Analysis of Functional Requirements to Ensure Authenticity and Integrity of Archiving Norwegian Electronic Public Administration Records

Tadele Tedla Damessie

Oslo University College
Faculty of Journalism, Library and Information Science

Master's Thesis
International Master in Digital Library Learning

Supervisor:
Dr. Thomas Sødring

Co-Supervisor:
Dr. Ragnar Nordlie

June 27, 2011

Declaration

I certify that all materials in this thesis 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.

Signature: X Tadele Tedla

Jun 2011

Abstract

Authenticity and integrity are crucial elements of trust in physical or electronic document archiving. This thesis analyzed the functional requirements of authenticity and integrity and how to ensure them in the context of the Norwegian public administration records. NOARK is the Norwegian recordkeeping and archiving standard and Fedora Commons an open source archival repository software are used as a record management system and archival system respectively to establish the case of the study.

For the purpose of meeting the objectives of the study, standards, literatures and previous studies on the area of trusted recordkeeping and archiving are analyzed; on the basis of which an archival framework addressing authenticity, integrity and trusted chain custody is proposed and prototype is developed as a proof of concept. The validation is carried out by purposely compromising the authenticity and integrity of the electronic records in the process of transferring from NOARK to Fedora Commons and detecting the failure in either of authenticity or integrity or both before and after archiving the records.

The study found out that records archived using our framework have met the authenticity and integrity requirements of archival objects. Records archived using the proposed archival framework are found to improve the evidential value of records for court cases.

Acknowledgements

This thesis adds a very special value to the learnings that has been occurring to me for the past two years in the DILL program. The following are some of the persons contributed in a way that getting this opportunity to thank them is a pleasure and an honor at the same time.

Thomas Sødring and Ragnar Nordlie are my two supervisors. An indeed many thanks to Thomas Sødring who boosted my enthusiasm, confidence and believe during the whole process of writing this thesis. In addition to the comprehensible inputs in the form of comment feedback, the extracurricular support from Ragnar Nordlie makes him more than just a supervisor.

The DILL professors at Tallinn University, Oslo University College and Parma University together with the visiting scholars shared the top education over the period of two years. Merje Songe, Kersti Ahrén Heløe and Christin Mollenhauer gave administrative support. It is a pleasure to thank you.

Special thanks goes to Bendik Rugaas for the simple paper cutting in to two demonstration yielding infinitely a bigger circle paper every time instead of two smaller pieces. That is a big time lesson worth more than four lines of gratitude for me. Thank you indeed.

The Erasmus Mundus program funded the whole study. The learning experience and the thesis would not have been possible with out the Erasmus funding. My DILL 2009 classmates also deserve a line of gratefulness for the two years life we shared together. A heartfelt genuine thanks to you.

My long time friend Amanuel gave me several constructive comments as well as tips in using LaTeX specially for reference management which is one of the painful tasks in research writing. Thank you so much Aman.

The thesis could have turn out to be different without your support Rosy Rosy. I owe my deepest gratitude to you for rhyming with the thesis; and the Lord who see this before ages which of course my gratitude stays ever **bold** for leveling my way since I not know and see.

Thank You all!

Contents

This page is left blank deliberately

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the problem in which the study is grounded and why the problem is so important to be addressed. In addition, the objectives, the scope and the significance of the study are described.

1.1 Motivation

Milakovich and Gordon (2008) defines public administration as:

> Public administration may be defined as all processes, organizations, and individuals (the latter acting in official positions and roles) associated with carrying out laws and other rules adopted or issued by legislatures, executives, and courts(p. 11).

Records serving as active registers of these processes, organizations and individuals are public administration records. Policies, laws, legislations and other transactional records between the state and the citizens are some examples of public administration records. Public administration records document the event triggered the process, the whole process between the event and the final decision and the final decision to the event. The purpose of public administration records is to document activities , process and decision of the state and make them available to citizens. Archiving of the public administration records is one way of ensuring the availability of the records to current and future citizens.

