TALLINNA ÜLIKOOL høgskolen i oslo UNIVERSITÀ DEGLI STUDI DI PARMA

Education and Culture DG

ERASMUS MUNDUS

MIREMBE FLORENCE

Digital Preservation in Norway's Record Keeping and Archiving Traditions: an exploration of authenticity practices using mixed methods research

Supervisor: Thomas, Sødring

Oslo University College

Faculty of Journalism, Library and Information Science

Master thesis
International Master in Digital Library Learning
2010

Declaration

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

fmirembe

………………………………………(Signature of candidate)

Abstract

This work is a study of the current authenticity practices of records in Norway. The study is limited to public administration records at municipality level, specifically municipality and city archives. Norway has a well structured records management tradition but when transferring records from the records creators to the archives, integrity is still inadequate for the current information demands with regards to authenticity, a prerequisite for Trust. This has been observed with the Noark records, as an instrumental case study.

The results portray a picture of what is happening on the ground. The study found out that the archives are semi electronic with manually driven work processes. It is interesting that archives do not have access to the original full electronic records they are supposed to preserve apart from referential data of the records. This puts archives at cross roads since they do not have control over the original databases until 25 years later when they fully take ownership of these records. Therefore with this prevailing scenario, recommendations have been made urging for the need to close this gap in a more trust worthy manner rather than relying on the traditional goodwill assumption which has no scientific verification. The ABM group and the archival community in general is urged to consider revising this time span period and make it shorter, implement file sharing through reliable authenticated systems to meet reliable information demands of the 21st century.

This research is of significance to the ABM group and general archival community in Norway and beyond that can identify with Norway's current authenticity management of archival records.

Keywords: Digital preservation, Authenticity, Records integrity, Noark, Record keeping – Norway.

Dedication

To my loving mum who has strived all the way to make me all that I am today.
Your efforts were and are worth it. I am very grateful!

Acknowledgements

This work is a result of the following persons and institutions that contributed a “special brick” to the “wall” that is standing today.

I am very grateful for all the support and guidance that my supervisor Thomas SØdring provided during the research process. Words are few to express my thankfulness but let me say “Tussen Takk”!

The archival institutions that willingly participated in this research study. Your time and input is very much appreciated.

Dr. Gillian Oliver, for being there for me at times when I needed guidance and support along the way.

Professor AnnMaria Tamarro for all the encouragement during my study time and helping me maximize the potential that you see in upcoming scholars, thank you!

The professors at Oslo University College, you have all contributed in a special way, right from when I started this course for you equipped me to produce what I have today. I am thankful to Kersti Ahren, our administrator for always being there for us when we needed help; you made our learning process much easier.

The Professors at Tallinn University, for the brain storming sessions that you provided, helping me think more critically.

The Professors at Parma University and the visiting professors for enriching my study time.

The European Commission for the funding through Erasmus Mundus programme, without which this entire course would have been impossible for me. Thank you!

My DILL classmates 2008, for making the learning experience enjoyable throughout the two years, and encouraging me during hard times. Baru, for the encouragement during the thesis writing.

My friend Faustine Nakazibwe, for your support and desire that I excel in academics.

To the Great IAM for without Him, I can do nothing.

TABLE OF CONTENTS

List of Figures and Tables

Abbreviations

AIP	Archival Information Package
ALM	Archives Libraries and Museums
CASPER	Cultural Artistic and Scientific Knowledge for Preservation Access and Retrieval
CRL	Center for Research Libraries
DIAS	Digital Archive Package Structure
DIP	Dissemination Information Package
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment
GDFR	Global Digital Format Registry
FAG	Fagsystemer (For Fag system in Norwegian)
IKA	Interkommunal Arkiv (an archival institution in Norwegian language)
InterPARES	International Research on Permanent Authentic Records in Electronic Systems
LONGREC	Long Term Records Management
NOARK	Norwegian Record Keeping System
OAIS	Open Archival Information System
OCLC	Center for Online Computer Library Center
PDI	Preservation Descriptive Information
SIP	Submission Information Package
TRAC	Trustworthy Repositories Audit and Certification: criteria check list

CHAPTER 1

1.1 Introduction

This chapter gives an introduction to the research study, aim and objectives and the scope within which this research is undertaken.

1.1 Motivation

In the 21st century, a lot of information is now born digital while at the same paper based information is also being converted to a digital form with the aim of increasing its preservation ability and to ease its management as well as access. Today, access to some previous popular files like a WordStar[1] document is difficult and perhaps in some cases impossible. In addition, when particular files are used and accessed on various operating systems, they might not open correctly or their general layout and some content can be lost. Therefore, as an Information worker interested in meeting information needs of users at all times, it is important to consider access of such resources.

1.2 Statement of the Problem

Can your ten year old digital record be used as proof in a case where its authenticity needs to be ascertained? Is its current state meaningful and does it convey its original meaning?

The information age has led to the creation of vast amounts of digital documents or records with the aim of facilitating effective and wide access if needed, while saving storage space too. At the same time, the rate of evolution in technology makes software and hardware reach their point of obsolescence much earlier than information users expect. This has created the need to preserve these digital records mainly through either a migration/conversion process (where the digital object is changed as it is being migrated or converted from one file format to another) or emulation (where the environment of the digital object is changed with the aim of retaining its functionality or access), (Bussel, 2007). Migration/conversion processes can have negative effects on the digital object; sometimes these changes could be negligible depending on the kind of object while at times, they could greatly change the original object. Consequently, this brings

[1] **WordStar** was a popular word processing system that was originally written for the CP/M operating system and ported to DOS . It was used during the 1980's.

in the need to ensure that as digital preservation is done, authenticity and integrity is maintained, so that the users of the digital objects or records can trust them to be the real (authentic) objects that they once were. So how is this done? How effective are the authenticity tools? Preservation efforts that incorporate a conversion can see the original byte structure of the file changing. This is acceptable as long as the original meaning of the contents of the file is retained (Factor, et. al., 2009).

A number of collaborative initiatives from various parts of the world like, PLANETS, CASPER, the National Library of Australia, just to mention a few, are still striving to attain best preservation practices but this can only be reached if authenticity and integrity is ensured. Some studies in Norway by projects like LongRec and institutions like ALM (Norwegian Archive Library and Museum Authority) have come up with recommended best practices (LongRec, 2010). How far has the implementation of these practices come when it comes to authenticity and integrity? It is clear that Norway has a clear and strong tradition[2] in ensuring authenticity during the records management stage of a documents lifecycle but what is the situation with regards to preservation? As records management has been regulated for quite a while now Norway, it makes it an interesting case study. This research therefore seeks to explore authenticity practices in Norway's municipality and city archives.

1.3 Aim and Objectives

The research aims to investigate the current authenticity practices within city and municipality archives in Norway. Identify gaps where possible within the Norwegian context and identify an approach that can best suit these archives as they endeavor to ensure that authenticity and integrity of their digital collection is retained. The following research questions will be used to attain the above;

- What are the current authenticity practices in digital archives of Norway?
- How is authenticity maintained in the digital archives of Norway?
- What are the best recommended authenticity practices for digital records in Norway?

[2] Since 1984 the records management of electronic public administration documents has been regulated by the Noark standard.

1.4 Scope

This research will cover the city and municipality archives in Norway and specifically look at Noark records, based on the Noark standard. Noark is a standard used in the Norwegian electronic record keeping system ("Norsk arkivsystem") for Norway's public administration records. It started off in 1984 and after a series of revisions; Noark 3 was introduced in 1994 followed by Noark 4 in June 1999(Riksarkivet, 2000). Noark 3 is therefore 16 years old while Noark 4 is 11 years old. Noark 4 is currently the most widely used system in public administration for electronic record. The major specifications cover the following:

- Information content (what information should be recorded)
- Data Structure (the design of the individual data elements and the relationship between them)
- Functionality (what functions the system must support) (Riksarkivet, 2010).

Today, we are now seeing Noark 5 approved systems although there should be at least another 4-5 years before we begin to see Noark 5 documents deposited at archival institutions.

The Norwegian Records Management and transfer to an archival institution practice is regulated. After 5 years, documents are required to be submitted to an archive. For the next 20 years, documents will be stored by the archival institution but the administration entity is required to maintain its own copy. The primary reason for a deposit after 5 years is to increase the chances for data authenticity as the system will most likely still be active. After another 20 years, it is unlikely the original computer system will be active. Interestingly, the Norwegian National Archive only receives documents from state public administration. Public administration at municipality level is required to use the Noark standard for records management but not required to deposit documents to the national archive or any archival institution. To make life easier for municipalities, many municipalities have created archival institutions called IKA (Interkommunal arkiv) to benefit from scale and reduced costs.

The timescale of 25 years and the expected technological evolution, changes and obsolescence raises questions with regards to preserving the authenticity of these records This is of interest to this research to find out present authenticity practices and possibly how effective they are, so that best practices can be identified for Norway's archives.

The Trustworthy Repositories Audit and Certification: criteria check list, (TRAC) by OCLC[3] and RLG enumerates a number of issues that have to be used as guidelines for a repository to be trusted, right from when records are received at ingest, up to when they are accessed by their respective primary consumers. It further looks at the organizational infrastructure and policies (OCLC and RLG, 2007). In this research, the authenticity aspect of digital records or objects will be the major point of focus to guide the research in ascertaining whether the archived records Noark 3 and 4 can be trusted as authentic records today and for years to come based on the current authenticity practices in Norway.

Authenticity is one of the core requirements for digital preservation repositories as outlined in the 10 principles by the Center for Research Libraries. The authenticity element among the principles states that an archive should:

Maintain/ ensure the integrity, authenticity and usability of digital objects it holds over time. (CRL, 2007).

Elaboration of the above is given further in the TRAC check list; the following are the major aspects that will be used as a bench mark regarding authenticity. These are;

B1.3 – Repository has mechanisms to authenticate the source of all materials.

B1.4 -Repository's ingest process verifies each submitted object (i.e., SIP) for completeness

B1.6 - Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.

B1.8 -Repository has contemporaneous records of actions and administration processes that are relevant to preservation.

B2.5 - Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects.

B2.7 - Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).

[3] OCLC stands for Center for Online Computer Library Center, while CRL stands for Center for Research Libraries

B4.4 -Repository actively monitors integrity of archival objects (i.e., AIPs)

(OCLC and CRL, 2007).

1.5 Intended audience

This study will be relevant to both the archival community in Norway as well as communities in other countries that have similarities to the Norwegian archival context as the research will point out both strengths and weaknesses. The research will also give them an idea on how best to work with the TRAC standard since they have already expressed the need to use it in the digital archival practices as documented in digital and authentic ("digital og autentisk" in Norwegian) report, 2010. The study will also be of interest to archivists working in municipalities around Norway, giving them a better understanding of current practices and showing where the challenges lie.

The Archives Libraries and Museums (ALM) (ABM in Norwegian[4]) group will benefit from this study too. This is because Norway's digitization programs take on coordinated efforts through ALM (ABM, 2006), and therefore once good practice is attained in one sector, it can easily be passed on to another sector, and customize it to its digitization and preservation needs. In fact if information from libraries and other information institutions is well defined as far as metadata and provenance is concerned, archives can easily absorb this information with less effort for long term preservation while maintaining authenticity of the received data. This study and discussion may also be of interest to record managers and archivists in other countries that would like to get a better understanding of how records management and preservation from the authenticity perspective are dealt with in Norway.

1.6 Significance of the research

This research will assist archives in Norway specifically in planning for the proper implementation of trust and authenticity right from the authoring institutions of public service information to the archives for long-term preservation. Other international archival institutions that can relate to Norwegian archives can borrow a leaf from this study as well. The true significance of this work lies in the fact that it is, to the best of my knowledge, the first time anyone has looked at this issue from the perspective of the transfer of public administration records to archival institution and how the institution maintains trust and authenticity in these

[4] ABM in Norwegian is Statens senter for Arkiv, Bibliotek og Museum.

records. Given the Norwegian records management tradition, one expects Norway to be at the forefront of this work.

CHAPTER 2 - LITERATURE REVIEW AND BACKGROUND

2.0 Introduction

This chapter provides related literature on authenticity in relation to digital preservation. A lot of work in this domain has been found to be undertaken by collaborative initiatives like CASPER[5] and PLANETS (European projects), InterPARES[6], OCLC and the National Library of Australia, just to mention but a few. For Norway in particular, the National Library is key in as far as preservation research is concerned, together with the LongRec project. A number of the articles referred to are from such initiatives and institutions. The main key words used while searching for literature are "digital preservation and authenticity".

Authenticity is defined as "the quality of being authentic, or entitled to acceptance", while the term authentic means "worthy of acceptance or belief as conforming to or based on fact" (InterPARES, 2001, p.2). It can also be referred to as the "trustworthiness of a record that is what it purports to be, un-tampered with and uncorrupted" (Duranti, 2009, p.16). The International records management standard (ISO 15489) has the same definition and further adds that "an authentic record is one that can be proven to have been created or sent at the time purported"(ISO, 2001, p.7). Duranti (2009) adds that authenticity is based on identity, integrity, and reliability of the system. The ISO 15489 recommends organizations to ensure authenticity by implementing "document policies and procedures which control the creation, receipt, transmission, maintenance and disposition of records to ensure that record creators are authorized and identified, protected against un authorized addition, deletion, alteration, use and concealment". Factor et. al. (2009) states that authenticity of a record must be supported by evidence in relation to its history, that is, the preservation treatments that the record has gone through over time. Therefore its reliability is based upon its complete documentation since creation plus the chain of processes it has gone through over time. Therefore authenticity

[5] CASPER stands for Cultural Artistic and Scientific Knowledge for Preservation, Access and Retrieval while PLANETS stands for Preservation and Long-term Access through Networked Services.

[6] InterPARES stands for the International Research on Permanent Authentic Records in Electronic Systems.

considers the information resource plus the provenance information. In practice however, it is more practical and easier to prove authenticity from the point you took ownership of the records.

Authenticity is a crucial aspect in digital preservation, without it, preservation efforts are greatly pre-empted. This is backed up by a number of scholars in their definitions of digital preservation. For instance, according to Ross(2007, p.1) "Digital preservation is about maintaining the semantic meaning of the digital object and its content, about maintaining its provenance and authenticity." Pennock (2006) defines it as a series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value. These definitions clearly state the importance of authenticity within digital preservation.

2.1 Authenticity elements: Identity and Integrity

Identity distinguishes a record from all other records. It refers to "the whole of the attributes of a record that characterize it as unique and that distinguishes it from other records" (Duranti, 2009, p. 17). It further includes the general context as well, for instance legal and technological. Factor et al. (2009) describes it in relation to Preservation Description Information (PDI) which includes Context, Provenance, Fixity and Reference as defined in OAIS model. The PDI elements have to be maintained together as a cluster of relationships defining the resource or object within particular boundaries, yet maintaining relationships which provide complete meaning to the object.

