high costs of taking action and continuing rapid changes in the availability of hardware, software and other technology required for access.

A problem of repetitive/cyclical task is anticipated in the memory institutions. For example, metadata information for rights of digital objects is not currently given attention in the NAE even if it believes its necessity.

Significant properties are characteristics of the digital object that should be preserved through the chosen preservation strategy. The determination of significant properties may be a repository-wide decision adhering to all materials in a particular class (Caplan, 2006). Thus, defining significant properties that have to be maintained for different digital objects is a difficult task and a huge problem for all three memory institutions.

The diverse and frequently changing range of file formats and standards, and the widespread use of relatively unstable media have a lot of impact on their processes of preservation. In addition, the administrative complexities in ensuring timely and cost-effective action are other challenges. As it is discussed in section 2.2, these challenges have an impact on the implementation of preservation metadata because they are parts of the process and those

challenges should be taken into consideration in deciding which metadata element is going to be recorded or which significant properties of the digital object is preferred over the other.

4.8. Discussion

The three interviewed memory institutions are recording a wide range of metadata elements. They group it mainly as descriptive, structural and administrative metadata. Administrative metadata can include rights, provenance, and technical metadata.

These institutions use metadata elements from various schemas. However, the implementation of preservation metadata within these memory institutions differs in scale, data management practices as well as heterogeneity of metadata recorded (cf. section 4.4.1).

Though these memory institutions use a variety of metadata standards/schema and it is a good trend, a great difference in the level of exploitation of those standards is observed. Among them, METS is the most widely used. This is a good sign for institutions to exchange their information and overcome the problem of interoperability issue (cf. section 4.4.2. and 4.5.).

The institutions included in this case study use different software and tools for tasks like capturing metadata, format identification, validation, and characterization of digital objects. However, the exploitation of these software and tools varies within these institutions. JHOVE, PRONOM and DROID are the mainly used externally available tools for preservation metadata creation and extraction of technical metadata.

One institution, NAE, has developed an in-house tool called UAM for the preparation and transfer of digital documents extracted from electronic records managements systems for agencies. The UAM has a large component dealing with metadata and converting the records' metadata into archival description that can be ingested into the digital archive. So UAM is clearly supporting the ingest of metadata and the schema it uses is matching the current thinking within the NAE for what metadata is needed to support digital archiving.

NLW is currently investigating options for online submission tool. It also began on developing the CDAS system (CD Accessioning System). NLE is looking to upgrade its existing digital repository software DIGAR.

All three institutions store multiple versions of every digital object in their care. For example, NAE and NLW store originals, migrated versions and backup copy/master copy; the NLE stores the originals, converted version and backup copy/master copy in the repository and their metadata is stored in an XML file within a database.

These memory institutions record metadata about different digital objects like books, WebPages, photographs, audio, video, files, bitstreams. However, the level of implementation of each object varies within institutions. For example, bitstreams and file are taken the same in most cases and bitstreams are implemented less commonly. This practice is likely to lead to risks in the future like not being able to distinguish between files and bitstreams and their properties.

The study revealed that PREMIS is not the only standard that institutions are looking at and that they are very likely at adopt a "pick-and-mix" strategy to suit their own metadata needs, rather than adopt straightforwardly just one standard even though PREMIS is associated with the OAIS reference model and institutions generally like its multiple-entity data model. This is maybe because they just could have a difficulty to understand the PREMIS standard fully for practical implementation (see section 4.4.3). Thus, the application of PREMIS entities varies from institution to institution and ranges from reviewing/analyzing stage to implementation.

Significant difference has been seen between national libraries and archives in mission, process of ingest of materials to their digital archive, influence or connection of their traditional cataloguing practice and type of standards used for the development of their metadata specification.

The results of this study support the findings of the survey of the PREMIS working group conducted in 2004 and a survey on the implementation of the PREMIS data dictionary by Woodyard-Robinson in 2007. Particularly, it come out with similar results on the level of

implementation of PREMIS entities, the type of preservation strategies, standards/schemes software and tools used as well as in understanding of preservation metadata elements. The results also showed to some extent similar problems and challenges to the results of previous research (see section 2.6.1).

4.9. Chapter Summary

This chapter has provided an analysis of the data and findings obtained in the research project. It explored the main themes that corresponded with the objectives and research questions of this study. It started by presenting background information of institutions and their respective digital archive. The digital archiving process with comparison of each other was presented. Then the discussion of implementation of preservation metadata was followed. Software, tools, standards and / or schema in use at memory institutions were investigated and discussed. To what level the PREMIS entities are implemented in these institutions was also covered in this chapter. Later, some comparison between national libraries and archives was made and then problems and challenges faced in the implementation of preservation metadata were identified and stated. Finally, the chapter was concluded with discussion on basic issues of the research.

CHAPTER FIVE: CONCLUSION AND FUTURE WORK

This final chapter of the thesis presents conclusions about the findings of this research. It summarizes the key findings drawn from the interviews and document analysis. It focuses on the main issues learnt from the study. This has been done by answering the research questions in a summarized form as well as pointing the implications of this research and possible future research ideas.