Archiving plays a vital role in documenting historical developments of a nation. For instance in Europe archives are created to document property, in Australia to document locations of fresh water and in North-America to document movement of herds or relationship between deities and man (Hoeven, Albada, Información, & UNISIST, 1996). The archived documents

are created to witness accounts of their time for future users of the archive. World War I document archive[1] and World War II documents[2] are archives containing records about the two major wars of the world. The documents in these archives serve as an account or witness of what has happened in those times.

However; the documents stored in archives could be lost due to natural disasters like earthquake, flood, cyclones; carelessness, fire or war. A particular example is the contents in the library of Leuven[3] which are "reduced to ashes" as a result of war. The problems mentioned above are common to both digital and physical archives. In addition to the shared common problems, the information in digital archives are vulnerable to disappearance or lose due to technical or non technical failures (Waugh, Wilkinson, Hills, & Dell'oro, 2000). The technical failures includes deterioration of the media on which the information is stored or the software reading the information. Inability to preserve aspects of the information that makes the digital information useful; which include the information's status, its ownership, its reliability, its authenticity, and its retrievability fall under the non technical failures.

Despite the problems mentioned above in electronic archiving, a study finds out that users wished to have electronic versions of the most important documents for reasons other than trust(Hart & Liu, 2003). This shows the gap between electronic documents and trust. This raises a serious problem when it comes to using digital documents from the archives of the public administration records as a reference to address a particular public issue. In addition to this, the gap between trust and electronic documents will put the witness value of the archived documents in question.

Records are increasingly produced and archived electronically. The gap between electronic documents and trust puts the archived records in question when it comes it using them as a witness of their time. It is this necessity of addressing trust in electronic archives triggered the undertaking of the current study.

1.2 Statement of the Problem

Trust is defined as an action on the basis of a subjective personal judgment(Arne-Kristian Groven, 2008). Though trust is a subjective action, the process of arriving at the outcome of trusting a particular digital information can be

[1] `http://wwi.lib.byu.edu/`

[2] `http://www.paperlessarchives.com/world_war_ii.html`

[3] `http://www.kuleuven.be/about/history.html`

supplemented with constructive evidence. As Hart and Liu (2003) indicated in the study trust needs to be embraced in every process and component of digital preservation. With the view of preserving trust in digital information, the study considers authenticity and integrity as a means of preserving trust in every chain of the digital objects and records transaction.

Authenticity deals with verification of the accompanying claims associated with a digital object (Muir, 2001). The claims are usually the description of the digital object expressed in some form like metadata. For instance a digital object (D_o) claiming to be written in the year (Y') by author (A^{D_o}) is authentically acceptable unless either Y' is proved to be not the year the object is written or A^{D_o} is proved to be not the author of the object. In other words, authenticity concerns more on proving whether the object is what it purports to be or not. In a more realistic sense of currently existing commercial applications, there exists a third party (P^T) who manages these issues, and the trust we assign on the objects relies partly on the confidence we assign to the third party (P^T). Integrity on the other hand, deals with the actual content of the digital object. That is whether the object(D_o) is still the same today as it was first created in the year (Y') by the author (A^{D_o}). That means, the content of the object (D_o) remains unchanged intentionally or unintentionally over time or in transit(Lynch, 2000).

In archiving objects from records or producers of digital contents a particular care for addressing these trust requirements – authenticity and integrity- should be taken into account; otherwise a content or an object in the archive which failed to pass authenticity and integrity test will have no legal value or unacceptable to be used as evidence in a court case (Florence, 2010). If an object in the archive has no legal value or acceptance in its basic form, the effort and resources spent in preserving that content is as good as a capital waste. This implies that addressing these trust requirements at the basic stage of archival development is an absolute survival requirement of the archival objects. It is this basic survival necessity of archival trust which formulates the problem of this study. In addition, many of the researches undertaken in the field of archiving are chief theoretical as Waugh et al. (2000) pointed out; which as experience taught there are times of ultimate sacrifices while experiencing theoretical accuracies into practice[4]. So to address the above mentioned problems, this study considers authenticity and integrity of electronic archiving of digital objects having evidential value the prime focus of the study. To balance the theory with practice and to validate the proposed solution a prototype is developed.