A record has integrity when the message it is meant to communicate in order to achieve its purpose is unaltered (Duranti, 2009). Integrity aims at "ensuring that a data record is accurate, complete and not modified in an unauthorized way" (Groven, et. al, 2008, p.40). The essential characteristic of an object therefore should be unchanged in spite of technological obsolescence (Factor et al., 2009). It is important to note that much as the bit stream might change, the content should be retained. This therefore demands understanding the resource, its characteristics and evaluating their role so as permit certain changes during preservation without losing integrity at the same time. Some of the tools used in authenticity management are described below.

2.2 Examples of Authenticity Management Tools

To manage identity and integrity of digital records, measures and tools need to be in place across the entire chain of custody right from creation. Such tools should be able to assess the level of authenticity that is, the completeness or changes that a record or digital object has undergone. Verification of authenticity is of paramount importance and maintaining authenticity as well (Factor et. al, 2009). Authorship and provenance are some of the key elements here.

From a technical point of view, identity and verification of files and records must be in place. If you are not able to identify or verify a file or document, what it is meant to be, integrity and authenticity become difficult to reason about. For the identification of file formats, a number of tools have been developed:

a) PRONOM, an online technical registry that provides authoritative information about data file formats and their supporting technical requirements, including supporting software products. It was developed by the preservation department of United Kingdom National Archives with the aim of supporting accession and long term preservation of electronic records (National Archives UK, 2010).

b) DROID (Digital Record Object Identification) is a software tool for automated batch identification of file formats with a link to a central registry of technical information that provides more information about the identified file format and its dependencies.. It is under the umbrella of PRONOM. It is java based and platform independent and freely accessible under the GNU Lesser General Public License (LGPL) (National Archives UK, 2010).

c) GDFR (Global Digital Format Registry) – aims at providing sustainable distributed services that facilitate discovery, storage and delivery of representation information about digital objects. It is being spearheaded by Harvard University Library (GDFR, 2010). GDFR has joined hands with PRONOM to form the Unified Digital Formats Registry (UDFR).

Verification of files is also an extremely important aspect as it is important to not only be able to identify a digital object, but also to verify the file and its contents. The following tools can be used for verification:

a) JHOVE – Jstor\Harvard Object Validation environment. JHOVE provides functions to perform format-specific identification, validation, and characterization of digital objects. It has been developed by Harvard University and freely available under the LGPL. It is a java tool as well (National Archives UK, 2010).

Some projects like PLANETS have developed tools for file identification and verification, all in one suit. In the case of PLANETS, it is PLATO, a tool based on PRONOM (Billeness, 2007). PLATO contains a service registry, migration tools like CRiB from Portugal and MiniMEE which does both migration and emulation functions. It is also a test bed that can be used by institutions to try out possible preservation actions before they undertake work on their collections. It can facilitate the entire preservation planning process (Becker, 2010), considering authenticity as well through its characterization process.

Identification and verification are two important aspects of authenticity. Other technologies supporting the authenticity aspect include the following;

a) **Digital Signatures**

These provide integrity checks through the use of hash algorithm technology. An example of this is got from digital signatures as used on Fedora[7] repository where digital signatures are computed for digital masters as well as derived objects. The signature is stored in the technical metadata of the object. Periodical re-computation of the hash of each byte stream is done and compared with the original computed hash. In case of any differences, they are reported and offline storage or mirrored repositories are used to restore the integrity of the object (Jantz and Giarlo, 2005). Lynch (2000), views digital signatures as a computation on data using a private/public key pair. The public key enables verification of known data to have been computed by a particular entity holding the key pair. Off course this all depends on trust given to the public key infrastructure operator since “trust is not necessarily an absolute, but often a subjective probability that

[7] Fedora stands for Flexible Extensible Digital Object Repository Architecture. It is a software for management and preservation of digital repositories

we assign case by case" (Lynch, 2000, p.46). The LongRec check list on preservation of trust states that the "evidential value of digital signatures might decrease because of life time (expiry, revocation) of keys and certificates used or signing methods" (LongRec, 2010). Therefore nothing is absolute with technology and therefore constant updates and revisions are required along the way to ensure that authenticity is maintained.

b) **Persistent Identifiers (PID**)

Web references are very unstable unless Uniform Resource Names (URNs) are used. The URN is a generic form of persistent identifiers that can be used for the entire lifetime of a digital object and it is therefore permanent and unique. This brings in an element of referential integrity or citation persistence as outlined by Jantz and Giarlo (2005). A URN comprises of a Namespace Identifier (NID) code that identifies the system being used, and a Namespace Specific String (NSS) which identifies a specific document. For example the 'ISBN' and 'ISSN' are registered as NIDs for URNs by the international ISBN and ISSN agencies. The persistent identifiers or "handles" are assigned, managed and resolved by a Handle System for managing digital objects and other resources on the Internet. A local handle service can also be integrated with the global system (NLA, 2002).

However the continued success of URNs greatly depends on the ability of organizations like the Corporation for Research Initiatives (CNRI) in charge of the Handle system, and others that avail URNs to preserve them forever and the repository staff to implement sustainable preservation policies and work flow practices. Never the less, the ideology of persistent identifiers goes a long way in facilitating referential integrity and therefore authenticity practices for digital content. Other types of persistent identifiers include:

- Digital Object Identifier (DOI)

  These are under persistent identifiers, but are mainly for commercial purposes through electronic commerce and copyright management for the publishing community. The DOI was initiated by the association of American publishers and currently managed by the International DOI foundation (NLA, 2002).

- Archival Resource Key (ARK)

This is a persistent identification designed for custodians of archived digital objects. The principle of provenance is key plus naming schemes over time. It is protocol independent. The scheme consists of three requirements: a link from the object to a stewardship promise; a link from the object to its metadata describing it; and another link to the object itself (NLA, 2002).

c) **Naming conventions**

To facilitate the proper use of persistent identifiers above, the naming convention of PIDs should be independent of technology, protocols and local naming conventions. Examples of naming conventions include CNRI handle syntax by CNRI Global Registry and Archival Resource Key (ARK) by the California Digital Library (Jantz and Giarlo, 2005).

d) **Digital water marks**

These have been mainly used to protect intellectual property by including a copyright claim as a water mark. However, they tend to intentionally corrupt objects where they are applied just like in cases of lossy compression (Lynch, 2000). On the other hand, if the water mark can easily be removed (due to bad water marking systems); still getting to know any other aspects that could have been corrupted in an object becomes difficult too. Therefore in light of the authenticity aspect, Lynch (2000) recommends using it for asserting a claim on the digital object and then have this claim verified through a digital signature. However, digital water marks do provide evidence of provenance which is crucial to authenticity as well.

e) Secure storage and access

To ensure data or record safety, technologies like Storage Area Networks **(**SAN) have been deployed by some information institutions and archives. For instance the Norwegian National Library (Riksarkivet, 2010). This is important to ensure that integrity and authenticity efforts are not put at risk, therefore safe storage measures are of paramount importance too. Secure storage is also realized through secured access control as highlighted by Groven et. al.(2008). That is, only authorized persons should have access to stored data and this is through the use of authentication technologies like digital signatures as applied in online banking transactions.

Digital preservation has taken on the major options of migration/conversion (change in the digital object as it is being migrated or converted) and emulation (change in the environment of the digital object), (Bessel, 2007). Therefore "the authenticity of digital resources is threatened whenever they are exchanged between users, systems or applications or any time technological obsolescence requires for an updating or replacing of the hardware or software used to store, process, or communicate" (Factor et. al, 2009, p.3). Again, the ease at which electronic records are created, modified and transferred emphasizes the importance of maintaining their integrity (Hirtle, 2002). These changes need to be captured over time within their context thus bringing in the aspect of provenance. Provenance refers to a cumulative record, describing the events in the life of content data since its creation (Factor, et al., 2009).

2.3 Authenticity and provenance from a historical perspective

Some studies have traced authenticity especially in ancient diplomatics and therefore tried to use its theoretical framework within digital preservation. Diplomatics, a core tool in archival science seeks to answer or provide a theoretical framework to provenance questions like who created digital content, when it was created, where, by whom among others (Ross, 2007). Diplomatics is a discipline or study originally developed in the seventeenth and eighteenth centuries, with the aim of ascertaining the integrity and authenticity of documents (Hirtle, 2000). However, Diplomatics "has been criticized for being very traditional in its record conception, therefore quite limited when applying it to electronic systems and the variety of entities contained therein" (INTERpares, 2001, p.33).

Never the less, the International Research on Permanent and Authentic Records in Electronic Systems(INTERpares), is among those that have used archival science and diplomatics in finding answers to authenticity issues (Hirtle, 2000). Ross (2007), highlights major principles in archival science and diplomatics that are relevant to any information object and therefore includes digital objects as well. These are authenticity, provenance, trust and context. He further adds description and arrangement plus repository design and management.

The InterPARES project did set up an authenticity task force and produced its report in 2001 which successfully developed a conceptual framework for establishing the requirements for preserving authentic electronic records. The Authenticity Task Force successfully "developed two sets of requirements that support the presumption of the authenticity of electronic records before they are transferred to the preserver's custody; while the second set includes requirements that support the production of authentic copies of electronic records"(InterPARES, 2001, p.1). Therefore this highlights the fact that it is important to have particular characteristics of records to be regarded as authentic, before they are preserved digitally (at ingest), maintain them as authentic during storage (Archival Information Package – AIP) and deliver them as authentic (Dissemination Information Package - DIP). These concepts are mainly used in relation to the OAIS model, an internationally accepted archival model.

Another relatively recent school of thought is from the Pittsburgh project which aims at setting up systems that can capture metadata automatically (Hirtle, 2000).This project mainly emphasizes metadata capture and not provenance as the later. The metadata approach is seen in works on Preservation Metadata Implementation Strategies (PREMIS) and a dictionary on the same has been generated with working manuals as well (PREMIS, 2005). It is also important to highlight the fact that the PREMIS working group incorporated both provenance and metadata approaches for authenticity purposes and builds on the OAIS model as well. The PREMIS data model is practical and independent of any metadata type or syntax (PREMIS, 2005). A closer look at the OAIS model will clarify and give a good quick start in comprehending authenticity aspects.

2.4 The Open Archival Information System

The development of the Open Archival Information System (OAIS) model was spearheaded by the Consultative Committee for Space Data Systems (CCSDS) in 1995, having realized that there was no standard by then to cater for digital preservation over a long time. A number of drafts were made 1997 to 2000 with a number of reviews, and later on adopted as an ISO standard 14721 in January 2002 (Lavoie, 2004). The OAIS model or OAIS archive type can well be defined by the following responsibilities it has to accomplish.

i. *Negotiate for and accept appropriate information from information producers*
ii. *Obtain sufficient control of the information in order to meet long-term preservation objectives*
iii. *Determine the scope of the archive's user community*
iv. *Ensure that the preserved information is independently understandable to the user community, in the sense that the information can be understood by users without the assistance of the information producer*

v. *Follow documented policies and procedures to ensure the information is preserved against all reasonable contingencies, and to enable dissemination of authenticated copies of the preserved information in its original form, or in a form traceable to the original*

vi. *Make the preserved information available to the user community*

(Lavoie, 2004, p.3).

Considering the above mandate in section (v) that looks at authenticity of the preserved information to the designated community. It guides librarians and archivists to preserve, be more careful with authenticity elements in regard to their designated primary community or users. The designated community refers to particular persons within a particular discipline or category of people. For instance they could be lawyers, medical personnel or architects. Therefore if the archive is primarily for medical personnel, the lawyers might find it a little difficult to comprehend and vice versa. The archive always considers the designated community first before any other user group.

The OAIS model thrives in the environment of the producers of information, the archive (in this case the OAIS archive type), consumers, primarily the designated community, and management that oversees the archive.

The OAIS functional model

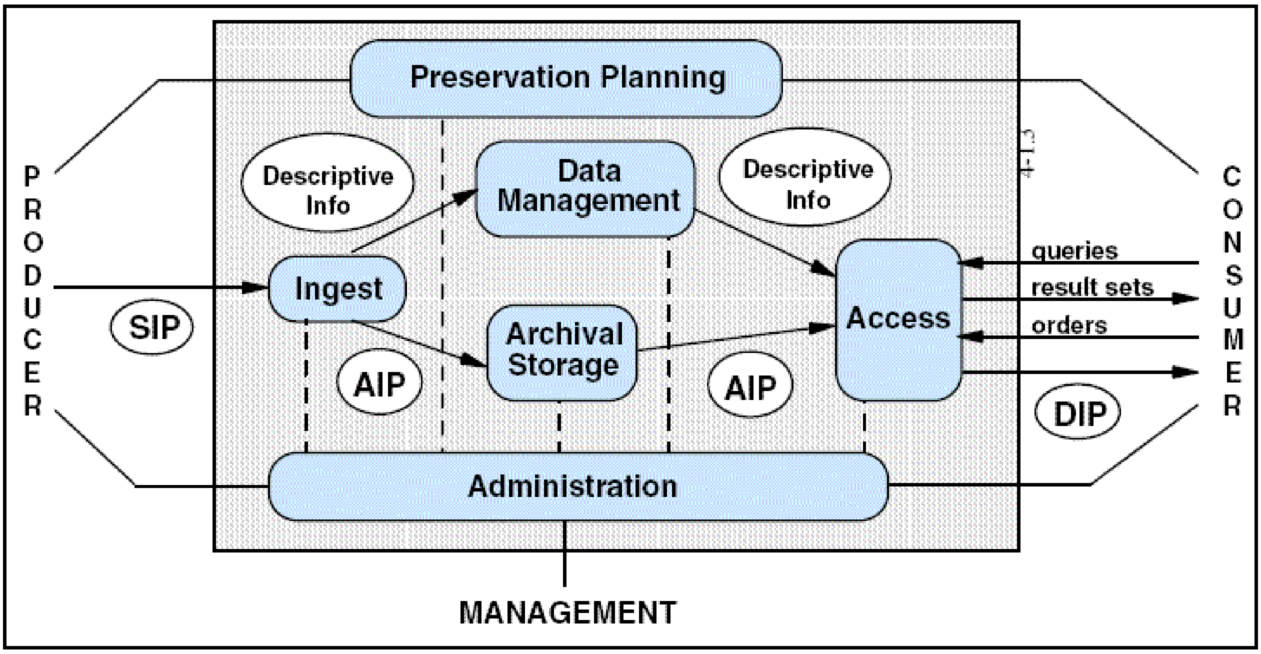


Figure 2.4.1 (OAIS Model) From Reference ***Model for an Open Archival Information System (OAIS) CCSDS 650.0-B-1 Blue Book)***

As illustrated in Figure 2.4.1, the first stage is at Ingest when the producers bring information to the archive, at this point referred to as Submission Information Package (SIP). Quality checks are carried out at this stage for the submitted information to ensure that it is not corrupted and is complete and updates coordinated. An Archival Information Package (AIP) is created for Archival storage and descriptive information extracted to aid retrieval functions of the archive (Briguglio and CASPER foundation team, 2008).