This study adopts a qualitative approach and uses the case study strategy. The data collection method consisted of semi-structured interviews and document analysis. Metadata experts/specialists were interviewed in three memory institutions (the National Library of Estonia, the National Archives of Estonia and the National Library of Wales) that practice digital preservation.

The literature discussed in chapter 2 revealed that there are gaps in the implementation of preservation metadata standards from theory to practice and as a result has its own future challenge from the very aim of digital preservation. In addition, there has been little research which has shown the application and therefore a number of case studies are expected to report on both implementation and use in carrying out preservation strategies even though metadata management in the process of digital preservation for long-term accessibility of digital objects has been a critical discussion point internationally. Thus, this study examined the extent of implementing standard preservation metadata into practice at memory institutions. Identifying the extent to which international metadata standards have been adopted for the preservation process will allow to analyze the extent of which metadata is used to support the digital preservation processes as well as to investigate problems and challenges that could be faced in the current practice of metadata usage for the preservation of digital objects. Therefore, the intent of this study was to add the case study researches that show about the application of preservation metadata standards in to practice along with the problems and challenges in the process and to provide some potential ideas for future research.

5.1. Conclusion to the Research Questions

The following section provides a summary of the findings in relation to the four research questions of this study.

Q1. How effective are preservation metadata theories into practice?

According to the results of this study, memory institutions use metadata elements from various standards and/or schema to suit their purposes and recording a wide range of metadata elements from different metadata categories such as descriptive, structural and administrative metadata. The study revealed that for these memory institutions administrative metadata includes rights, provenance, and technical metadata and preservation metadata is found within all other categories of metadata. However, the study revealed that there has been a dissimilarity/discrepancy in the level of exploitation of preservation metadata standards, understanding and progress of preservation metadata practice and use of the PREMIS standard within the memory institutions. The implementation of preservation metadata within these memory institutions differs in scale, data management practices as well as heterogeneity of metadata recorded. For example, NLE and NAE are working to have their own metadata specification and at least NAE is considering PREMIS along with other standards and the NLE is only using PREMIS to inform the development of its own preservation metadata schema indirectly. The NLW, on the other hand, is using PREMIS fully for its preservation metadata implementation. Therefore, the adoption of PREMIS entities ranges from reviewing/researching stage to implementation and its five entities are implemented partially.

The study also revealed that significant differences exist between national libraries and national archives in terms of mission, materials accepted, the ingest process and the connection between their traditional catalogue and/or OPAC to their metadata for the preservation of digital objects.

Q2. What tools, standards and strategies are adopted for metadata management and why?

The study revealed that a range of software and tools are utilized in the memory institutions for different tasks in order to assist in the preservation of its digital object, e.g., for capturing

preservation metadata, format identification, validation, and characterization of digital objects. Among them JHOVE, PRONOM and DROID are the most used ones.

The study showed that even though some institutions are researching and reviewing the existing metadata standards and/or schemas for the development of their own metadata specifications, as it is stated in the summary of Question 1, a variety of standards and/or schema are in use to record and manage different categories of metadata elements. The study revealed that PREMIS is not the only standard that institutions are looking at and that they are very likely at adopt a "pick-and-mix" strategy to suit their own preservation metadata needs, rather than adopt straightforwardly just one standard. METS is the most used one. The study showed significant discrepancy in the use of metadata standards and/or schema in memory institutions. In this regard, the NLW is in a better state as compared to NAE and NLE.

The study revealed that migration as a preservation strategy is implemented at least in two of the three institutions and the third one is also planning to use it in the near future in addition to conversion. Institutions are worried about the uncertainty of the future even though they are trying to make the best decisions they can with an eye towards the future.

Q3.What is the level of granularity (e.g., representations, files, bitstreams) that preservation metadata is applied in practice?

The study showed that the application of preservation metadata in describing object entities is varying. They record a range of metadata elements for all object entities (representations, files and bitstreams). Representations and files are implemented more than the rest. However, files and bitstreams are taken as the same in most cases and bitstreams are implemented less commonly. The study found that there is institution that is preparing different metadata descriptions for each type of object entity. Institutions practiced exchange of metadata or information packages (metadata together with the object).

Q4. What type of risks can be anticipated when preservation metadata implemented only partially in practice?

The study discovered that institutions have faced several problems and challenges in managing the metadata on the digital preservation process for ongoing accessibility of digital objects. Largely, metadata management in the process of digital preservation is a challenging task for all three institutions. Settling decisions about preservation medium, defining significant properties, preservation strategies, standards, software and tools to automate those preservation actions, etc are problematic because of the wide range of know-how required for these decisions. The study showed that institutions have faced shortage of funding for their digital preservation process and rapid change of information technology is their concern. The study also showed that a problem of repetitive/cyclical task is anticipated in the memory institutions because of not taking actions to minimize risks on time for example, metadata elements for rights are not given attention in some institutions at the moment, though they believe in its necessity.

5.2. Implications of the Research

The implication of this study is that the results can be used for people, agencies/organizations or for any one that are responsible in developing preservation metadata standards, software and tools to notice the application of them at memory institutions. The implementation of theoretical standards to practice is imperfect. From the results one can get the information about the gaps that are likely happened in the process of implementation of theory into practice at memory institutions.