[4] `http://en.wikipedia.org/wiki/Software_bug`

1.3 Objectives of the Study

The research aims to investigate the issues involved in the archival of trusted digital objects having evidential value to court cases. The study analyzed the functional requirements of authenticity and integrity for trusted archival of objects using the Norwegian archival standard – NOARK as a record management and Fedora Commons as an archival repository case. To address the objective, the following questions are asked:

- what are the functional requirements of authenticity for long term archiving of trusted objects?
- what are the functional requirements of integrity for long term archiving of trusted objects?
- what are the functional requirements of trusted chain of custody for archiving of trusted objects?
- how do we ensure authenticity and integrity during the process of transferring records to archive?
- how do we ensure authenticity and integrity after objects are being archived?

1.4 Scope of the Study

The study is about authenticity and integrity in archiving of digital objects. Specifically, the study focused on the functional requirements of trusted digital objects for submission into the archive from records. The process of maintaining trust in the record side specially in NOARK-5 is a complex process of tracking down records for twenty five years before the records passed onto the archive(MoReq, 2009). As the result the study focused more on the submission side of the problem i.e. once after records are marked finalized. In addition, studying users of the archive is also out of the scope of the study as the result the consumers side or the dissemination information package trust maintenance issues are also out of the scope of the study. The process of maintaining trusted objects with in the archive and the process of trusted submission including chain of custody handling is what this study focused on.

1.5 Significance of the Study

The study might have the following significances:

- it might help in raising awareness of archival bodies towards trusted archiving
- it might help in informing archival software developers on the process of maintaining trusted chain of custody and archiving of trusted digital objects
- it might also help in informing archival bodies on the process of maintaining integrity within the archive
- maintaining archives with evidential value for court cases is a delicate matter. One of the significance of the study might be imparting a method of achieving such to archive maintainers
- the output of the study might also be incorporated in the development of archival software
- it might also serve as a spring board to researchers along the same line of theme.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews literature conducted within the area of trust in records management and archiving of digital objects. Authenticity and integrity are the elements used to describe trust in this study. The chapter operationalizes trust and the elements authenticity and integrity; and further reviews the requirements in the maintenance of trusted records and archiving of objects from studies conducted within the area. In addition, the Norwegian e-signature act is summarized to examine the admissibility of digital documents in court cases. The Open Archive Information System(OAIS) and the mechanisms of handling trust requirements in the model concludes the chapter.

In addition to Google, ACM digital library, Emerald, IEEE Xplore and ScienceDirect online databases are used to gather literature for review. The search terms used contain the combination and variant of the words "trust", "authenticity", "integrity" "archiving", "digital" "object" "record". For example a typical search using "integrity + authenticity + digital + objects" on Google produces more than seventy six thousand results during the time of undertaking the study. In addition to the archival search terms, phrases like "Evidential weight and legal admissibility of electronic information" is also the common phrase used in the online databases to relate the evidential value of records and objects archived digitally.

2.2 Definition and Operationalization

This section defines the most common words related to trust and contextualizes their usage in this study. The words include trust, trustworthiness,

provenance and non-repudiation. The necessity of highlighting the related words is for the sake of inclusiveness of the literature review section with regards to concepts related to trust and also it helps to establish the scope of the study.

One of the basic shifts in the understanding of trust is the shift in viewing trust as a property to an assessment. The property view of trust implies that trust is something that can be built into the systems which users are supposed to rely on. However; the way people develop their trust to a system or an object or another person or organization is mostly based on their day-to-day interaction rather than the trust component built and incorporated within the system. This view of reality in trust development by researchers lead to the shift in the paradigm of of trust; that is, trust as property to trust as an assessment (Denning, 1993).