At the second functional stage, which is the Archival Storage, the safety and maintenance of the AIP is catered for to ensure that it can be accessed over a long period of time (Briguglio and CASPER foundation team, 2008.). To ensure that this objective is met, preservation motivated actions like refreshment and migration are done plus error checking procedures to ensure that AIP is still in a good state after the preservation procedures or mitigate certain risks that could have come up as a result of these actions. The Archival storage also provides the function of responding to access queries from information consumers (Lavoie, 2004).

The Data Management stage, deals with populating descriptive information, maintaining it and providing access to it (Briguglio and CASPER foundation team, 2008). This entails performing queries and generating reports in response to requests from other functional components within the OAIS; as well as updating the databases (Lavoie, 2004). As such, search and retrieval of archived content is supported.

At the Preservation planning stage, the OAIS environment is monitored in relation to external environment especially in terms of technological changes, mapping out preservation strategies and providing recommendations (Briguglio and CASPER foundation team, 2008). This stage facilitates detection in primary users and technological changes "impacting the OAIS's ability to meet its responsibilities, designs strategies for addressing these changes, and assists in the implementation of these strategies within the archival system" (Lavoie, 2004, p.9).

The Access function deals with query processing, retrieval and delivery of information to consumers. Data Security of the archive is provided plus access control (Briguglio and CASPER foundation team, 2008). The Access function provides an interface that facilitates access of its archived content to the user community.

The last functional component, administration is responsible for managing the day-to-day operations of the OAIS, as well as coordinating the activities of the other five high-level OAIS services. Other responsibilities include liaising with producers for instance when negotiating submission agreements[8], and consumers for customer care, Management for policy and standard management (Lavoie, 2004).

The OAIS model purposes to maintain the integrity of information received. This can further be seen in examining the AIP package.

2.4.2 Authenticity and the Archival Information package (AIP)

The OAIS deals with preservation of AIP which is finally delivered to the user as a Dissemination Information Package (DIP). A closer look at the components of this package will help highlight some authenticity issues for better comprehension.

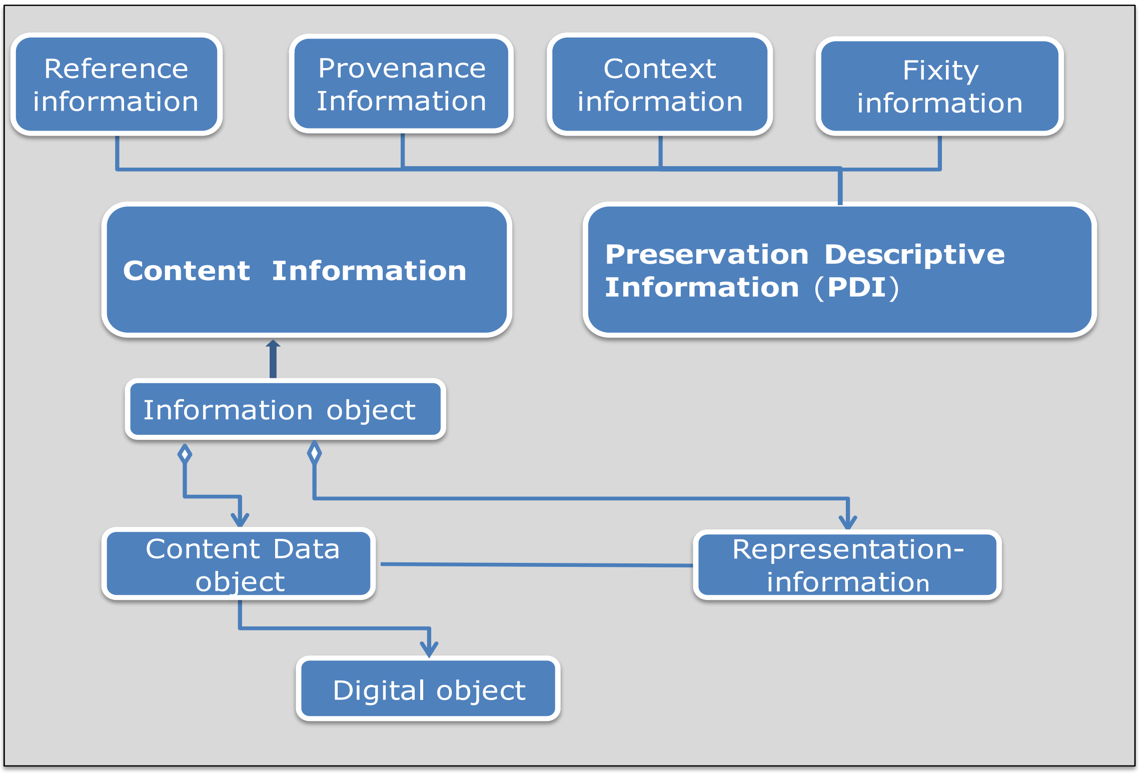


Figure 2.4.2 AIP structure (illustration from class notes, 2010).

The archival package is comprised of the content data object initially which is derived from the digital object and this might be any type like text, image or sound. Actually this is the object that needs to be preserved over a long period of time. To make it understandable to the designated community, representation information is supplied and this may take the form of technical

[8] A submission agreement is an understanding between the information producer and the archive specifying a data model for the data submission session (CCSDS, 2002).

information. For instance hardware or software needed for the file to open or program to run and a brief explanation of the content data object is also given. The content data object and its representation information create the Content Information.

To preserve the Content Information over time, additional supporting metadata is required and this is collectively referred to as Preservation Description Information (PDI). The PDI is comprised of reference information which gives a unique identity to the Content Information within the OAIS and externally (this might be a digital object identifier), the context describes the relationship of this data object to other data objects within the archive, provenance documents the history of the data object since its creation and subsequent preservation changes that it has undergone or any change of custody over time. Fixity Information validates the integrity or authenticity of the Content Information. This might be through the use of checksums, water marks or digital signatures (Lavoie, 2004).
However the OAIS model is taken to be more of a reference model than an implementation model when dealing with preservation aspects especially the authentication element. Therefore PREMIS and METS standards come in handy to deal with authentication metadata at the actual implementation level, when dealing with container packages (Riksarkivet, 2010). Already some Nordic Libraries have embraced PREMIS and METS, for instance the Swedish National Archives. (Riksarkivet, 2010). Never the less, the OAIS model gives a good understanding of major preservation aspects as highlighted above.

2.5 Authenticity from an organizational risk management perspective

More recent research (from 2006 to 2008) by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE) has led to the development of “Digital Repository Audit Method based on Risk Assessment” (DRAMBORA). This repository audit methodology encourages institutions to have comprehensive self assessment, right from institutional objectives, policies and looking at digital curation as a risk management activity within their institutions. A digital curator’s central role under DRAMBORA has been defined as “rationalizing the uncertainties and threats that inhibit efforts to maintain digital object authenticity and understandability, transforming them into manageable risks” (DCC and DPE, 2008). A DRAMBORA toolkit has been developed to facilitate this process and provide provision for assessing risk from physical to human resources and the entire operation of an institution. Examples of risk assessment preservation aspects in this toolkit include evaluation of:

- Destruction of primary documentation
- Effectiveness of technical infrastructure and Security
- Completeness of submitted packages
- Externally motivated changes or maintenance to information during Ingest (Does repository obtain full physical and intellectual control of submitted content?)
- Loss or maintenance of authenticity of information
- Loss or maintenance of integrity of information (DCE and DPE, 2007).

With such a risk management strategy embedded within institutions, it is more likely that proper curation of information is well taken care of, with a preservation perspective for long term access. Information is treated as a more valuable resource and therefore worth ensuring that authenticity aspects are well taken care of for reference, historical with evidential value over a long period of time. A number of libraries and archives have used DRAMBORA and these include;

- The British Library
- CERN Document Server, Switzerland
- National Archives of Scotland, Edinburgh, UK
- National Archive of the Netherlands e-Depot
- National Library of the Czech Republic
- National Library, Florence, Italy
- Netarkivet (Danish Internet Archive), among many others (DCC and DPE, 2008).

2.6 Background information and authenticity practices in Norway

The Public administration of Norway has a defined electronic record keeping system tailored to meet the functional requirements of public administration and it is referred to as Noark (Norwegian record keeping system) or "Norsk arkivsystem" in Norwegian. The current Noark 4 standard is a revision of the first Noark standard that began in 1984. Noark 4 came into existence in 1999, while the later Noark 3 was introduced in 1994 (National Archives of

Norway, 1999). Therefore a well regulated record system has been in existence for at least 25 years so far.

The Archives and Regulations act[9] of Norway groups record systems into three main categories basing on functionality, namely;

- Record keeping systems without fully electronic storage of documents, the minimum records should therefore be Noark 3 or Koark[10].
- Record keeping systems with fully electronic storage of documents and these are meant to follow Noark standard according to the ABM guidelines of 2007.
- Databases and document management systems without record keeping and without electronic storage of documents. The National Archivist may require a built in export functionality (abm-utvikling, 2007).

Based on the above information, Norwegian archives are most likely, semi electronic archives. This is further confirmed in module 5 of the Noark 4 standard which states that, "it should be able to handle both paper-based and electronic storage of documents since Noark allows for combined storage of cases with paper documents and cases with electronic documents" (National Archives of Norway, 1999, p.5).

In reference to the above, minimum standards have been stipulated in the Noark 4 standard specifying that email documents should be "based on SGML[11] syntax, with the name of the sender (organization), case title, case and document number, date and description of contents as well as a unique reference to the registry entry" (National Archives of Norway, 1999, p.4). This information is resourceful for authentication purposes since identity and provenance data is captured. A look at chapter 10 of the Noark standard states that internal authentication is well implemented through Noark's automated registration of persons responsible for performing key activities plus activity logging functions. The system further provides for the option of applying digital signatures to document versions as part of the internal processing activity (National Archives of Norway, 1999).

[9] Archives and Regulations act can be accessed at http://www.lovdata.no

[10] KOARK was a specific records management standard for municipalities. KOARK became part of Noark 4

[11] SGML stands for Standard Generalized Markup Language

The digital and authentic planning report by the National Archives of Norway (2010) highlights authenticity as a cardinal requirement for archival materials for they are unique products of actions and events and these can be used as evidence. Therefore integrity of digital documents must be maintained continuously during migrations, conversions and other maintenance operations carried out in an archival repository (Riksarkivet, 2010).

In light of the above information, some institutions in Norway seem to be implementing authenticity practices to some extent based on the fact that public institutions are using Noark 4 standard which incorporates such measures. However, at times they are not able to put in place facilities that can enable this throughout the preservation process. For instance, the National Library of Norway seems to be keen in following up the authenticity aspect as revealed in one of the case reports by LongRec stating that, "there is no integrity checking in the transfer of files from the production stage to the storage area, whether the files are produced in-house or externally, except for in-house digitization of photos" (Cerrato, et. al, 2008, p.7). This report however provides integrity check practices when data integrity is monitored on file movement through fixity checks. This shows that integrity is desired though at times cannot be ensured throughout each preservation process due to various reasons.

CHAPTER 3- METHODOLOGY

3.1 Purpose of the research

The research explores current authenticity practices within digital archives of Norway. The results will consequently identify current practices, identify gaps and identify an approach in which best authenticity practices can be adopted in Norway's city and municipal archives from the point of view of digital records preservation and authenticity.

3.2 Theoretical framework

The research uses pragmatic knowledge claims in its philosophy. That is practical approaches to the research problem (Denscombe, 2007). The pragmatic view seeks to find answers to the research problem. The main focus is on what works and solutions to the problems, Patton, (as cited in Creswell, 2007). The pragmatism world view "focuses on the consequences of the research, the primary importance of the question asked rather than the methods and the multiple data collection informs the problems under study" (Creswell and Clark, 2007, p. 23). This view looks at "what" and "how" to research (Creswell, 2009). A number of writers have embraced this world view and these include Patton (1990), Murphy (1990), Rorty (1990), Cherryholmes (1992) and Tashakkori and Teddlie (2003), (as cited in Creswell, 2007). In this case, the research seeks to answer the following questions.

- What are the current authenticity practices in digital archives of Norway?
- How is authenticity maintained in the digital archives of Norway?
- What are the best recommended authenticity practices for digital records in Norway?

3.3 Research Design/Approach

Today, the three major research designs are quantitative, qualitative and mixed methods research approaches. Quantitative tends to consider the objective reality of social facts, qualitative design considers social construction of reality (Gorman and Clayton as cited in Pickard, 2007) while mixed methods uses both a quantitative and qualitative form of inquiry. The mixed methods research embracing the pragmatic philosophy has been used in earlier studies by ethnographers like (LeCompte and Schensul, 1999) and case study researchers: Luck, Jacksson and Usher, 2006; Yin, 2003 (as cited in Creswell, 2007).

The mixed methods approach has four major designs. That is, triangulation design, embedded design, explanatory design and exploratory design, (Creswell and Plano, as cited in Creswell, 2008).

This research will take on the explanatory mixed design method (also called a two phase model) which "consists of collecting quantitative data and then qualitative data to help explain or elaborate more on quantitative results"(Creswell, 2008, p.560). The quantitative data from the survey gives a general picture of authenticity practices on Noark 3 and 4 records in city and municipality archives of Norway, while qualitative data from the interviews provides an in-depth exploration on authenticity practices as well as filling in any missing gaps that could have arisen from the first quantitative data collection. The explanatory design is a more clear design for mixed methods since it is easier to implement. A particular set of data is collected at a time and therefore a single researcher can easily manage it (Creswell and Clark, 2007). According to Denscombe (2007), mixed methods approach gives room for validation of findings, provides a more comprehensive picture as complementary data could be generated from different methods and provides a way of compensating strengths and weaknesses of methods used.

3.4 Research Strategy

The research uses mainly a case study strategy which aims at providing a "holistic account of the case and in-depth knowledge of the specific through rich descriptions situated in context. This may lead to an understanding of a particular phenomenon" (Pickard, 2007 p.86). Yin defines a case study as an empirical investigation on contemporary phenomenon within its real life context (Yin, 2009). In this case, the authenticity phenomenon of digital records. According to Stake (as cited in Creswell, 2008) case studies explore an event, program, process or activity in depth and are restricted by time and activity and therefore researchers have to collect detailed data from multiple sources. Samset (2000) too views case studies as having a combination of various data collection methods. For instance, the use of questionnaires and interviews. Denscombe (2007) supports this view when he says that a variety of research methods can be used or are rather encouraged. This research aims to investigate authenticity as the phenomenon in this case, using the case of city and municipality archives in Norway, and Noark 3 and 4 in particular to understand this aspect. *Authenticity in this case is the unit of analysis as applied to Noark 3 and 4 records.*

An instrumental case study has been chosen as the best option, for the research is interested in the phenomenon of authenticity as applied to Noark records and not the case study site as such (Pickard, 2007). The municipality archives have been used as a channel for this investigation. This study realized the need to use the survey strategy too as explained below.