5.3. Future Research Ideas

This study considers three memory institutions, two national libraries and one national archive. It would be interesting to conduct further research by taking and considering more memory institutions in number and variety like museums, cultural heritage institutions, educational institutions and all other kinds of institutions those practicing preservation of digital objects.

The study has been seen the use of PREMIS as information for the implementation of preservation metadata in the memory institutions from the general level focused mainly on the PREMIS entities with some semantic units. This study can be extended thorough consideration of all PREMIS data dictionary semantic units and constraints on them.

This study has focused on the application of PREMIS data dictionary preservation metadata standard. Research can be done to include the influence of other preservation metadata standards in the implementation of preservation metadata in memory institutions.

This study has discovered different tools adopted in the implementation of preservation metadata. It would be worthwhile to study to what extent these tools are automating the tasks of the preservation metadata processes and how they match to the preservation metadata standards and to what extent they satisfy the practical needs of memory institutions.

This study has discovered that some memory institutions are trying to come up with their own preservation metadata specifications. It would be interesting to study the cooperation level and its need between different memory institutions for the development of better specification that can cope up with the wide range of standards and formats.

This study has also shown that memory institutions have looked at different preservation metadata standards/schema to record metadata elements as well as developed their own metadata specifications. It would also be interesting to study the comparison and harmonization of various metadata specifications as well as the cooperation between the many metadata initiatives that have an interest in digital preservation.

References

Alemneh, D.G., Hastings, S.K. and Hartman, C.N. (2002). A metadata approach to preservation of digital resources: The University of North Texas Library's Experience. *First Monday*, 7(8). Retrieved on January 5, 2010 from http://firstmonday.org/issues/issue7_8/alemneh/index.html.

Anderson, D., Alemu, G.A., Delve J. and Pinchbeck, D. (2009). Preliminary document analyzing and summarizing metadata standards and issues across Europe. 85pp. Retrieved on February 22, 2010 from http://www.keep-project.eu/ezpub2/index.php?/eng/content/download/4124 /20617/file/KEEP_WP3_D3.1.pdf.

Anderson, G. (1993). Fundamentals of educational research. Falmer Press, London, pp:152-160.

Anderson,C., Hallahan, J., Kays, S. and Whitworth, E. (2009).OAIS and PREMIS. Accessed on June 3, 2010 from https://webspace.utexas.edu/ecw494/www/metadata/digipres_masterslides_v2.pdf.

Bennett, J. (1997). A framework of data types and formats, and issues affecting the long term preservation of digital material. British Library Research and Innovation Centre. Available at: http://www.ukoln.ac.uk/services/elib/papers/supporting/#blric.

Besser, H. (2000). Digital longevity, in: handbook for digital projects: a management tool for preservation and access. Andover, Mass. Northeast Document Conservation Center. Retrieved on February 22, 2010 from http://www.nedcc.org/resources/digitalhandbook/dman.pdf.

Byrne, M. M. (2001). Sampling for qualitative research. *AORN Journal*, 73(2), 497-498.

Calanag, M.L., Sugimoto, S. and Tabata, K. (2001). A metadata approach to digital preservation. Retrieved on January 14, 2010 from http://www.nii.ac.jp/dc2001/proceedings/product/paper-24.pdf.

Calanag, M.L., Tabata, K., and Suginoto, S. (2004). Linking preservation metadata and collection management policies. *Collection Building*, 23 (2), 56-63. Retrieved on February 5, 2010 from http://www.emeraldinsight.com/10.1108/01604950410514730.

Caplan, P. (2006). Preservation metadata: DCC Digital Curation Manual, S.Ross, M.Day (eds). Retrieved on January 12, 2010 from http://www.dcc.ac.uk/resource/curation□manual/chapters/preservation□metadata/preservation□metadata.pdf.

Caplan, P. (2007). Ten years after. *Library Hi Tech*, 25(4), 449 – 453. Retrieved on January 22, 2010 from http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&contentId=1640691.

Caplan, P. (2009). Understanding PREMIS. Accessed on February 24, 2010 from http://www.loc.gov/standards/premis/understanding-premis.pdf.

Carignan, Y. et al. (2006). Best Practice Guidelines for Digital Collections at University of Maryland Libraries. Retrieved on March 10, 2010 from http://www.lib.umd.edu/dcr/publications/best_practice.pdf.

CCSDS (Consultative Committee for Space Data Systems) (2002). Reference Model for an Open Archival Information System (OAIS). Blue Book, Issue 1. Washington, DC: CCSDS Secretariat. Available at: http://public.ccsds.org/publications/archive/650x0b1.pdf.

Cordeiro, M. I. (2004). From rescue to long-term maintenance: preservation as a core function in the management of digital assets. *VINE: The Journal of Information and Knowledge Management Systems*, 34(1), pp 6-16. Retrieved on February 21, 2010 from

http://www.emeraldinsight.com/Insight/viewContentItem.do;jsessionid=D3262D8532E2D5B8AFF206682016BB13?contentType=Article&contentId=1503758.

Creswell, J. (1994). Research design: qualitative and quantitative approaches. London: Sage.