Viewing trust as an assessment retains the subjective nature of trust; which is based on an individual assessment of an object under investigation or declaration of others assessment about the object under investigation whom the investigator trusts (Denning, 1993). An individual may choose to trust a particular object to be written by the claimed author on the basis of personal assessment. An assessment can be carried out on the basis of previous personal experience or else certain expected behavioral attributes. For instance, we chose to trust amazon on the basis of previous personal transaction experience with amazon or trust a particular hand writing belonging to the claimed author examined by an organization which we trust. Ultimately an object is trusted if and only if users of the object trusts it which makes trust a subjective matter(Arne-Kristian Groven, 2008).

On the other hand, Lekkas (2003) identified four kinds of trust in a current society. Calculus-based Trust, Information-based Trust, Transitiveness-based trust and trust within a social system. Calculus-based trust is established between involved parties on the basis of possible risk analysis. The developed trust is a result of careful mathematical calculation. A typical example of calculus-based trust is business partnership where partners carefully analyze profit loss before committing trust to one another. This type of trust might seem an objective attempt to materialize trust; however; ultimately despite the result of the mathematical calculation, partners may still chose to trust and venture or not. This is what makes trust a complex matter to deal objectively. Information-based trust is an incremental relation resulting from continuous assessment and interaction overtime. As the result uncertainties decrease and the trust relationship develops further. This is trust based on interaction and intimacy. Transitiveness-based trust is trust developed through a trusted third party. The trust between participant entities is developed through a trusted third party, which the participant entities

trust. This kind of trust is the most common form of trust in the electronic world. For instance certificate authorities are trusted third party organizations where users of their services chose to trust both the organization and other users of the certificate authorities. Trust within the social system is a trust developed because of participation within a social system; as the result participants develop trust exclusive to the system. For instance trust developed based on membership is a typical example of this type of trust.

The context of trust in this study is the kind of trust which is developed through trusted third party which is the Transitiveness-based trust and its contextual understanding in this study is the one adopted from (Lekkas, 2003):

> The notion of trust in a TTP(Trusted Third Party) could be defined as the customer's certainty that the TTP is capable of providing the required services accurately and infallibly, a certainty which also expresses the customer's faith in its moral integrity, in the soundness of its operation, in the effectiveness of its security mechanisms, in its expertise and in its abidance by all regulations and laws, while at the same time, it also contains the acknowledgement of a minimum risk factor, by the relying party (p. 4).

On the basis of the above description, a trusted archival body is expected to meet users requirement in the maintenance of the services it provides; and assurances that the services it provides are intact either by introducing a trusted third party which producers and users of the archive trust and rely on or by becoming one of the trusted third parties meeting producers and users expectation itself. However; in this study, an archival body which works with an independent trusted third party is considered and modeled. The trusted public third party is the one ensures trust for users, producers and the archive trust and which complies with open set of standards facilitating sustainability and cooperation. In light of which, integrity and authenticity are considered as elements of trust that are particularly investigated in the study.

A similar but slightly different concept related to trust is trustworthiness. Unlike trust which is an action, trustworthiness is characteristic of objects to be trusted indicating the object is worthy of trusting(Toma, 2010). The development of trust can be facilitated by social interaction whereas judgment of trustworthiness can be carried out in the absence of social interaction or indirectly. To materialize the difference between trust and trustworthiness in the context of this study, trust comes from the archival body action. The archival body can choose to trust or not during archiving of a particular

object; however; trustworthiness comes from the characteristic of the object being archived. An object might exhibit trustworthiness but may not be trusted. However; other studies in the field of socio-economics indicated that the higher the degree of trustworthiness, the higher the trust(Tullberg, 2008).

Another concept related to trust is provenance. Provenance is information regarding the creation, the chain of stewardship, or the version a particular object or record has undergone including the context of use (Buneman, Chapman, & Cheney, 2006; Chapman, Jagadish, & Ramanan, 2008; Cheney, 2009). The provenance information generally comprise the version history of a record or an object mostly in various format. The common provenance information is kept in the form of metadata of a record or object. Provenance provides further information regarding the sequence of events that lead to the creation of the content, the context of the content creation and usage; which can further be used to address issues that might arise after the creation of the content. Trust is described as one of the motivations behind invoking provenance information (Cheney, Chong, Foster, Seltzer, & Vansummeren, 2009).