The survey strategy is used too to reach out to the entire population of the city and municipal archives in Norway. The aim of the survey is to collect and analyze information on a representative or entire population (Pickard, 2007). According to Denscombe (2007), surveys have a characteristic of having a wide and inclusive coverage to present a particular picture prevailing at a particular point in time). In this case the survey is used to present the current state of authenticity practices in city and municipal archives of Norway. A descriptive survey is used as it can describe a situation within a defined representative population or entire population. The fact that survey research embraces both qualitative and quantitative research(Pickard, 2007), it is therefore suitable for this particular research study since both research methods are used as described in the research design above.

3.5 Instruments design

a) The questionnaire provided brief background about the research stating its objectives to the respondents. The questions were both closed (for quantitative data generation) and open ended (for qualitative data elaborating various aspects). The questions were derived from TRAC, considering record integrity and authenticity elements of TRAC and then adjusted to meet the Norwegian context. The pilot phase with one of the city archives made the adjustment possible. The detailed questionnaire is available in appendix I. The following aspects were covered following the same sequence as in the questionnaire.

Background information on electronic records within the archive

Archives had to specify whether they had Noark 3, 4 or Koark records and the period of time they had taken care of those records. The size of their repositories in gigabytes and also ascertain whether their collections had undergone any preservation process.

Record integrity measures upon receipt of new archival records

This section was interested in tools used for ascertaining record integrity such as checksums, policies, mechanisms for validation like submission agreements or digital signatures and the extent to which such tools were reliable. Handling of records at the ingest stage was also important to this research.

Electronic storage of records

The format of deposited files was investigated, management of the conversion process to match the archival file format was considered, the use of persistent identifiers and the maintenance of trust within the archival repositories.

The use of international standards in record keeping

The respondents had to clarify whether they use any international registries like PRONOM, the use of the OAIS model and clarify this by describing any of the model principles. The respondents finally had the option of giving any other data that could be resourceful to the research.

b) Interview guide

This was structured in such a way that it was probing further what had been realized from the questionnaire. For instance, at this point the researcher had established that Noark records were actually in paper form, so there was need to find out the actual structure of Noark records in relation to what had been obtained from the survey since they were using checksums and some of the OAIS principles. This showed that some aspects were in electronic form. The interview therefore investigated the following;

- The procedure for handling incoming electronic data including the integrity check measures and the challenges experienced while endeavoring to maintain the authenticity of electronic records. This was in question 2 of the interview guide.
- Question 3 dealt with the strength of the submission agreement as a validation tool for deposited data plus its components.
- The identity of individual files was investigated and the handling of various file formats in questions 4 to 6.
- Question 7 addressed the maintenance of integrity during conversion while question 8 explored how OAIS model is used in the archives.

- The researcher was interested in understanding the entire work flow in question 9 and any other comments from the interviewees relevant to the research in question 10.

  (Refer to appendix III)

Using the above instruments the researcher was able to get sufficient information for the research study.

3.6 Data Collection techniques and analysis

Being an explanatory mixed design method, online questionnaires were first distributed to city and municipal archives within Norway, after which interviews were carried out within a few selected archival institutions that were willing to participate. Initially, with the guidance of my supervisor, I got the list of all city and municipal archives. The research supervisor did the initial contacts at the archives, introducing me for the data collection exercise. This was through email and telephone calls after which I followed up the communication with the archives. The data collection and analysis process is described below:

i) A web based questionnaire – this was the first data collection tool used and it comprised of mainly closed and a few open ended questions. It was administered through quest back, a web based survey tool. The closed questions provided quantitative data while the open ended questions provided qualitative data also referred to as categorical or discrete data by Pickard (2007). The web based questionnaire is much easier for the respondents since they have predefined answers to choose from in case of closed questions and submission is just a click away. On the other hand, the researcher can easily transfer results to a spreadsheet quickly with accuracy too (Denscombe, 2007) just like it was done in this study.

   To ensure that the questionnaire is well designed, a pilot phase was done through the participation of one archivist from one of the city archives and one professional in archives. After some amendments, it was distributed online to all city and municipal archives in Norway. Telephone calls were made to remind them to respond to the online questionnaire.

   To make the questionnaire user friendly to the respondents, the questionnaire was administered in the Norwegian language, and a copy of every response was automatically available via the researcher's email through quest back. One of the faculty staff at Oslo University College (a Norwegian native speaker) helped the researcher with translating questions into Norwegian, after which the researcher uploaded them to questback.

The responses were translated to English through Google scholar, coded through the development of concepts and categories. The researcher then had to go through the data in an iterative process (Denscombe, 2007), that is going over data again and again to generate useful information which was fully analyzed through an excel worksheet.

ii) Interviews - These followed after collecting responses from the questionnaire. Six archivists from six different institutions were interviewed, two of which were face to face interviews with key informants within the archives. A structured interview guide in both English and Norwegian languages was sent out to respondents a week earlier to enable them get familiar with the questions. The interviews provided a follow up advantage to some of the responses from the online questionnaire. The two face to face interviews held lasted between one to one and a half hours each. These interviews enabled the researcher to seek more clarification on pertinent issues and adjust questions easily based on the responses of the interviewees since they were held in English. The interview questions were semi structured and this provided the interviewees the opportunity of developing ideas and elaborating more on them since the interviewer is flexible in topic (Denscombe, 2007) or question order. The interviewees were recorded with their permission using a tape recorder and notes were made during the interview sessions too. The researcher later transcribed the responses to Microsoft Word and coded the responses using keywords from the interview responses.

The other four interviews were administered through e-mail following the face to face interviews. This is because it was not easy to get archivists to participate in person and fortunately the email interviews were very informative too. Since these were held after the face to face interviews, the researcher got an opportunity to probe further, based on the responses received from the earlier interviews and affirming certain responses. The email interviews were administered in both English and Norwegian languages. The respondents chose to reply in their native language Norsk or Norwegian and these responses were translated through Google scholar. For some inaccurate translations, the services of a Norwegian information scientist were sought to ensure that all the information is well translated without losing the intended meaning.

iii) Document analysis – this comprised of some of the city and municipal archive websites, the ABM handbook on digitization for long-term access, Riksarkivet website, Noark 3 and 4 standard documents. The references from these documents are well captured in the discussion section of chapter 4.

Triangulation in data collection has been used to affirm consistency of findings (Neto, 1997) thus providing a way for checks and balances.

3.7 Population

The target population included all city and municipal archives in Norway. These are: Oslo byarkiv, IKA Østfold, Fylkesarkivet i Oppland, IKA Kongsberg, Fylkesarkivet i Vestfold, Aust-Agder Kulturhistoriske Senter, IKA Vest-Agder, IKA Rogaland, IKA Hordaland, Fylkesarkivet i Hordaland, Fylkesarkivet i Sogn og Fjordane, Bergen byarkiv, IKA Møre og Romsdal, IKA Trøndelag, Arkiv i Norland, IKA Troms and IKA Finnmark. However, only 5 archives responded to the online questionnaire.

3.8 Limitations of the Study

The study was affected by the public administration strike where in some cases, members of the city archives were on strike during the data collection period. It lasted for two weeks and therefore even some archivists, who had committed themselves to respond upon communicating with them on telephone initially, finally did not respond after returning from the strike. We believe this to be a result of the fact that they had a lot of work to catch up on and unfortunately could not prioritize partaking in this survey.

The communication system in Norwegian archives being centralized became a hindrance instead of an advantage. I was advised to use 'postmottak' mail – a central mail address for each archive to ensure coverage. However getting feedback very much depended on who was in charge of this mail and his or her willingness to respond or pass it on to the relevant person. Unfortunately "postmottak" hinders the identification of the relevant people to talk to at the institutions. On calling the institution to find out where to direct the survey and questions, the answer is that the institution prefers all communication to go through postmotakk where it will be dealt with in a timely manner. The institutions seemed reluctant to identify relevant personnel. These two factors explain why the feedback was not as high as it could have been.

Language was also a hindering barrier for the researcher to some extent since she had to spend a lot of time translating documents from Norwegian to English. This therefore implies that some relevant literature to study might not have been used though the most essential documents have been used in this study.

As discovered from the study, these archives do not have full electronic records, they consequently did not feel that they were obliged to respond since the authenticity aspect has not been fully tackled in their daily work. The Noark 4 standard came in 1999 and was the first standard that supported full electronic records management. Given the deposit rule where records must be deposited after 5 years, we expected to see some electronic records at the archival institutions. This was not the case, and was a surprise and something dealt with in the qualitative work of this research.

3.9 Ethical Considerations

The anonymity of respondents and archives has been reserved in the research study as illustrated in the data analysis and presentation of findings. The gathered information has been used to meet the needs of the study without abusing the privacy rights of the archives that participated in the study.

CHAPTER 4 - FINDINGS AND DATA ANALYSIS

4.0 Introduction

This chapter provides a presentation of data from the survey and interviews. Having used an explanatory mixed method, the quantitative results are presented first, followed by the qualitative data. This was the very sequence that was used during data collection. Deductions will then follow basing on the data analysis.

4.1 Survey data response

An online questionnaire (comprising of closed and open ended questions) was sent out to all 17 archives in Norway, that is city and municipal city archives. These are: Oslo byarkiv, IKA Østfold, Fylkesarkivet i Oppland, IKA Kongsberg, Fylkesarkivet i Vestfold, Aust-Agder Kulturhistoriske Senter, IKA Vest-Agder, IKA Rogaland, IKA Hordaland, Fylkesarkivet i Hordaland, Fylkesarkivet i Sogn og Fjordane, Bergen byarkiv, IKA Møre og Romsdal, IKA Trøndelag, Arkiv i Norland, IKA Troms and IKA Finnmark.

Only five of the archives responded. This could be attributed to the fact that archival emails arrive at one central address- postmottak at every archive after which they are passed on to the relevant archivist. Therefore getting feedback greatly depends on the person in charge of postmottak mail, forwarding it to the appropriate person and the appropriate person responding to this mail. It could also be due to the fact that archives have not fully developed the aspect of authenticity and trust and therefore they could have felt that the online questionnaire was not in line with what they do. Much as it is a small sample response, it can give us an idea of what is happening in city and municipality archives in regard to authenticity practices. The results are given below under four major sub themes namely;

i. Background information on Noark records

ii. Record integrity measures upon receipt of new databases

iii. Electronic storage of records/databases within archives

iv. Electronic record keeping and international standards.

4.2 Background information on Noark records in city and municipality archives

It is important to highlight the fact that the research initially was interested in Noark 3 and 4 electronic records mainly but it turned out that they actually had referential databases and not full electronic records as such.

Possession of Noark Referential Databases in Archives

∑ = 5 Archives

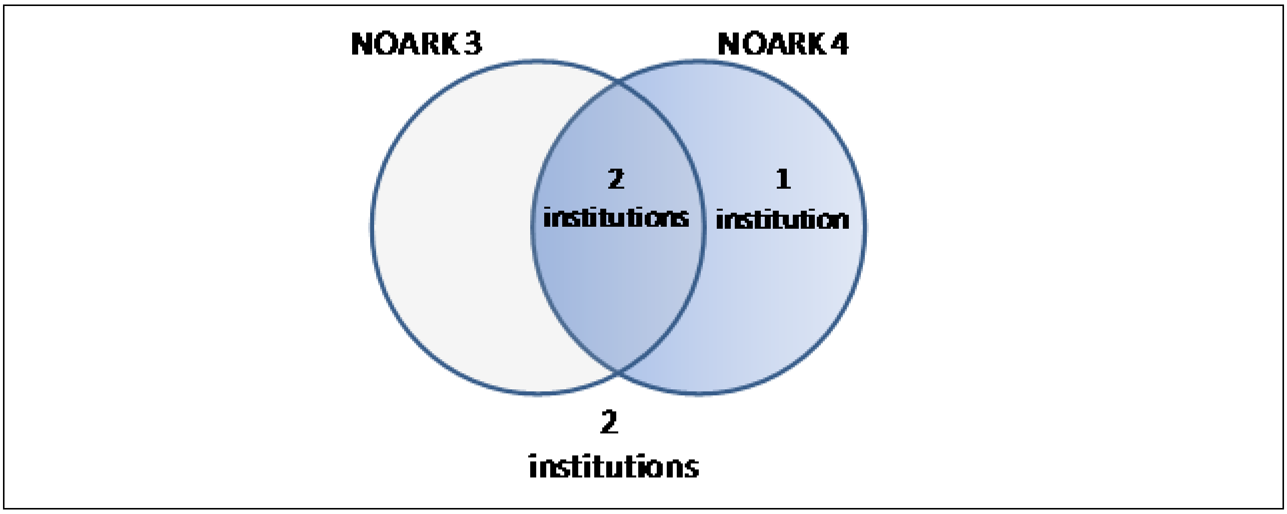


Figure 4.2.1 - Archives with Noark databases

2 institutions had neither Noark 3 nor Noark 4 records. 1 institution had only Noark 4 records and 2 institutions had both Noark 3 and Noark 4 records

As illustrated in Figure 4.2.1, two of the archives had both Noark 3 and 4 databases. One institution had Noark 4 and Koark, while the remaining 2 did not have any of these databases. One of the archives reported that it had in custody submission of Noark 3 and 4 in an improper format while one stated that it had only paper records. One archive also stated that it had only database tables of Noark 4. The improper file format gives an impression that the database could be erroneous. On average, all archives have approximately 10GB of data. This is quite low. However, the fact that a few of them have referential databases in mainly text and xml, this explains why the size is small.

Preservation processes

Two of the Archives had done migration; two had done refreshment while only one had done conversion. This is an interesting finding as it shows that conversion and migration is not an issue that will be dealt with in the future but is an issue the archives have to deal with now.

4.3 Record integrity measures upon receipt of new archival data

The above is summarized in the table below, (Table 4.3.1) based on responses received.