Dappert, A. and Farquhar, A. (2009). Implementing metadata that guides digital preservation services. Retrieved on January 13, 2010 from http://www.planets-project.eu/docs/papers/Dappert_MetadataAndPreservationServices_iPres2009.pdf.

Das, T. K., Sharma, A. K. and Gurey, P. (2009). Digitization, strategies & issues of digital preservation: an insight view to Visva-Bharati Library. Retrieved on January 20, 2010 from http://www.inflibnet.ac.in/caliber2009/CaliberPDF/5.pdf.

Day, M. (2001). Metadata for digital preservation: a review of recent developments. Paper presented at the ECDL2001, 5th European Conference on Research and Advanced Technology for Digital Libraries. Retrieved on March 3, 2010 from http://www.ukoln.ac.uk/metadata/presentations/ecdl2001-day/paper.html.

Day, M. (2003a). Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal. Retrieved onn juanuary 24, 2010 from http://www.siderean.com/dc2003/101_paper38.pdf.

Day, M. (2003b). Preservation metadata initiatives: Practicality, sustainability, and interoperability. Retrieved on March 11, 2010 from http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf.

Day, M. (2005). DDC/Digital curation manual instalment on metadata. HATII, University of Glasgow; University of Edinburgh; UKOLN, University of Bath; Council for the Central Laboratory of the Research Councils. Retrieved on January 28, 2010 at http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata.

Denzin, N., and Lincoln, Y. (1994). Handbook of qualitative Research. Sage Publicatio, California, pp: 3-5.

DigitalPreservationEurope. (2006). DPE Research Roadmap, DPE-D7.2. Retrieved on February 2, 2010, from http://www.digitalpreservationeurope.eu/publications/dpe_research_roadmap_pdf.

Gartner, R. (2008). Metadata for digital libraries: state of the art and future directions. Bristol: Technology & Standards Watch. Retrieved on March 14, 2010 from http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf.

Glesne, C. and Peshkin, A. (1992). Becoming qualitative researchers: An introduction. White Plains, NY: Longman.

Gray, E.D. (2004). Doing research in the real world. Sage publication. London. Thousand Oaks. New Delhi.

Griffee, D. (2005). Research tips: interview data collection. *Journal of Developmental Education*, 28(3), 36-37. Retrieved on March 14, 2010 from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/2a/1c/bb.pdf.

Groenewald, R. and Breytenbach, A. (2009). The Use of metadata and preservation methods for continuous access to digital data. Retrieved on February 17, 2010 from http://www.ais.up.ac.za/digi/docs/groenewald_paper.pdf.

Guenther, R. (2009). Understanding and Implementing the PREMIS Data Dictionary for Preservation Metadata. Accessed on May 7, 2010 from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp09/index.html.

Honey, M.A. (1987). The interview as text: Hermeneutics considered as a model for analyzing the clinically informed research interview. *Human Development*, 30, 69-82.

Jana, S., Mondal, M.K. and Marjit, U. (2090). Digital preservation with special reference to the open archival information system (OAIS) reference model: an overview. Retrieved on February 22, 2010 from http://www.inflibnet.ac.in/caliber2009/CaliberPDF/3.pdf.

Johnson, R. B. and Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33 (7), 14-26. Accessed on June 3, 2010 from http://www.aera.net/uploadedFiles/Journals_and_Publications/Journals/Educational_Researcher/Volume_33_No_7/03ERv33n7_Johnson.pdf.

Kaplan, B. and Maxwell, J.A. (1994). Qualitative Research Methods for Evaluating Computer Information Systems, in Evaluating Health Care Information Systems: Methods and Applications, Thousand Oaks, CA. pp. 45-68.

Knight, S. (2005). Preservation metadata: National Library of New Zealand experience. *Library Trends*, 54(1), 91-110. Retrieved on March 2, 2010 from http://muse.jhu.edu/login?uri=/journals/library_trends/v054/54.1knight.html.

Lavoie, B. and Gartner, R. (2005). Technology watch report: preservation metadata, Oxford University Library Services and Digital Preservation Coalition. Available at: http://www.dpconline.org/docs/reports/dpctw05-01.pdf.

Lavoie, B. F. (2004). Implementing metadata in digital preservation systems: The PREMIS activity. *D-Lib Magazine*, 10(4). Retrieved on March 2, 2010 from http://www.dlib.org/dlib/april04/lavoie/04lavoie.html.

Lazinger, S. H. (2001). Digital preservation and metadata: history, theory, practice. Libraries Unlimited, Englewood, CO.

Lee, K., Slattery, O., Lu, R., Tang, X. and McCrary, V. (2002). The state of the art and practice in digital preservation. *Journal of Research of the National Institute of Standards and Technology*, 107(1), 93-106. Retrieved on March 7, 2010 from http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf.

Ludäsher, B., Marciano, R. and Moore, R. (2001). Preservation of digital data with self-validating, self-instantiating knowledge-based archives. *SIGMOD Record*, 30(3): 54-

63. Retrieved on February 15, 2010 from http://www.sdsc.edu/NARA/Publications/Web/kba.pdf.

Lynch, C. (1999). Canonicalization: a fundamental tool to facilitate preservation and management of digital information. *D-Lib Magazine*, 5(9). Retrieved on January 17, 2010, from http://www.dlib.org/dlib/september99/09lynch.html.