Another related but slightly different concept to trust is non-repudiation. Non-repudiation is a means of establishing proof that the producer of a particular content will not find a way to deny the fact of creating the content or the archive body will not deny the reception of a digital content or committing a certain action on a digital object; proof of origin, proof of integrity and verifying them using a third party are the mechanisms for establishing non-repudiation to an electronic document (Peiris, Soysa, & Palliyaguru, 2008). Proof of origin establishes an irrefutable relationship between the content and the originator where as proof of integrity ensures the content is not altered after creation which the third party uses to establish the non-repudiation of the electronic content.

On the basis of the above theoretical assumptions, the likelihood of a content to be trusted is facilitated if all historical information leading to the creation of the content and a mechanisms for not denying them is intact. Due to the policy of of archiving records actively used for twenty five years in NOARK, incorporating the provenance in the prototype from the record side needs a separate study by itself.

2.3 Evidential role: Authenticity and Integrity

This section describes authenticity and integrity as elements of trust and ways of maintaining them from a general perspective in the preservation of electronic contents. In addition, the evidential role authenticity and integrity

in legal cases is analyzed and reviewed. Authenticity as described in the first chapter, is the verification of the digital document's accompanying information in the form of metadata or the claims of the content; whereas integrity is about the actual content being intact or unmodified without proper recognition since origination or creation.

Evidence according to Gladney (2004, p. 407) is "information indelibly recorded by two or more people who are unlikely to have colluded in misrepresentation". The recorded information includes the actual content, the metadata and any other closely related information recorded to provide evidence that the content is trustworthy throughout the history of the object. Since integrity is considered to deal with the actual content and authenticity to the information regarding the content, their evidential role is significant. However; highlighting "evidential integrity" and "evidential authenticity" provision and maintaining mechanisms imparts a developmental role for the study.

Evidential integrity refers to integrity of digital content acceptable as evidence in court cases; so does evidential authenticity referring to metadata of the digital content having legal value(Irons, 2006). However; the legal side information requirement for admissibility of digital objects for evidence needs an in-depth analysis by itself so as to come up with a fully-fledged evidence requirement for electronic resources by legal institutions. This study, concentrates particularly on authenticity and integrity as a means of providing evidence to digital resources.

Creating a digital fingerprint of the electronic content is the common method of ensuring integrity. Hashing methods are used to generate a unique checksum or message digest out of the digital content which is further signed using the identity of the producer of the digital content. The algorithms ensure integrity in a way that even a single addition of a space or changing letter case results in a completely different message digest; which helps to easily detect modification of the electronic content. Furthermore, authenticity is information regarding the source of the content which can be expressed in standard format using XML markup languages. However; in this case, authenticity or source of the content is verified using the private identity used to sign the message digest of the digital content. The mechanics of integrity and authenticity verification is detailed in Figure 2.1. Keeping track of the chain of custody of the whole process of archiving including authenticity and integrity of the digital content is a crucial asset to the evidential value of the archival objects as it properly logs all the actions that are carried out on the object and by whom and when.

Use of public methods are the methods employed to assert authenticity, integrity and custodianship which are used to model the proposed framework

in the study. The public methods certify deposits of original objects, registration of unique identifiers, fingerprints, metadata or proofs in a publicly available way(Bearman & Trant, 1998). Details of authenticity and integrity mechanisms in NOARK-5 and Fedora Commons is discussed in the information modeling chapter of the study. In addition, the chapter also details the architecture of the proposed framework.