Question on Usage of:	Archive A	Archive B	Archive C	Archive D	Archive E
Integrity tools					
Checksums	√		√	√	
Policies			√		√
Chain of custody documentation			√		
Others		Privacy for an electronic archive. No electronic documents therefore no integrity protection.			
Verification tools					
Submission agreement	√	√	√	√	
Authenticity logs					√
Digital signatures					√
Reliability of chosen verification tools above (%)	90	50	75	25	50
Other suggestions towards improving authentication process	Loosen proprietary controls on Noark 4	Use of checksums for metadata and document files	Training in handling electronic records	No Response	Use of checksums with logs
Incomplete records handling					
Report error	√	√	√		√
Reject files			√		
Suspend processing			√	√	
Is the receipt process automated?	No	No	No	No	No
Would automation of receipt process be more reliable?	Yes	Yes	Yes	Yes	No

Table 4.3.1 *Survey Data response*

From table 4.3.1, it is clear that work practices are quite different across all archives. Some seem to be more ahead of others in managing electronic databases. This could be attributed to the fact that some have mainly paper records. For instance Archive B seems to be a paper archive mainly because it states so, and relies on submission agreements only. All Archives do agree that the receipt process is not automated but they would prefer an automated process for receiving archival electronic databases or records as a way of ascertaining trust and authenticity. Archive C has highlighted training as an important aspect in empowering them to do their work more optimally as far as electronic records and databases are concerned. Four archives use submission agreements as a verification tool. However the reliability of the submission agreement gives quite diverse reliability opinions. As seen above, Archive A seems to be satisfied with it at 90 % while Archive D is not at 25 %. This gives us a range of (90 -25) % = 65%.

This should not be seen as dismissal of submission agreements in general. Submission agreements in other settings, forms or contexts, perhaps even in other countries might be very successful, but on the basis of this survey the archives that responded show that there is no clear understanding on whether or not submission agreements are useful for verification purposes. In fact the individual archives might have totally separate approaches to the implementation of submission agreements. The results show a disparity and one reason for this is that agreements are not standardized and therefore vary case by case.

However to make a basis for this opinion based on only 5 archives suggests there is need for more research to find out how relevant the current submission agreement guidelines are to archives when doing their work.

4.4. Electronic Storage of Records/ Databases within Archives

Only Archive D is using persistent identifiers while Archive C is yet to decide on a particular filing system and the kind of persistent identifiers to use. This could probably be due to the fact that the electronic collection is not yet developed in all archives. The Archives receive various file formats including those that are not accepted by the National Archive. These range from MS word, text files, PDF files to proprietary databases as reported by two archives. It is the responsibility of the archives to convert them into an appropriate format as reported by Archive A. Only Archive A has intentions of converting files into the appropriate file format, for instance from PDF to PDF /A. Conversion tools used include Open Office batch processing and Adobe Pro as reported by two archives. Archive D does not intend to do any conversions while the remaining archives do not have files that need conversion. This again is probably due to the fact that they hardly receive electronic records and therefore conversion is not done on a large scale apart from a few databases that they work with. To ensure that trust is retained, the chain of custody and work flow documentation is used as reported by three archives. Checksums on both metadata and document files ensures quality control in case of conversions or receipt of database dumps as reported by Archive B. Contrary to other archives, Archive C stated that they do not have authenticity checks but rely on trust entrusted to depositors that the data holds integrity and is authentic upon deposit. However this is not always the case and this explains why they reject some deposits or report errors. For more reliability to ensure trust and authenticity, Archive C recommends training for depositing institutions in appropriate delivery formats and archival staff in extraction and appropriate handling of deposited electronic data.

4.5 Electronic record keeping and international standards

None of the archives have subscriptions to or use international registries apart from Archive C that uses PRONOM and other tools for identification of files. Only two Archives C and E feel the OAIS model reflects their work. This is further seen by one of the archives stating that, they use the National Archives regulations, chapter VIII reflecting OAIS model. Again this fact is well reflected in the ABM skrift 43 handbook on methods for digital long term storage in the municipal sector (minnehåndtering: metode for digital langtidslagring i kommunal sector). However, it is quite interesting that this handbook was released in 2007 as a guide for municipality archives but three years later, it is clear that some archives have not yet embraced this guide. Could it be due to the fact that they are not handling electronic documents? More investigation is required to find out why certain recommended practices are embraced by some archives but not all. One of the Archives stated that they do not have electronic records apart from the DIAS (Digital Archive Package Structure) project which is working on a repository management system to handle file packages and their authenticity. In fact, it was further reported that very few archives have electronic documents including the national archives. On trying to get more evidence on authenticity practices in archives, the researcher tried to reach the national archives, unfortunately, they were too busy to attend to this inquiry. Therefore future research could also include an investigation into authenticity practices at the National Archives of Norway.

4.6 Interview data response

Six key persons from six archives were interviewed, two of which were face to face interviews, while the other four were email interviews. The email interviews were used because it was difficult to engage them in a face to face interview. Below are the responses on various aspects.

4.6.1 Current state of Noark 3 and 4 records

All the archivists confirmed that Noark 3 and 4 are journal records in paper form. They do not exist as electronic documents in the archives but rather as reference records. These references contain mainly the source, the year and brief subject heading about the content of a particular paper record. The Noark data is not that much as stated by one archivist, "we have about 120 deposits, of which 90% are from FAG[12] system which is not approved for full electronic preservation." Therefore this leaves only 10 % of Noark referential databases. The journals are still at the authoring institutions and archives take over ownership after 25 years.

Noark 3 is in a text file format with identification references while Noark 4 is in an xml format. For instance, this ID can be **1996 NOR1: public health**. The reference data is normally received as an oracle dump which is imported into oracle and a Noark 3 format is generated. The data deposit may also be a text file or any doc file. According to one archivist, "Noark 3 is quite an old standard and therefore lacking many additional fields". On the other hand, Noark 4 takes on all fields or tables thus making it more detailed. It is a relational database and all the relations are available. For clear visibility rather than reading the xml format, the data is imported to a PHP platform using Mysql. Some archives are promoting Mysql as a standard for deposits or they endeavor to convert various database platforms to a Mysql dump, which is converted to xml, and this facilitates the maintenance of data integrity.

However, the fact that Noark 4 is comprehensive, it becomes more of a disadvantage during conversions (for instance conversion of information structure into a relational database results in poor quality data as stated by one archivist). This is further aggravated by the fact that the files are quite big, with each database holding 95 tables and therefore more prone to distortion in case of conversion errors with far reaching consequences as far as integrity is concerned. The only records that are in electronic form are the FAG records within the FAG system. The FAG system incorporates specialized databases in various fields with in the social sector for streamlining administrative procedures. For instance the OSCA database is full of text and provides information on health treatment and support. Other systems may be for maps, agriculture, just to mention but a few. However, it is quite interesting to discover from one archivist that the Noark standard is meant "to account for both full electronic archive formation as well as preservation unlike the FAG system". The Noark 4 standard further confirms this as

[12] A FAG system is a Norwegian records management system used by public administration, mainly for a particular subject area. For instance, health.

stipulated in chapter 12, with details on remote storage and transfer to archival repository. So the challenge here is why the Noark system is not effectively doing the authenticity management in archives, a role it is meant to achieve primarily as revealed in this study.

4.6.2 Maintenance of Trust

The data is preserved in xml or text format as explained above and then put on CD-R or DVD, making two copies of each data set. Since Noark 4 is the current standard, data is received in a Noark 4 format (xml) from Noark 4 systems. Authenticity at Noark 4 level may be based on the provenance information accompanying a database, though it cannot be used as evidence in legal terms. As stated by one archivist "we have no procedures or technical solutions that reveal whether data has been modified". Therefore, authenticity is not keenly followed apart from correcting errors where possible and in cases where the errors cannot be corrected, the data is retained in the former state. In fact, one of the archivists stated that, "we do not change the data that we receive, the depositors have never complained about the data that they deposit at the archive." However, this is a bit contrary to another archivist who said that they correct errors if they are able to, and give the depositors a copy of the corrected database. The archives always keep the original deposit, make changes if needed and give the authoring institutions a copy of what they have in the archive. Never the less, it seems that the authoring institutions are happy with the corrected databases and this explains why they have never complained. In trying to ensure trust, the archives face the following challenges.

Trust challenges

A possible change to stored data can be discovered through manual routine checks that are every two to five years, after which more backup copies are made. This is labor intensive and time consuming at the same time as stated by one archivist. For some records in the FAG system, trust is enhanced through the conversion of documents to PDF/A, accompanied by checksums. However this is not quite accurate at times as reported by one archive since dates change on checksums as a result of computer date changes, thus creating incorrect checksums yet the object versions do not change, thus creating false data and uncertainty. It is also important to note that "the Noark standard is capable of full electronic records management as well as preservation unlike the FAG system which is designed to increase efficiency in the daily work of municipalities", as stated by one archivist. The FAG system it is assumed to be doing the preservation role as well by the public administration according to one archivist. This misconception could explain why municipalities are comfortable with depositing the referential databases at the moment, assuming that preservation is well taken care of in the FAG system and therefore ignoring the deposit of full Noark records at the archive since FAG is meeting their needs in their daily work. At the same time, they have full text Noark records with them. The 25 years legislation for archives taking over ownership of records further promotes such work practices as municipalities are not obliged to deposit these records at the archives before 5 years, yet archives believe that the record creators are not in position to preserve and ensure their integrity and authenticity over such a long period of time.

Therefore, to ensure integrity of the received data, original databases are preferred as stated by one archivist so that comparisons between the original and derived databases can be compared and contrasted when doing integrity checks.

Noark 3 in particular has been reported to have many errors. For instance, some reference numbers are missing and therefore archivists are not certain about which documents are restricted for public viewing in future. To ensure privacy of these records, more journals are likely to be restricted after 25 years for public access.

One archive reported that Noark 5, the most recent version which is supposed to be better than all earlier versions still leaves archivists in a dilemma. Data is classified as an "ARK DEL" (which might be according to organization or time period) and mapped for referential integrity purposes. However, if a particular classification happens to go missing in a table, it means that all records under that classification cannot be retrieved because of the hierarchical aspect in classification. The National archives are trying to get a solution to this problem. With this problem, one of the archivists emphasizes the need of having an original database and then generate Noark 5 and other Noark formats instead of receiving them as Noark referential data at the archive.

The inefficient repository management environment at the moment has been cited as a big challenge at the moment. This is clearly revealed by one archivist saying that, "we need a storage facility that has the necessary capacity and security". Most processes are manually driven and therefore time consuming yet prone to more errors at the same time especially with regards to information integrity and security.

4.6.3 Further exploration of authenticity management

a) Identity management of records

Files are kept as batch files and not individual record references as reported by one archive. On arrival at the archive, a particular batch is given an identification number (ID) and is kept as a batch file in ASTA[13] database and stored on CDs. During retrieval, particular batch identification on CD or DVD is retrieved then particular keywords are searched from this database. Two of the archives declared a general absence of record identification due to poor documentation and repository management while another archivist says, "there has not been a need for this". This could be attributed to the fact that they are not handling full text electronic documents at the moment. In addition, not many data extractions have been done from database deposits probably due to the current state of repository management.

b) Authenticity practice

Since archives are receiving referential databases of Noark records, authenticity has not been a priority as declared by two archivists. One of them said that "not much authenticity check is done on deposits because the work has not come very far". Another archivist declared the general absence of authenticity solutions. However, some of the archivists have come up with their own tools to deal with error checking and this is the integrity check they are doing at the moment. One of them is the Universal Relations Database (URD), a tool that can read XML and shows information of the deposited databases. It is PHP based and tests everything that the archivist is interested in. The National Archives of Norway have also come up with a tool known as ARK4; it is based on pearl and runs on a Linux platform. Checksums are frequently used and some archives use the hash check shell extension to flag errors http://code.kliu.org/hashcheck/. However all the above tools are manually driven thus rendering archives inefficient in their preservation efforts since timely preservation tasks cannot be carried out to ensure integrity protection over time.

[13] ASTA is an archival information system developed in Norway.

c) Chain of custody practice and workflows

None of the archives was able to provide the researcher with any documentation on their work flow or chain of custody practice. This is because Norway does not use this term in archival repositories as it was explained by one of the archivists. That is “provenance is maintained in paper records which are able to provide the documents history from formation to present day” according to one archivist. The workflows are different and not standardized in archives since the Norwegian archives lack a uniform definition for an Archival Information Package, its contents and management. In practice, AIPs vary across archives as revealed by two archivists. The National Archives DTD[14] metadata scheme has elements to support chain of custody documentation but it is not used when handling Noark extracts at the submission and ingest stages since there is no standard in workflows, thus making it dysfunctional, according to one archivist. Consequently, this has led to the need to define a Norwegian AIP through the Digital Archive Package Structure (DIAS) project. It began its duties in May 2010 in conjunction with the National Archives of Norway and it is expected to handle authenticity and integrity aspects as well.

d) Storage facilities

The inadequate storage facilities further hinder the proper management of AIPs and chain of custody documentation since they lack repository management systems according to one archive. Consequently this affects integrity and authenticity management. However, some archives have already realized this and are therefore planning to use repository management and preservation soft ware like Fedora[15] or Dspace[16] to address integrity issues. These will enhance preservation aspects which can further be enriched with automatic error checks incorporated within for timely interventions when required.

[14] DTD (Document Type Definition) is meant to define the document structure and the legal building blocks of an xml document.

[15] Fedora (Flexible Extensible Digital Object Repository Architecture) for repository and preservation management.

[16] Dspace is repository management software with preservation abilities.

e) The use of submission agreements

A submission agreement formalizes the relationship between the file creator and custodian, according to one archivist. However, these are not very structured agreements as stated by two archives. They mainly contain the name of the institution depositing the data, the system (probably Noark version, 3 or 4) the number of records, and the total number of CDs, DVDs or Memory sticks deposited. The agreement also covers liability and access issues. Costs involved and implementation procedures could also be included as well. The receiving archive has the obligation of retaining the integrity of the deposited data. At the moment some archives seem to be happy with this agreement since it was weighted at 90% as a verification tool by one archivist.

4.6.4 OAIS Model Usage

The OAIS model is followed by some archives since it has been given as a guideline in the ABM skrift 43 handbook on methods for digital long term storage in the municipal sector. One archive reported that they have established good features concerning ingest and preservation planning. Another archivist also agreed that OAIS model is reflected in their work to some extent and not fully since all work practices are disintegrated and manual with some packages missing like the Dissemination Information Package. According to this archivist "the OAIS environment requires an interaction between systems with automation functions of security to enhance integrity and authenticity over time". Storage was highlighted as inadequate for preservation and authenticity purposes. The dissemination information package was declared absent by three archives since they do not disseminate any Noark records at the moment. However other archives do not really follow it and one of the archivists said, "am not familiar with OAIS, though the Noark deposits follow the OAIS guidelines," the archivist further added that, "I have not used it in my work because there has been a lot of work to do." This could probably be due to the fact that they are not receiving full electronic records and therefore mainly concentrate on correcting errors from referential deposits. Basing on the above survey and interview results, the model (Figure 4.6.5) can highlight the scenario at the moment.