Mack, N., Woodsong, C.,.Macqeen, K. M , Guest, G.and Namey. E. (2005). Qualitative Research Methods: A Data Collector's Field Guide. North Carolina, Family Health International. Retrieved on March 13, 2010, from http://www.fhi.org/NR/rdonlyres/etl7vogszehu5s4stpzb3tyqlpp7rojv4waq37elpbyei3tgmc4ty6dunbccfzxtaj2rvbaubzmz4f/overview1.pdf.

Marketakis,Y., Tzanakis, M. and Tzitzikas, Y. (2008). PreScan: Towards Automating the Preservation of Digital Objects. *ACM*. Retrieved on January 7, 2010 from http://portal.acm.org/citation.cfm?id=1643898&dl=GUIDE&coll=GUIDE&CFID=91251763&CFTOKEN=12484501.

Marshall, M.N. (1996). Sampling for qualitative research. *Family Practice*, 13: 522-525. Retrieved on January 10, 2010 from http://spa.hust.edu.cn/2008/uploadfile/2009-9/20090916221539453.pdf.

Maxymuk, J. (2005). Preservation and metadata. *The Bottom Line: Managing Library Finances*, 18(3): 146-148. Retrieved on March 1, 2010 from http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&hdAction=lnkhtml&contentId=1513289&ini=xref&history=false.

MILES, M.B.and Huberman, A.M. (1994) Qualitative data analysis: an expanded sourcebook (2nd ed), Sage:London & Thousand Oaks, California.

Myers, M.D. (1997). Critical ethnography in information systems, in information systems and qualitative research. London, pp. 276-300.

National Information Standards Organization. (2004). Understanding Metadata. Bethesda, MD: NISO Press. Available at http://www.niso.org/standards/resources/UnderstandingMetadata.pdf.

National Library of New Zealand. (2003). Metadata standards framework –preservation metadata (revised). Retrieved on March 5, 2010 from http://www.natlib.govt.nz/downloads/metaschema-revised.pdf.

Noor, K. B. M. (2008). Case study: a strategic research methodology. *Am J Appl Sci*, 5: 1602–4. Retrieved on April 2, 2010 from http://www.scipub.org/fulltext/ajas/ajas5111602-1604.pdf.

OCLC/RLG PREMIS Working Group. (2004). Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community. Report by the joint OCLC/RLG Working Group Preservation Metadata: Implementation Strategies (PREMIS). Dublin, Ohio: OCLC Online Computer Library Center, Inc. Available at: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf.

OCLC/RLG PREMIS Working Group. (2005). Data dictionary for preservation metadata: final report of the PREMIS Working Group. Dublin, Ohio: OCLC Online Computer Library Centre; Mountain View, Calif.: Research Libraries Group. Available at: http://www.oclc.org/research/projects/pmwg/

OCLC/RLG PREMIS Working Group. (2008). Data dictionary for preservation metadata: final report of the PREMIS Working Group. Dublin, Ohio: OCLC Online Computer Library Centre; Mountain View, Calif.: Research Libraries Group. Available at : http://www.oclc.org/research/projects/pmwg/

OCLC/RLG Working Group on Preservation Metadata. (2001). Preservation metadata for digital objects: a review of the state of the art. Retrieved on January 26, 2010 at http://www.oclc.org/research/pmwg/presmeta_wp.pdf.

OCLC/RLG Working Group on Preservation Metadata. (2002). A metadata framework to support the preservation of digital objects. Retrieved on January 26, 2010 http://www.oclc.org/ research/pmwg/pm_framework.pdf.

Oltmans, E., and Wijngaarden,H.V. (2004). Digital preservation in practice: The E-Depot at the Koninlijke Bibliotheek. *VINE*, 34(1), 21-26. Retrieved on February 5, 2010 from http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&hdAction=lnkpdf&contentId=862527.

Patton, M. (1987). How to use qualitative methods in evaluation. Sage Publication, California, pp: 18-20.

Patton, M. (2001). Qualitative research and evaluation methods (2nd Edition).Thousand Oaks, CA: Sage Publications.

Research Libraries Group. (2002). Trusted digital repositories: Attributes and responsibilities. Retrieved on April 10, 2010 from http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf.

Rosenthal, D.S.H., Robertson, T.S., Lipkis, T., Reich, V.and Morabito, S. (2005). Requirements for digital preservation systems: A bottom-up approach. *D-Lib Magazine*, 11(11). Retrieved on January 23, 2010 from http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html.

Ross, S. (2007). Digital preservation, archival science and methodological foundations for digital libraries, keynote address at the 11th European Conference on Digital Libraries (ECDL), HATII at the University of Glasgow. Retrieved on April 1, 2010 from http://www.ecdl2007.org/Keynote_ECDL2007_SROSS.pdf.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-75. Accessed on June 5, 2010 from http://iospress.metapress.com/content/3ccttm2g59cklapx/

Srivastava, P., and Hopwood, N. (2009). A practical iterative framework for qualitative data analysis. *International Journal of Qualitative Methods*, 8(1), 76-84. Retrieved on April 2, 2010 from http://ejournals.library.ualberta.ca/index.php/IJQM/article/viewFile/1169/5199.