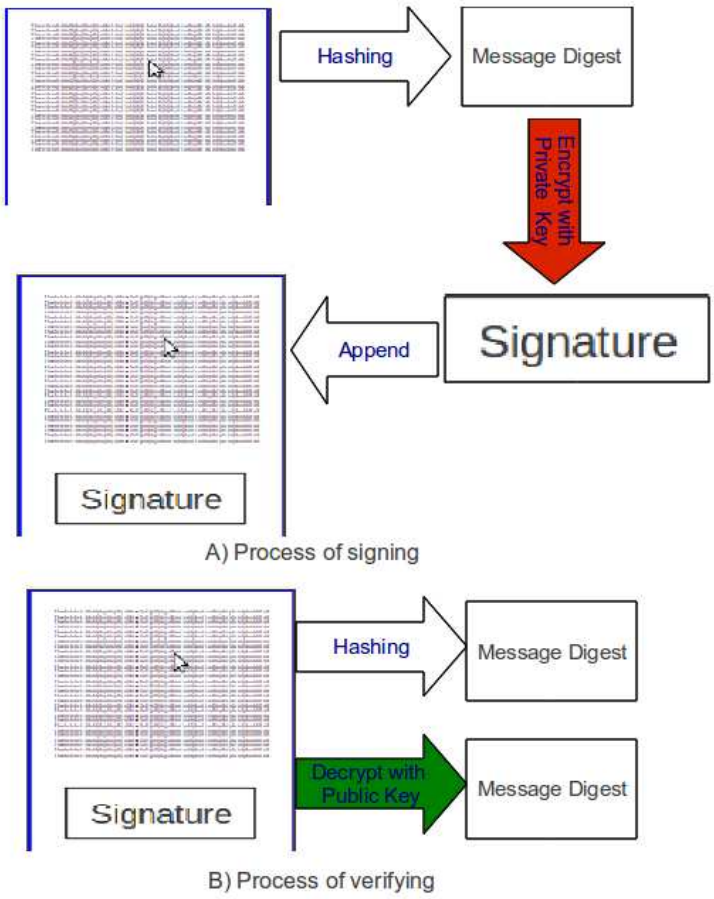

Figure 2.1: Process of Digitally Signing and verifying a Document

Figure 2.1 demonstrates a process of digitally signing and verifying a document. As shown in Figure 2.1 A, the signing process is carried out by applying a hashing algorithm on the document to be signed. The application of hashing algorithm on the document to be signed creates a message digest. The message digest is a one way compact fingerprint or unique representation of the digital document. It is one way because the algorithm makes it impossible to recover the digital document out of the generated message digest(Mel & Baker, 2001). The length of the message digest is determined

by the type of hashing algorithm used for instance MD5 is 128 bit and SHA-1 160 bits and so on.

Since the proposed solution uses a public key cryptographic method (see Chapter four for detail), the digital document signing party is assumed to have two related but different keys initially before engaged into the signing process. The two keys are the private key which is kept secrete and used only by the signer; and the corresponding public key which can be distributed publicly. Only the corresponding public key decrypts what is encrypted by the secrete private key. Being able to decrypt a document with a public key proves that the document is signed by holder of the private key. In addition to this, the decryption of the document with the public key gives back the message digest. The verifier of the document can calculate a message digest out of the received document and compare the calculated message digest with the message digest resulted from the decryption of the document. If the two message digests are equal, then the document is not modified. In this way public key ensures both authenticity and integrity of digital documents at the same time.

As shown in Figure 2.1 A, message digest is created for the digital object using hashing algorithms. Encrypting the message digest using the secrete private key of the signer makes a digital signature. Appending the digital signature on the digital document produces a signed document. It is possible to verify authenticity and integrity of a signed document.

The process of verifying the signed document is shown in Figure 2.1 B. As explained above, being able to decrypt the signed document using the public key verifies that the document is signed by holder of the corresponding private key. This proves the authenticity of the document. As show in Figure 2.1 B, decrypting the signed document using the public key produces a message digest. This message digest is used to verify the integrity of the digital document. That is, the verifier first calculates a separate message digest out of the digital document using a similar hashing algorithm of the signer. If the decrypted message digest and the calculated message digest are equal, then the integrity of the document is intact.

As shown in Figure 2.1, using the public key method provides evidence regarding the source and the content of a document. That is, it proves the authenticity and integrity of the digital document with evidence.