Overview of current records management in city and municipal archives

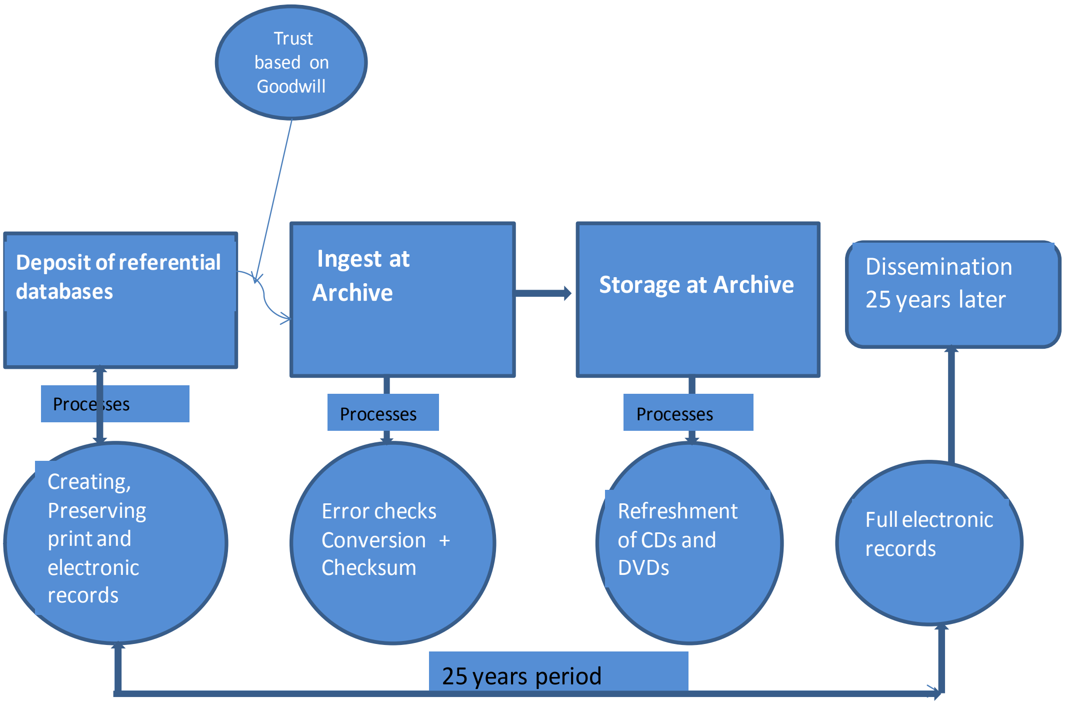

Trust based on Goodwill
Deposit of referential databases
Ingest at Archive
Storage at Archive
Dissemination 25 years later
Processes
Processes
Processes
Creating, Preserving print and electronic records
Error checks Conversion + Checksum
Refreshment of CDs and DVDs
Full electronic records
25 years period

Figure 4.6.5 - Derived model from data collected

4. 7 Deductions from Survey and Interviews

Only two archives willingly participated in both the survey and interview. The total participating archives from the survey and interviews are 9 out of the 17 archives. Basing on the above collected data, the following has been deduced.

a) Authenticity state of Noark records in archives

Looking at authenticity from the two key aspects, identity and integrity, it is clear that the identification system still has some weaknesses. That is, identity is taken at a batch level and not individual record level. This is not necessarily the fault of the archives as such but the authoring institutions. From this study, it seems that retrieving a particular record is not very practical. Therefore, there is need to streamline the identification of individual records. On the other hand, integrity is not a big problem at the moment since records are still in print form and therefore are less prone to alteration or changes. So the archives will eventually have the original records with integrity but effective retrieval and access will be a problem since the deposited referential databases are often with errors which the archives are not in position to highlight at times. This is because archives receive deposits on the trust assumption for authenticity and preservation management in general.

b) The current archival possession time frame is a setback

The current time frame of 25 years in which archives can possess the public records fully is not realistic in today's world of drastically changing technology. In just a decade, a lot changes in hardware and software and therefore if the depositing institution had referential errors in its deposit, those errors will never be rectified after such a time frame. In addition, the authoring institutions have not displayed best practices in records management. This is evidenced by the fact that archives always have to deal with errors and rejecting some referential databases because of poor records management. For instance Noark 3 that covers records from 1994 has errors that cannot be rectified in 2010, and these are just sixteen years old. What will happen in 25 years when archives take over ownership of these records? It may be impossible to rectify these errors.

c) The trust question at the submission stage

The study reveals that, the submission stage where depositors bring in their deposits at the archive leave archivists in a dilemma in that, they do not have a choice but to trust even the untrustworthy databases. That is, as long as a database can be accessed, it is unlikely that archivists can point out referential errors since they do not have access to the original databases. A preservation and authenticity aware environment should be created between the authoring institutions and archives.

d) Archival practices vary greatly

From the above results, it is clear that work practices across archives vary greatly because all archives seem to be at various stages. Some seem to be quite more advanced than the others mainly based on the amount of electronic documents or databases in their possession. This is further seen with OAIS practices. Some seem to be following the guidelines in ABM handbook on methods for digital long-term storage in the municipal sector which describes basic principles to follow, including OAIS while others are not. It is also interesting to note that these guidelines were produced in 2007 but nearly three years later, many archives are not following them. Others could be following the guidelines without actually being aware that they are doing so. For instance from the questionnaire response, some archives denied using OAIS model at all but they all listed some of the OAIS principles of reference, provenance and context. However, they could be other models that have similar principles though they were not defined by the respondents.

e) Current archival records management is inadequate

A number of weaknesses have been identified in this study right from acquisition of databases by archives to preservation planning and storage. For instance, the manual error detection hinders timely authenticity and preservation interventions at large thus exposing archives to failure to achieve their main mandate of preservation for posterity. This quotation by one archivist summarizes these observations, "the survey has revealed to us that we lack methods, tools and practices concerning integrity and authenticity protection". We do not meet TRAC because we have not yet established trusted repositories".

CHAPTER 5- RECOMMENDATIONS

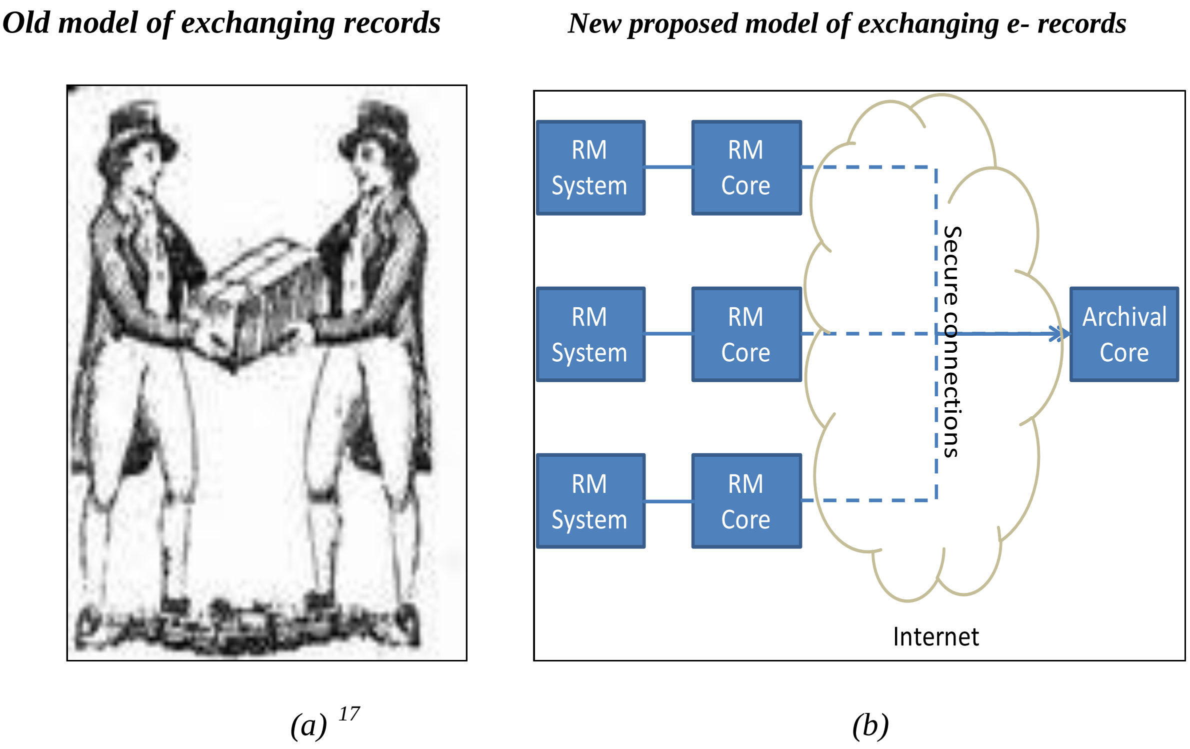
This chapter provides recommendations based on the survey and interview results. The recommendations are also based on authenticity check list points on the TRAC check list and the International records management standard ISO15489. It also uses the major authenticity principles of identity and integrity as a bench mark.

5.1 Archives should have full electronic documents sooner than later

At the moment archives mainly receive referential databases and therefore they do not have full control over the identity of records. The identity and integrity currently lies with the authoring public institutions, and there is no guarantee that these institutions are able to ensure long term preservation and authenticity in the 20 to 25 year time period that they store electronic documents. As revealed in the study, archives at the moment do not have individual file identification of records but rather, they have identification by batches which again are not very authentic since it is unclear what the contents of the original databases are. For instance, what implicit functionality is embedded within the application that is lost when it comes to a database extraction? It is well known that various databases have differing support for stored procedures. Do the extractions cater for this? This is very unclear. It is apparent that the authoring institutions are mainly concerned with their daily business and not the preservation of electronic information. Perhaps this is a reflection of the fact that records management is well regulated but the preservation aspect has not been as well defined. As a consequence we believe city and municipal archives should have a provision where they can access full electronic records and take care of them. For instance, every two to five years, public institutions should deposit their records to the archives since their primary mandate is long term storage and retaining authenticity of the records.

5.2 Need for answering the trust and authenticity question at the submission stage

The current method of depositing CDs, DVDs or memory sticks do not guarantee that all that is meant to be deposited is really deposited. As mentioned earlier, the archives accept what is given to them as long as it is accessible on their computers. Therefore since they do not have access to the original content, even when some records are missing, they will not know what is missing. Therefore automation of the submission process between authoring institutions and archives will go along way if particular standards are well prescribed in the system with secure access and authentication measures. That is, the system should have authorization facilities so that access transactions are well captured and can be tracked. This data later becomes part of the provenance data, and therefore providing evidence when needed at any one time. Another consideration can be to have a database mirror function or record deposit of complete semi active records from authoring institutions to archives. This will empower archives to manage preservation of such records in a more efficient and timely manner. This proposal is illustrated in the model below.



Old model of exchanging records

New proposed model of exchanging e- records

(a) [17]

(b)

Figure 5.2.1()   Exchange of records (old and new models)

The traditional model as shown in Figure 5.2.1 (a) is still reflected in today's archives where documents are collected and passed out to an archival institution. Figure 5.2.1 (b) shows a potential solution that is more in line with the technological advances we have seen by the general IT industry. This may solve some of the conversion errors and insecurities prevailing at the moment at the submission stage. This is because; the new proposed model provides access to original Noark record systems.

The main problem with the traditional model as it is practiced today is that it first requires an export of the data from an existing records management system to a file structure (xml representation of the original tables). This data is then imported into the archive institution and

[17] From www.freeelpaso.com/images/barter.png Image assumed to be in the public domain

processed into a new format for long term preservation. There is potential for loss of meaning when data is exported to the temporary format. There is also a potential that the records can be altered at this stage. When importing the records into the new archival toolkit, there is also a potential loss of meaning or records. The archives are following a traditional model that has too many points of weakness. The Noark 5 standard does to some extent help this situation but it still follows the traditional model.

The abstraction prevalent in the Noark 5 standard, namely that of a "core" where the data that is to be exported to the archive is stored, opens up for the possibility to apply the exchange of documents from authoring institution to archives in a more trusted manner. By increasing trust, we increase authenticity. The model shown in Figure 5.2.1 (b) is a theoretical solution to this problem. A Noark 5 core contains the data the archives want. Creating a submission from a core is a relatively straight forward job. What is required at the archival institution is a scaled up super core that is securely connected to the original core in the authoring institution using a public/private key security mechanism. When a record is assigned a "finished" status it is automatically extracted from the authoring core and moved to the archive core. It leaves an empty metadata shell in the authoring core to indicate that the record is still available, just stored offsite. If the authoring institution requires access to the records, it is possible; it is copied from the archives core back to the authoring core. The archives core retains a copy of the original record. Any changes to the record in the authoring core get recorded. In essence this strategy extends the concept of trust from the archival institution out to the records management system.

This solution answers two aspects that we believe are a hindrance to authenticity, time and export/import. The combination of 5 + 20 years leaves ample time for problems to arise and is dealt with by the archival institution taking control of a record as soon as its status is marked "finished" and transferred to the archive. Any problems with quality of the records can be dealt with at the time the record has been finalized, rather than waiting 5 years before a submission of a large number of records to an institution. The export/import problem falls away as the system is implemented as a large distrusted archive system.

5.3 There is need to standardize work practices across all archives

Currently, all archives have their own work practices depending on their resources and expertise which is a setback in preservation and authenticity management. As revealed from the study, work tools and practices vary across archives as revealed in section (a) and (b) of the data analysis, and the ABM skrift handbook. As such, some archives could have better tools than the others and therefore creating varying quality databanks. In addition, the work tools are very much manually driven. For instance, if an archivist does not check particular CDs or DVDs for errors on a regular basis, the data can be lost if problems with the media arise or readers of the media are no longer available for purchase. Repository databases with automated file format identification errors should be used and embedded in these repositories. For instance DROID and PRONOM software tools are platform independent and can therefore be incorporated in most digital repository architectures.

In addition to the above, the archival institutions should come to a common understanding of OAIS model usage. As revealed from the study, some are using it but denying it, while others declare outright that they do not use it. This is quite confusing because the ABM skrift handbook for archives (2007) uses this model, the digital and authenticity report by the National Archives (2010) recommends this standard as well.

5.4 Current position of city and municipal archives in relation to international standards

Based on the TRAC standard and looking at authenticity section of TRAC standard, I can rightfully assert that archives are currently fulfilling the following;

i) They have mechanisms for authenticating the source of all materials as revealed in the submission agreement and therefore implementing the following;

B1.3 – Repository has mechanisms to authenticate the source of all materials.

ii) The archives do try to verify the SIPs for completeness to a certain extent. There is still need to capture and verify complete SIPs as full electronic records, at the moment only referential databases are received creating incomplete SIPs. This is based on : *B1.4 -Repository's ingest process verifies each submitted object (i.e., SIP) for completeness*

iii) The archives implement B1.6 (as defined below) when they agree, reject or suspend some incomplete deposits or deposits with vast errors.

B1.6 - Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.

However the following TRAC authenticity aspects are missing in archives as revealed in the study results.

B1.8 -Repository has contemporaneous records of actions and administration processes that are relevant to preservation. (This cannot be fully confirmed due to the shortcomings described in data analysis, especially the fact that documents are co-stored with authoring institutions for over 20 years).

B2.5 - Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects. (The survey and interviews confirmed absence of this).