Strodl, S., Becker, C., Neumayer, R. and Rauber, R. (2007). How to choose a digital preservation strategy: evaluating a preservation planning procedure. Retrieved on January 12, 2010 from http://www.ifs.tuwien.ac.at/~strodl/paper/FP060-strodl.pdf.

Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in coming years. In the state of digital preservation: an international perspective. Conference Proceedings. Institutes for Information Science Washington, DC. Accessed on April 22, 2010 at http://www.clir.org/pubs/reports/pub107/pub107.pdf.

Woodyard, D. (2004). Significant property: Digital preservation at the British Library. *VINE,* 34(1), 17-20. Retrieved on February 1, 2010 from http://ninetta.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&contentId=862526.

Woodyard-Robinson, D. (2007). Implementing the PREMIS data dictionary: a survey of approaches. Retrieved on February 5, 2010 from http://www.loc.gov/standards/premis/implementation-report-woodyard.pdf.

Yin, R.K. (1989). Case study research. Sage Publication, California, pp: 22-26.

Yin, R.K. (2009). Case study research: design and methods. 4th ed. Thousand Oaks: Sage publication.

Website References

American Library Association. (2005). http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html, accessed on April 23, 2010

Doubline core. http://dublincore.org/

DROID (Digital Record Object Identification). http://freshmeat.net/projects/droid.

Encoded Archival description. http://www.loc.gov/ead/

Estonian National Library. (2008). http://www.abbyy.com/adx/aspx/adxgetmedia.aspx?MediaID=588&Filename=2008E_CS_RecServer_ENL_Est.pdf. Retrieved April 7, 2010

European Semantic Elements. http://www.europeanalocal.eu/eng/Documents-Reports

Fedora version 2.0. http://www.fedora-commons.org/software

HTTrack website copier. http://www.httrack.com/

ISAAR CPF :International Standard Archival Authority Record for Corporate Bodies, Persons and Families. http://www.paradigm.ac.uk/workbook/cataloguing/isaarcpf.html.

ISAD(G):General International Standard Archival Description. http://www.ica.org/sites/default/files/isad_g_2e.pdf.

JHove. http://hul.harvard.edu/jhove/

Library of Congress Audiovisual Metadata. http://www.loc.gov/rr/mopic/avprot/avprhome.html.

MARC21 (MAchine-Readable Cataloguing). http://www.loc.gov/marc/

Md5sum. http://manpages.ubuntu.com/manpages/karmic/en/man1/md5sum.1.html.

Mediainfo. http://mediainfo.sourceforge.net/en.

Metadata Encoding and Transmission Standard (METS). http://www.loc.gov/standards/mets/.

Metadata Object Description Schema (MODS). http://www.loc.gov/standards/mix/

'Metadata Rules' - a report from the Open Forum on Metadata Registries. (2003). http://www.webservices.org/categories/technology/registry_uddi/metadata_rules_a_report_from_the_open_forum_on_metadata_registries/(go)/Articles. Accessed on May 22, 2010

METSRights. http://www.loc.gov/standards/rights/METSRights.xsd.

National archives of Estonia. http://www.ra.ee/?id=11991.

National Library of Estonia. (2007). http://www.cdnl.info/2008/CDNL-Quebec-2008-country-report-Estonia.rtf. Retrieved on April 16, 2010

National library of Estonia. http://www.nlib.ee/17606

National library of Estonia. http://www.nlib.ee/584

National library of Wales. http://www.llgc.org.uk/index.php?id=2

NLW. (2008). Digital Preservation Policy and Strategy. http://www.llgc.org.uk/fileadmin/documents/pdf/2008_digipres.pdf. Accessed on April 21, 2010

PLOP (PDF Linearization, Optimization, Protection). http://www.pdflib.com/products/plop/

PREservation Metadata: Implementation Strategies (PREMIS). http://www.loc.gov/standards/premis/

PRONOM. http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

PROTAGE Project. (2010). http://www.protage.eu/partners.html. Retrieved on April 26, 2010

Technical Metadata for Text (textMD). http://www.loc.gov/standards/textMD/

Tessella Safety Deposit Box system. http://www.tessella.com/wp-content/uploads/2009/02/e_tessella-sdb1.pdf.

Text Encoding Initiative (TEI). http://www.tei-c.org/index.xml.

Universal Archiving Module (UAM). http://www.ra.ee/en/universal-archiving-module/&i=6.

Appendix A: Interview Questions

1. What is the mission of the preservation repository?
2. How are materials obtained by the preservation repository?
3. What preservation strategies is your preservation repository implementing now? Why you chose this one?
4. What is your institution trend of preservation, keeping the original and also store several preservation copies of the object , i.e., normalized or migrated version of the content object, each with related metadata?
    a. What is the relationship between the original and preservation copies of the object stored in your preservation repository?
    b. Does an access copy/original copy get preservation treatment (e.g. migration or else)?
5. What software and tools are used in your preservation repository? Why you choose those tools?
6. What metadata standards are in use in your preservation repository? What is the reason for the choice?
7. Could you explain about the traditional metadata/ description standards/catalogues used in your institution?
8. How it gets implemented and influences the metadata implementation in your digital repository?
9. Does your institution use the PREMIS data dictionary as information you need for preserving digital objects? Could you explain?
    a. Is your institution used PREMIS as a checklist for evaluating the software and tools that you are using for preservation? Could you explain?
    b. Have you done any mapping in your existing metadata to the PREMIS Data Dictionary? Could you explain?
10. Is your repository able to cope up with the wide range of standards and formats that exits? How?
11. What categories of metadata are stored and used by your preservation repository?

12. How your repository handles the digital Provenance information, i.e., chain of custody and change history of digital object? What information is recorded?
13. How your repository handles the rights and permissions information? What information is recorded?
14. How your digital repository does understood, managed, created and verified the technical metadata? What information is recorded?
15. How your digital repository does understood, managed, created and verified the administrative and management information? What information is recorded?
16. How your digital repository does understood, managed, created and verified the bibliographic/descriptive metadata? What information is recorded?
17. How your digital repository does understood, managed, created and verified the structural metadata? What information is recorded?
18. Is there any other metadata element that is handled other than the above once in your digital repository? If so, could you explain them?
19. How is metadata obtained by the preservation repository?
20. How metadata stored and updated in your preservation repository? Please explain.
21. How the metadata and materials stored within the preservation repository?
22. What are the important preservation metadata elements /significant properties for your institution? What factors are considered to define them?
23. Do you think that the preservation metadata recorded adequate for the goals of the repository? How?
24. Which metadata encoding scheme are using for implementing the metadata element set? What is the reason for the choice?
    a. What about interoperability, i.e., could your repository be able to transfer metadata or information package containing metadata to other (e.g. object or metadata exchange)? How?
25. Is your preservation repository managed all types of intellectual entities/ the three levels of digital objects, i.e., representation, files, bit streams? If so,

a. Is there a difference for metadata relating to different types of objects and the information recorded indicating relationship between objects? If so, could you explain

26. Have you faced a problem like
    a. Interpreting or defining semantic units of PREMIS or mapping your metadata elements to the PREMIS Data Dictionary in different way.
    b. Misunderstanding of significant properties for your digital repository or mixing of other properties for example taking some technical properties of format specific information, inhibitors, etc as significant properties especially related to PREMIS.
    c. The extended information added by your digital repository, i.e., local metadata, could pose challenges for interoperability and/or do complicate the content structure in your digital repository.
    d. Not explicitly recording mandatory semantic units by policy or any other reason in your digital repository and not adhering to a data constraint in the PREMIS Data Dictionary.
    e. Not applying the obligation of a semantic unit as it is stated in the PREMIS Data Dictionary (e.g., not using explicitly some identifiers even though they are mandatory semantic units).
    f. In defining the important preservation metadata elements /significant properties for your institution by policy or other case.
    g. What other challenges and how would you solve these problems?
27. Could you give / show me examples of implementation in general, i.e., metadata element for each type of entity (intellectual entity, event, agent, right)?
28. Any comments that you would like to add about the practice of preservation metadata in your preservation repository?