B2.7 - Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).

The survey and interview results confirm the absence of B2.7 above since the real original digital objects do not exist in the archives. Only one archive has used PRONOM so far.

B4.4 -Repository actively monitors integrity of archival objects (i.e., AIPs).

Again based on the fact that complete digital objects or electronic records do not exist in archives the AIPs are incomplete or do not exist at all, that is why there are no electronic DIPs in archives as illustrated in Figure 4.6.5. Therefore B4.4 is not being implemented at the moment.

In addition to the above missing TRAC aspects, the Norwegian record keeping system in archives is missing the following records management principles as outlined in ISO 15489, section 7.1.

The city archives have not assessed the risks that would be entailed in case of failure to have authoritative records of activity as defined in section 7.1 (f). This is backed up by the fact that no study before this research has considered this risk and the research results did not realize such measures on the ground.

The security of the records is questionable at the moment as demanded in section 7.1(i) of ISO 15489 since the security measures at the authoring institutions are unknown.

5.5 There is need to have a proper definition of AIP

This is in relation to the model standard that should be used. As revealed in 5.3 above, archives have varying opinions on OAIS model and work practices in general. In fact one archivist confirmed this further by stating that "there is no uniform definition of AIP and therefore each archive has its own AIP composition". This has consequently led to the formation of the DIAS project to develop an AIP standard for government and repository institutions". This is very much in the right direction for sorting out the inconsistencies prevailing at the moment and hence facilitating authenticity and preservation management at large.

5.6 Conclusion

The current authenticity practices are manually driven- and authenticity currently only covers integrity at the authoring institutions to a great extent since they have paper records. The unique identity element is currently missing within the archives. Therefore there is need to embrace all the authenticity elements of identity, integrity and following international standards like TRAC and ISO15489 as a backbone in authentic records and preservation management.

5.7 Future research

The National Archives of Norway should spearhead the research on authenticity management in their own archives and pass on best practices to the city and municipality archives. This is because it has a traditional and national role and has existed much longer with better financing when compared to the city and municipality archives. A lot of the municipality archives are 10 years or less in age and experience.

Further research should also consider exploring why some archives do embrace certain guidelines while others ignore them. The information flow for implementation of best or recommended work practices should be investigated in city and municipal archives. This is because all archives work differently yet they have similar guidelines as discussed in chapters 4 and 5.

Archives have spent a lot of resources looking at the import problems and not that much on how to deal with their collections over time. It is clear that automatic error detection software on records will become a requirement as the software data collections grow in size. This kind of software will go a long way in identifying errors in records and their structure over time. The research should look at how such tools can be developed so that they can be incorporated into any repository architecture that archives have chosen to use and can be integrated with the potential solution described above to increase authenticity by extending an archives authenticity mechanism to records management system. Future research should look at how we can create practices with regards to the transfer of records and authenticity that are suitable to the Information age of the 21st century.

REFERENCES

Archives System To All. (n.d.). Stiftelsen Asta. Retrieved May 26, 2010, from http://www.stiftelsen-asta.no/ASTA5/asta_5.htm

Becker, C. (2010). Preservation_planning with plato. Retrieved May 14, 2010, from http://www.planets-project.eu/training-materials/6-becker-preservation_planning/

Billeness, C. (2007). *The Preservation Action Cycle: introduction to Planets*.. Retrieved from http://www.planets-project.eu/training-materials/2-billenness-planets_risk_management/

Briguglio, L. (2008). The Casper answers to the digital preservation issues in DPE: registry of training materials. Retrieved April 21, 2010, from http://www.digitalpreservationeurope.eu/registries/materials/?topic%5B%5D=24

Bussel, S. (2007). *How_to_preserve*. Retrieved from http://www.planets-project.eu/training-materials/3-van-bussel-how_to_preserve/

Cultural Artistic and Scientific Knowledge for Preservation Access and Retrieval Architecture Team: (2009). Caspar overall architecture components and interfaces. CASPER. Retrieved from http://www.casparpreserves.eu/Members/cclrc/Deliverables/caspar-overall-architecture-components-and-interfaces

Center for Research libraries. (2007). Center for Research Libraries - Ten Principles. Retrieved May 13, 2010, from http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re

Commonwealth of Learning (Canada). (2004). In *PREST: Practitioner Research and Evaluation Skills Training in open and distance learning*. Vancouver: Commonwealth of Learning.

Consultative Committee for Space Data Systems. (2002). In *Reference model for an Open Archival Information System (OAIS)*. Washington, D.C: CCSDS Secretariat.

Creswell, J. W. (2008). In *Educational research: Planning, conducting, and evaluating* *quantitative and qualitative research*. Upper Saddle River, N.J: Pearson/Merrill Prentice Hall.

Creswell, J. W., & Creswell, J. W. (2007). In *Qualitative inquiry & research design: Choosing among five approaches*. Thousand Oaks: Sage Publications.

Digital Repository Audit Method Based on Risk Assessment: About. (2008). Retrieved June 19, 2010, from http://www.repositoryaudit.eu/about/

Denscombe, M. (2007). The good research guide: for small scale social research projects. (3rd ed.). Berkshire: McGraw- Hill.

Dempsey, L., & Lavoie, B. F. (2004). *Thirteen ways of looking at--digital preservation*. (D-Lib magazine, 10.)Retrieved April 10, 2010 from http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

Digital Curation Centre Partners and DigitalPreservationEurope Partners. (2007). Digital Repository Audit Method Based on Risk Assessment. DRAMBORA. Retrieved from http://www.repositoryaudit.eu/download

Digital Preservation Coalition., OCLC., & Lavoie, B. F. (2004). In *The Open Archival Information System reference model: Introductory guide*. DPC technology watch series report, 04-01. London: Digital Preservation Coalition.

Digital Preservation Europe. (2006). DPE: registry of training materials. Retrieved April 22, 2010, from http://www.digitalpreservationeurope.eu/registries/materials/?topic%5B%5D=24

Duranti, L. (2005). In *The InterPARES Project: The long-term preservation of authentic electronic records : the findings of the InterPARES Project*. San Miniato (PI), Italia: Archilab.

Duranti, L. (2009). The trustworthiness of digital records. InterPARES. Retrieved from http://www.rinascimento-digitale.it/eventi/conference2009/slides14/duranti.pdf

Establishing Trust in a Chain of Preservation: The TRAC Checklist Applied to a Data Staging Repository (DataStar). (n.d.). . Retrieved March 12, 2010, from http://www.dlib.org/dlib/september09/steinhart/09steinhart.html

Factor, M., Henis, E., Naor, D., Cohen, S. R., & Reshef, P. (2009). Authenticity and provenance in long term digital preservation: modeling and implementation in preservation aware storage. IBM Research Lab. Retrieved from http://www.usenix.org/event/tapp09/tech/full_papers/factor/factor.pdf

Fedora Commons. (n.d.). About — Fedora Repository. Retrieved June 18, 2010, from http://www.fedora-commons.org/about/about

Global Digital Format Registry (GDFR) Information Site. (n.d.). . Retrieved April 22, 2010, from http://www.gdfr.info/index.html

Geser, G. (2002). Integrity and authenticity of digital cultural heritage objects. Thematic issue, 1. Salzburg: Salzburg Research Forschungsgesellschaft

Havard University Library, OCLC, . et. al. (2010). Global Format Digital Registry. In *GDFR*. Retrieved 20th March 2010, from http://www.gdfr.info/index.html

Hitchcook, S., Brody, T., Hey, J. M., & Carr, L. (2007). Preservation Metadata for Institutional Repositories: applying PREMIS. University of Southampton, UK. Retrieved from http://preserv.eprints.org/papers/presmeta/presmeta-paper.html

Interkommunalt arkiv i Vest-Agder. (n.d.). Interkommunalt arkiv i Vest-Agder IKS (IKAVA). Retrieved May 25, 2010, from http://www.ikava.no/

International Research on Permanent Authentic Records in Electronic Systems. (2001). Authenticity task force final report. InterPARES. Retrieved from http://www.interpares.org/documents/atf_draft_final_report.pdf

ISO 15489: Information and documentation - records management. (2001). Switzerland: International Standards Organization.

Jantz, R., & Giarlo, M. J. (n.d.). Digital Preservation: Architecture and Technology for Trusted Digital Repositories. Retrieved March 8, 2010, from http://www.dlib.org/dlib/june05/jantz/06jantz.html

Kanhabua, N. (n.d.). Extraction of Temporal Information from Documents. Norwegian University of Science and Technology. Retrieved from http://www.longrec.com/Intranet/ResearchResults/Articles/State-of-the-art%20Extraction%20of%20Temporal%20Info%20from%20Documents.pdf

LongRec (2009). NB Pilot: migration experiments and recommendations. LongRec. Retrieved from http://www.longrec.com/Intranet/ResearchResults/Case%20reports/2009_Longrec_NB-Pilot.pdf

LongRec (2007). *The National Library of Norway: challenges in search and retrieval*. Oslo: LongRec. Retrieved from http://www.longrec.com/Intranet/ResearchResults/Case%20reports/CaseReport_NB_FIND_Final.pdf

Longrec_Understand-SOTA_v1.01.doc. (n.d.). . Retrieved January 7, 2010, from http://www.longrec.com/Intranet/ResearchResults/StateOfTheArt/Longrec_Understand-SOTA_v1.01.doc

LongRec (2007). LongRec Case Study : repository records management. LongRec. Retrieved from http://www.longrec.com/Intranet/ResearchResults/Case%20reports/CaseReport_NB_READ_Final.pdf

Norwegian Archive Library and Museum authority. (2006). *Cultural heritage for all : on digitisation in the archive, library and museum sector.* Oslo: ABM- Utvikling.

National Archives, UK. *(2010).* The technical registry, *PRONOM.* In *The National Archives.*

Retrieved 20th March 2010, from http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

National Library of Australia. (2002). PADI - Persistent identifiers. Retrieved May 14, 2010, from http://www.nla.gov.au/padi/topics/36.html

Neto, S. (1997). Research Design & Mixed Methods. Retrieved May 4, 2010, from http://www.socialresearchmethods.net/tutorial/Sydenstricker/bolsa.html#Why%20Mixed

OCLC, & CRL. (2007). Trustworthy Repositories Audit and Certification: criteria and check list. OCLC. Retrieved from http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Pennock, M. (2006). Digital Preservation : Continued access to authentic digital assets. JISC. Retrieved from http://www.jisc-collections.ac.uk/media/documents/publications/digitalpreservationbp_rtf.rtf

PREMIS Working Group., OCLC., & RLG (Organization). (2008). In *PREMIS data dictionary for preservation metadata*. Washington, DC: Library of Congress.

Pickard, A. J. (2007). In *Research methods in information*. London: Facet.

Plano, C. V. L., & Creswell, J. W. (2007). In *Designing and conducting: Mixed methods research*. Thousand Oaks, Calif: SAGE Publications.

Richmond, L. (1999). Procedure for applying identifiers to documents. University of Glasgow. Retrieved from http://www.gla.ac.uk/infostrat/ERM/Docs/procref.htm

Riksarkivet. (2010). Digitalt og autentisk :Planlegging av ny depotløsning for Arkivverkets digitalt skapte arkivmateriale. Riksarkivet.

Riksarkivet. (1994). Noark 3: Standard system for EDB- Basert Journalføring i statsforvaltiningen. Retrieved April 26, 2010, from http://www.arkivverket.no/arkivverket/Offentlig-forvaltning/Noark/Tidligere-versjoner/Noark-3

Ross S. (2007). Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries, Keynote Address at the 11th European Conference on Digital Libraries (ECDL), Budapest (17 September 2007).

Samset, K. (2003). In *Project evaluation: Making investments succeed*. Trondheim: Tapir Academic.

SourceForge.net: droid. (2009). Retrieved June 9, 2010, from http://sourceforge.net/apps/mediawiki/droid/index.php?title=Main_Page

Silverman, D. (2010). In *Doing qualitative research: A practical handbook*. Los Angeles: SAGE.

The National Archives | PRONOM | Information Resources. (n.d.). *The National Archives*. Retrieved April 22, 2010, from http://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm

The National Archives of Norway (2000). Noark 4: Norwegian record keeping system. Retrieved April 26, 2010, from http://arkivverket.no/arkivverket/Offentlig-forvaltning/Noark/Noark-4/English-version

Yin, R. K. (2009). In *Case study research: Design and methods*. Los Angeles, Calif: Sage Publications.

APPENDIX 1 – Questionnaire (English and Norwegian)

AUTHENTICITY PRACTICES IN DIGITAL ARCHIVES OF NORWAY

Digital Records go through a number of changes over time due to technological changes like failures in hardware and software. This poses a challenge to information institutions like archives of ensuring that these records are readable and authentic. In one of the National Archive's reports "Digital og autentisk - Planlegging av ny depotløsning for Arkivverkets digitalt skapte arkivmateriale, 2010, a number of challenges are pointed out that need to be addressed. With this report in mind, we wish to ascertain the current practices amongst municipality and city archives. The research goals aim at:

- Exploring and Identifying present practices of maintaining authenticity and integrity of digital records within Archives.
- Identify best practices to uphold while at the same time recommend other possible practices that can be adopted in ensuring digital record integrity.

For clarity, ***authenticity refers to trust, worthiness or is concerned with ongoing control over a record including creation process and custody.***

This survey will approximately take 15 minutes of your time. Your cooperation is highly appreciated.

Thanking you,

Florence Mirembe
Digital Library Learning Student (Masters)
Oslo University College
E-mail: s153413@stud.hio.no
Telephone 46 27 27 18

Background of the state of Electronic Records in your Archive

1. What kind of electronic records does your Institution take care of?

    - o NOARK 3
    - o NOARK 4
    - o KOARK

a) For how long have you had custody of NOARK 3 records? (Please state the number of years or year in which that standard was brought under your care).

----

b) For how long have you had custody of NOARK 4 records? (Please state the number of years or year in which that standard was brought under your care).

____

c) Approximately, how many electronic records are taken care of by your Archive?

- o >100MB
- o >1GB
- o >10GB
- o >100GB
- o >1TB
- o >10TB
- o >Do not Know

2. Has your collection gone through any of the following processes?(Please select what applies in your case)

    - o Conversion
    - o Migration
    - o Refreshment of Media Storage
    - o None of the above

Record integrity measures upon receipt of new archival records

3. What tools do you use for information integrity measurements upon receipt of new records for storage/ custody within your institution?

   - ○ Use of Checksums
   - ○ Chain of custody documentation
   - ○ Policies
   - ○ Others(Please specify) ---

4. What mechanisms are in place to verify and validate the source of all materials?

   - ○ Submission /Deposit agreements
   - ○ Authentication logs
   - ○ File format validation
   - ○ Use of digital signatures
   - ○ Others (please specify) --

a) To what extent do you believe that your chosen tools above are reliable?