Appendix B: NLW METS Template Document

Example PREMIS metadata extracted from one of NLW METS template document.

```
<!-- PREMIS identifier for all the significant files in object - i.e. Archival,  Alto, Text -->
 <METS:dmdSec ID="dmdSec3">
  <METS:mdWrap MDTYPE="PREMIS:OBJECT">
   <!--For archival file use archival file name as unique identifier -->
   <METS:xmlData>
     <premis:objectIdentifier>
     <!-- use capital letters for consistency and in order to differentiate from the file name -->
     <premis:objectIdentifierType>WlAbNL_METS_AWJAD00100001</premis:objectIdentifierType>
     <premis:objectIdentifierValue>fileID1</premis:objectIdentifierValue>
    </premis:objectIdentifier>
    <premis:objectCategory>file</premis:objectCategory>
    <premis:originalName>awjad00100001.tif</premis:originalName>
   </METS:xmlData>
  </METS:mdWrap>
 </METS:dmdSec>

<!-- CRC metadata and format registry reference for Archive file with information about derived files -->
   <METS:mdWrap MDTYPE="PREMIS">
    <METS:xmlData>
     <premis:objectIdentifier>
      <premis:objectIdentifierType>WlAbNL_METS_awjad00100001</premis:objectIdentifierType>
      <premis:objectIdentifierValue>fileID1</premis:objectIdentifierValue>
     </premis:objectIdentifier>
     <premis:objectCategory>file</premis:objectCategory>
     <premis:objectCharacteristics>
      <premis:compositionLevel>0</premis:compositionLevel>
      <premis:fixity>
       <premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>
       <premis:messageDigest>03a101a0bae39047135e55206bd80dc1</premis:messageDigest>
      </premis:fixity>
      <premis:fixity>
       <premis:messageDigestAlgorithm>SHA-1</premis:messageDigestAlgorithm>
       <premis:messageDigest>ea0f440615ab7780ab7f055b88eb54ce4af19e47</premis:messageDigest>
      </premis:fixity>

      <!-- record format with reference to format registry -->
      <premis:format>
```

```
        <premis:formatDesignation>
          <premis:formatName>Tagged image file format (TIFF)</premis:formatName>
          <premis:formatVersion>5</premis:formatVersion>
        </premis:formatDesignation>
        <premis:formatRegistry>
          <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
          <premis:formatRegistryKey>fmt/9</premis:formatRegistryKey>
          <premis:formatRegistryRole>specification</premis:formatRegistryRole>
        </premis:formatRegistry>
      </premis:format>
    </premis:objectCharacteristics>
    <premis:significantProperties>
      <premis:significantPropertiesType>content</premis:significantPropertiesType>
      <premis:significantPropertiesValue>content only</premis:significantPropertiesValue>
    </premis:significantProperties>

    <!-- information about derived Reference file -->
    <premis:relationship>
      <premis:relationshipType>derivation</premis:relationshipType>
      <premis:relationshipSubType>source of</premis:relationshipSubType>
      <premis:relatedObjectIdentification>
        <premis:relatedObjectIdentifierType>WlAbNL_METS_awjad00100001</premis:relatedObjectIdentifierType>
        <premis:relatedObjectIdentifierValue>fileID2</premis:relatedObjectIdentifierValue>
        <premis:relatedObjectSequence>1</premis:relatedObjectSequence>
      </premis:relatedObjectIdentification>
      <premis:relatedEventIdentification>
        <premis:relatedEventIdentifierType>WlAbNL</premis:relatedEventIdentifierType>
        <premis:relatedEventIdentifierValue>CREATE_DERIVED_FILES-001</premis:relatedEventIdentifierValue>
        <premis:relatedEventSequence>2</premis:relatedEventSequence>
      </premis:relatedEventIdentification>
    </premis:relationship>

<!-- Fixity information for Reference image -->
  <METS:techMD ID="techMD6">
    <METS:mdWrap MDTYPE="PREMIS:OBJECT">
      <METS:xmlData>
        <!-- <premis:object> does not validate in latest version but can be included in MDTYPE above -->
        <premis:objectIdentifier>
          <premis:objectIdentifierType>WlAbNL_METS_awjad00100001</premis:objectIdentifierType>
          <premis:objectIdentifierValue>fileID2</premis:objectIdentifierValue>
        </premis:objectIdentifier>
```

```
        <premis:objectCategory>file</premis:objectCategory>
        <premis:objectCharacteristics>
          <!--  <premis:compositionLevel> is mandatory and is an indication of whether the object is subject to one or more processes of decoding or unbundling. Numbering goes lowest to highest (first encoded = 0). 0 is base object; 1-n are subsequent encodings.
              Use 0 as the default if there is only one compositionLevel.-->
          <premis:compositionLevel>1</premis:compositionLevel>
          <premis:fixity>
            <premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>
            <premis:messageDigest>d436f692e7e5d4be2ee650aaf4a04549</premis:messageDigest>
          </premis:fixity>
          <premis:fixity>
            <premis:messageDigestAlgorithm>SHA-1</premis:messageDigestAlgorithm>
            <premis:messageDigest>3124091ad36dc063fc90266b955a6958c61d23bb</premis:messageDigest>
          </premis:fixity>
          <!-- <premis:format> is mandatory in latest version of PREMIS -->
          <premis:format>
            <premis:formatDesignation>
              <premis:formatName>image/png</premis:formatName>
            </premis:formatDesignation>
          </premis:format>
        </premis:objectCharacteristics>
      </METS:xmlData>
    </METS:mdWrap>
  </METS:techMD>

    RIGHTS INFORMATION
    <premis:rights>
    <!-- persmissionStatement was replaced with rightsStatement in this version -->
    <premis:rightsStatement>
                                                    <premis:rightsStatementIdentifier>

          <premis:rightsStatementIdentifierType>WlAbNL</premis:rightsStatementIdentifierType>

          <premis:rightsStatementIdentifierValue>H015487</premis:rightsStatementIdentifierValue>
                                                    </premis:rightsStatementIdentifier>
                                                    <!-- rightsBasis tag is mandatory -->
                                                    <premis:rightsBasis>copyright</premis:rightsBasis>
                                                    <premis:copyrightInformation>

          <premis:copyrightStatus>publicdomain</premis:copyrightStatus>
```

```
<premis:copyrightJurisdiction>gb</premis:copyrightJurisdiction>
                                                    </premis:copyrightInformation>

                                                    <premis:rightsGranted>
                                                            <premis:act>all</premis:act>
                                                            <premis:termOfGrant>

<premis:startDate>2009</premis:startDate>
                                                            </premis:termOfGrant>
                                                            <premis:rightsGrantedNote>Digital object
created at LlGC/NLW and
                                                                    therefore LlGC/NLW has total
control over the
                                                            object.</premis:rightsGrantedNote>
                                                    </premis:rightsGranted>
                                                    <!-- ObjectIdentifier contained word "all" in default
METS profile -->
                                                    <premis:linkingObjectIdentifier>
                                                            <premis:linkingObjectIdentifierType/>
                                                            <premis:linkingObjectIdentifierValue/>
                                                    </premis:linkingObjectIdentifier>
                                            </premis:rightsStatement>
                                    </premis:rights>
```