   - ○ 25%
   - ○ 50%
   - ○ 75%
   - ○ 90% and above
   - ○

b) What other methods/ mechanisms do you think can be deployed to improve on the current authentication process within your archive?

____

____

----

5. How do you handle incomplete records?

- o Reject them
- o Suspend processing until missing information is received
- o Report errors
- o Other (Please specify)---

a) Is the current process of receiving digital archival records automated?

- o Yes
- o No

b) If yes, is automation more reliable in maintaining Trust as opposed to the manual method of receiving records?

- o Yes
- o No

Electronic Storage of Records within your Institution

6. Does your repository have persistent unique identifiers (especially for repository managers) for all archived records or objects?

- o Yes
- o No

7. Have you had to accept files in formats that are different from the ones specified by the National Archive?

- o Yes
- o No

a) Please give examples of these file formats if you said Yes above.

----

8. If you have files in formats that are (today) not valid file formats, do

You intend to convert them?

o Yes

o No

a) What “invalid” file formats do you hold in custody at the moment?

____

b) In case you convert them, what mechanisms or tools to you intend to use to ensure that the integrity of these records is retained?

____

____

9. How is trust of records maintained in your Archive?

o Chain of custody documentation

o Work flow documentation

o Others (please specify) --

Electronic Record keeping and International standards/trends

10. Do you have subscription to any international registries like the following registries? Select those that apply to your repository.

o Global Digital Format Registry (GDFR)

o PRONAM – UK National Archives file format registry

o Others ---

o None

11. Do you believe your repository/archive reflects the Open Archival Information System (OAIS) model?

○ Yes

○ No

a) If yes, please suggest any OAIS principles that your archive implements.

----

12. What are the components of your Preservation Description Information (PDI)? Select those that apply.

    ○ Reference (find-ability of information)

    ○ Context (Relationship to other information in a particular environment)

    ○ Provenance (life history of content data since its creation)

    ○ Fixity (ensures that content information is not corrupted)

    ○ Representation Information

    ○ None of the above

13. Please give any other comment that could be helpful in making this research more meaningful?

____

Thank you for your time!

HVORDAN AUTENTISITET PRAKTISERES I DIGITALE ARKIVER I NORGE

Arkivdokumenter går gjennom en rekke endringer over tid som følge av teknologiske endringer som svikt i maskinvare og oppdatering av programvare. Dette skaper en utfordring for depot institusjoner for å sikre at arkivdokumentene sine er lesbare og autentiske. I Riksarkivets rapport "Digital og autentisk - Planlegging av ny depotløsning for Arkivverkets Digitalt skapte arkivmateriale, 2010, blir en rekke utfordringer påpekt. Med denne rapporten i tankene, ønsker vi å fastslå gjeldende praksis blant kommune og byarkivene. Forskningen ser på å:

- Utforske og identifisere om nåværende praksis opprettholder ektheten og integriteten til digital-arkivdokumenter
- Identifisere beste praksis for å opprettholde, mens samtidig anbefale andre mulige fremgangsmåter som kan bli vedtatt i å sikre digital posten integritet.

Med autentisitet så mener vi tillit og pågående kontroll av arkivdokumenter, fra de ble skapt, brukt og avlevert

Denne undersøkelsen vil ca ta 15 minutter av din tid og vi setter stor pris på din samarbeid.

På forhånd takk.

Florence Mirembe

Digital Library Learning Student (Masters)

Høgskolen i Oslo

E-post: s153413@stud.hio.no

Telefon 46 27 27 18

Bakgrunnen til arkivdokumentene i depot

1. Hvilken elektroniske arkivdokumenter har dere?

- KOARK
- NOARK 3
- NOARK 4

a) Hvor lenge er det siden dere mottok den første KOARK arkivdokumentet?

----

b) Hvor lenge er det siden dere mottok den første NOARK 3 arkivdokumentet?

----

c) Hvor lenge er det siden dere mottok den første NOARK 4 arkivdokumentet?

----

d) Hvor stor (ca.) er samlingen av digitale arkivdokumenter?

- > 100MB
- > 1GB
- > 10GB
- > 100GB
- > 1TB
- > 10TB
- Vet ikke

2. Har elektroniske arkivdokumenter i samlingen din gått gjennom noen av de følgende prosesser?

(Vennligst velg hva som gjelder i ditt tilfelle)

- Konvertering
- Migrasjon
- Forfriskning av Lagringsmedia
- Ingen av de ovennevnte

Vanlig integritettiltak ved mottak av ny arkivmateriale

3. Hvilke verktøy bruker du for å få informasjon måle integritet ved mottak av nye arkivdokumenter ved avlevering?

- Bruk av sjekksummer
- Kjede av varetekt dokumentasjon
- Retningslinjer
- Annet (Vennligst spesifiser) -- -- ------------

4. Hvilke mekanismer er på plass for å verifisere og validere kilden til alle mottatt arkivmateriale?

- Mottaksavtaler
- Autentisitet logger
- Filformat validering
- Bruk av digitale signaturer
- Annet (vennligst spesifiser) -- -- --------

a) I hvilken grad tror du at de verktøyene ovenfor er pålitelige til å sike autentisitet?

- 25%
- 50%
- 75%
- 90% og over

b) Hvilke andre metoder / mekanismer tror du kan bli brukt til å forbedre den gjeldende godkjenningsprosessen i depotet ditt?

----

5. Hvordan håndterer du ufullstendig arkivdokumenter?

- Avvise dem
- Suspenderer bearbeiding til manglende informasjon er mottatt
- Rapporter feil
- Annet (Vennligst spesifiser)-- -- --------

a) Er den nåværende prosessen med mottak av digitale arkivmateriale automatisert?

- Ja
- Nei

b) Hvis ja, mener du at en automatisert prosess er en mer pålitelig måte å opprettholde autentisiteten til arkivdokumenter enn manuelle metoder?

- Ja
- Nei

Elektronisk Lagring av Arkivdokumenter i din institusjon

6. Har depotet ditt vedvarende unike identifikatorer for alle arkivdokumenter eller objekter?

- Ja
- Nei

7. Har du vært nødt til å motta filer i formater som ikke var godkjent av Riksarkivet?

- Ja
- Nei

a) Vennligst gi eksempler på disse filformatene hvis du sa ja ovenfor.

----

8. Hvis dere har filer i formater som tidligere har vært godkjent av Riksarkivet men som i dag ikke er gyldige filformater for langtidslagring, har dere planer til å konvertere dem til godkjente filformater

- Ja
- Nei
- Har ingen slike filer

a) Kan du gi en eksempel på et slikt filformat?

b) Hvis du skal konvertere slike filer, hvilke mekanismer eller verktøy har du tenkt til å bruke for å sikre at integriteten til disse dokumentene beholdes?

9. Hvordan opprettholder dere tillit til arkivdokumentene deres?

- Spore hvem som har forvaltet dokumentet over tid (chain of custody)
- Arbeidsflyt dokumentasjon
- Annet (vennligst spesifiser) --- --------

Elektronisk Journalføring og internasjonale standarder / trender

10. Abonnerer dere på noen internasjonale registre som følgende ? Velg det som gjelder til depotet.

- Global Digital Format Registry (GDFR)
- PRONAM – UK National Archives file format registry
- Andre ---
- Ingen
-

11. Tror du depotet deres reflekterer Open Archives Information System (OAIS) modellen?

- Ja
- Nei

a) Hvis ja, kan du fortelle hvilken OAIS prinsipper depotet utfører

----

12. Hvilken er Bevarings Beskrivelse Informasjon (PDI) bruker dere? Velg det som gjelder.

- Referanse (Unike identifikatorer (både internt og eksternt)
- Kontekst (Hvordan informasjon er relatert til annet informasjon)
- Proveniens (opphav og livshistorien til arkivdokumenter)
- Stabilitets informasjon (beskyter innholdet fra udokumentert endring)
- Ingen av de ovennevnte

13. Har du noen andre kommentar som du synes kan være nyttige for denne forskningen?

Takk for din tid!

APPENDIX II –Questionnaire data response

Questions	Response 1	Response 2	Response 3	Response 4	Response 5
1. Kind of records	Noark 3 and 4	Noark 3 and 4	Noark 4 and Koark	Not received	Not received
2. Noark 3 custody	Improper format/not delivered	2002- only paper	Koark -2005	Not received	Not received
3. Noark 4 Custody	Improper format/not delivered	2005 - only paper	2008	Not received	Only database tables
4. Size GB	10 GB and more	10 GB	10 GB	N/R	10 GB
5.Process - Conversion Migration	Conversion and Migration	Refreshment	Migration and Refreshment	N/R	N\R
6. Integrity tools	Checksums	Other- privacy for paper doc.	checksums, Policies, chain custody doc.	Checksums	Guidelines
7. Verification tools	submission agreements	submission agreements	submission agreements	submission agreements	Authenticity logs, digital signatures
8. What extent - reliability %?	90	50	75	25	50
9. Other methods- reliability?	loosen proprietary controls-Nk4	Checksums	Training in handling electronic doc.	N/R	Checksums with logs
10. Incomplete records handling?	Report error	Report error	Report error, suspend processing, reject	Suspend processing	Report error
11. Receipt process, is it automated?	No	No	No	No	No
12. Would automation of receipt process be more reliable	Yes	yes	Yes	Yes	No
13. Persistent identifiers?	No	No	No	Yes	No
14. Do you accept file formats not applicable to National Library?	No	No	yes	Yes	No
15. Examples of such file formats above	N/A	No such files	Proprietary databases, not elec.rec	MS Word(.doc) RTF	We get archival material from databases and convert them

Questions	Response 1	Response 2	Response 3	Response 4	Response 5
16. For invalid file formats; do you have plans to convert them?	Yes	N/A	No such files	No such files	No
17. Invalid format conversion examples	PDF to PDF /A	N/A	N/A	N/A	N/R
18. In case of conversion-tools used	Open office batch processing and Adobe	Checksums	N/R	N/A	Adobe Pro
19. How is Trust of records maintained?	N/R	N/R	Chain of custody, workflow doc.	Chain of custody, workflow doc.	Workflow documentation
20. Subscription to International registries?	No	No	No but use PRONOM and other sources for ident. of files	No	No
21. OAIS - Archive	No	No	Yes	No	Yes
22. OAIS Principles	N/A	National Archives reg. Chapt 8	We work with ABM font 43- memory management as a basis	N/A	conservation – planning, ingest and dissemination
23. PDI components	Context, Provenance	Reference, Provenance	Ref, context, provenance , fixity	Reference	Context and provenance
24. Any other comment	None	No electronic doc. DIAS project	Not decided on filing system and Perst. ID.	None	None

N/A – Not Applicable

N/R – No response

APPENDIX III– Interview guide (Norwegian and English)

INTERVIEW GUIDE ON AUTHENTICITY PRACTICES IN CITY AND MUNICIPALITY ARCHIVES OF NORWAY

Dear Respondent,

Thanks a lot for accepting to be a part of this study – Authenticity practices in digital archives of Norway. This study aims at ascertaining current authenticity practices in city and municipality archives and identify an approach that can best suit these archives as they endeavor to ensure that authenticity and integrity of their digital collection is retained over time.

This interview will take an hour or one and a half hours. Your archive will receive a copy of this final work.

Thanks a lot for your cooperation.

Yours sincerely,

Florence Mirembe

Digital Library Learning Student (Masters)

Oslo University College
E-mail: s153413@stud.hio.no
Telephone 46 27 27 18

1. Hvilke elektroniske datakilder lagrer dere? For eksempel, 3 NOARK, 4, relasjonsdatabaser (fra

fagsystemer).

a. Spørreundersøkelsen viser at de fleste datakildene var NOARK 3 avleveringer og uttrek

fra fagsystemer. Jeg vil gjerne vite litt mer om hvordan dere håndterer Noark 3

avleveringer og utrekk fra fagsystemer.

2. Hva er den normale prosedyren for å håndtere innkommende elektronisk arkivdokumenter?

a. Når det gjelder autentisitet. Hvordan ivaretar dere tillit / ektheten til innkommende

arkivdokumenter?

b. Hvilke personer er ansvarlige for disse arkivdokumenter? Er de interne eller eksterne

personer?

c. Hvordan forhindrer du eventuelle endringer i disse arkivdokumenter?

d. Hvordan oppdager dere endringer i disse arkivdokumenter (hvis endringer skulle

oppstå)?

e. Hvilke utfordringer står du overfor for å opprettholde ektheten av elektroniske

dokumenter?

3. Undersøkelsen har også avdekket at enkelte depot institusjoner har kun mellom 25% og 50%

tillit til mottaksavtaler. Har du en mening om hvorfor det er slik?

a. Hva er inkludert i en mottaksavtale? Vennligst gi noen korte detaljer.

4. Hvordan vedlikeholder dere identiteten til individuelle filer / elektroniske dokumenter over tid?

5. Hvordan håndterer dere ufullstendig arkivdokumenter?

6. Hvordan håndterer dere filer i formater som ikke er på Riksarkivet godkjent liste?

Når filer skal konverteres, hvordan sikrer dere at den opprinnelige budskapet i dokumentet

og/eller integriteten til dokumentet blir ivaretatt?

8. Mener du at institusjonen din gjenspeiler OAIS modellen? Hvis ja, hvordan?

a. Dersom institusjonen ikke følger OAIS modellen, hvilken modell følger dere?

9. Er det mulig for meg å få en kopi av arbeidsflyt dokumentasjon eller prosedyren som brukes til å

ivareta chain of custody ?

10. Eventuelle forslag eller kommentarer er velkomne.

1. What electronic data sources do you hold in your archive? For example, NOARK 3, 4, relational databases (FAG system).

   a) The recent survey indicates that NOARK 3 actual records are in paper form, much as a bibliographic database seems to be available. Please clarify more on this. What is the actual structure of NOARK 3?

2. Normally what is the procedure of handling incoming electronic records?

   a) A look at authenticity. How is trust/ authenticity maintained from when records arrive at your institution?

   b) Which persons are in charge of these records? Are they internal or external persons?

   c) How do you prevent any changes to these records?

   d) How do you detect any changes to these records in case they occur?

   e) What challenges do you face in maintaining authenticity of your electronic records?

3. The survey also revealed that some archives have a 25% to 50% trust in submission or receipt agreements. Why is this so or why do you have less trust in submission agreements?

a) What is included in a submission agreement? Please give some brief details.

4. How do you maintain Identity of individual files/ electronic records over time?

5. How do you handle incomplete records?

6. How do you handle file formats that are not on the list of National Archive?

7. When you do file conversions, how do you ensure that the meaning or integrity of these records is retained?

8. Do you think your archive reflects the OAIS model? If yes, how is this reflected?

b) If the archive is not using the OAIS model at all, what model are you using?

9. Is it possible for me to get a copy of your work flow documentation or chain of custody documentation procedure?

10. Any suggestions or comments are welcome.

TUSEN TAKK!