2.4 The Norwegian E-Signature Act

The Norwegian e-signature act is the act of 15 June 2001 No. 81 on electronic signatures which entered into force on 1 July 2001. According to MoReq

(2009) an electronic signature can be used to verify integrity, authenticity and is agreed to have the same legal function as hand written signature:

> An electronic signature can be used to verify that electronically transmitted information has not been altered during sending, to provide confirmation of who sent the information and as verification that the sender will not be able to deny that he sent it. These function are refereed to as securing of integrity, authenticity and non-repudiation... an electronic signature is accorded the same legal effect as a handwritten signatures(p. 21).

On the basis of the e-signature act describing electronic signature having evidential value in legal environment, the study proposes a framework embodying digital signature and develops a prototype in the context of a trusted public third party architecture. The trusted third party as described above adds value to the evidential value of digital documents by introducing an open and publicly available chain of custody.

2.5 OAIS and Trust

Open Archive Information System (OAIS) is a reference model standard for archiving information either in digital or physical form. The reference model is developed by the Consultative Committee for Space Data System (CCSDS) and approved for publication by the management council of the CCSDS and published in January 2002(CCSDS, 2002). The reference model is developed to standardize the terms and concepts in long term archiving. The common platform created by the OAIS enables implementers to understand and collaborate each other. In addition to the popularity of the model among the archival community, compliance to OAIS is one of the attributes of being a trusted digital repository(RLG-OCLC, 2002). Moreover, the standard also facilitates interoperability among archives. Allinson (2006) described OAIS as a useful model to ensure long term preservation. FurthermoreFlorence (2010) in her study finds out that archives in Norway follows the OAIS model. Plus, OAIS is the framework of development for Fedora commons which is the archival side software. It is due to the above reasons that OAIS in relation to trust is reviewed.

We examine trust in OAIS from two perspective. Trust inside the OAIS model and trust outside the OAIS model. Trust inside the OAIS model refers trust after the digital object is deposited within the archive. Trust outside OAIS refers trust before the digital object is archived which includes trust in the chain of custody of records to archive.

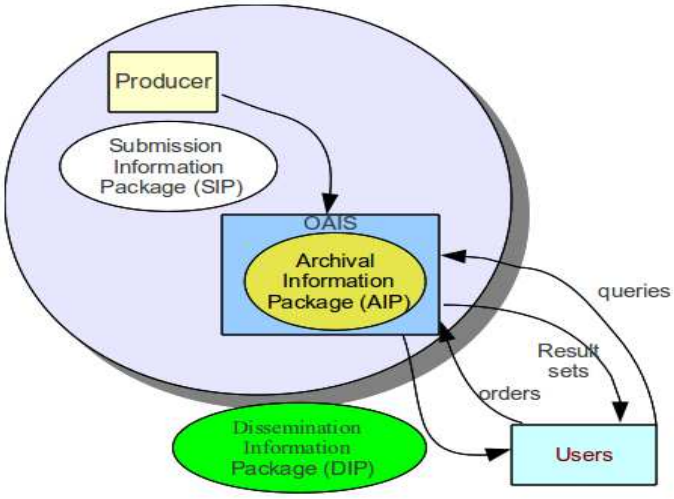

Figure 2.2: Problem domain in OAIS

Figure 2.2 depicts our problem space with respect to OAIS. The gray circle area is the problem domain of the study. Producers are content creators. In our case, the NOARK record management system is the producer of a content. See chapter four for detail on NOARK. The submission information package (SIP) as shown in the figure is an information package delivered by the producer of content to the OAIS(CCSDS, 2002). Contents of the SIP includes the records and its associated metadata. The OAIS then uses the SIP to construct the archival information package (AIP). AIP like SIP is an information package consisting the actual content plus the necessary metadata for preserving the content(CCSDS, 2002). The difference between SIP and AIP is a SIP is prepared primarily for transferring into the archive where as AIP is created to preserve the information in the archive. The AIP might include extra information. A SIP might be mapped into many AIPs during archiving. A request to the archive documents can be made using queries. The queries return back result or results to the users. The dissemination information package (DIP) can be received by users in response to orders; which is derived from one or more AIP(CCSDS, 2002).

The context of trust in the study is shown using a gray circle in Figure 2.2. The next section describes this context of trust in detail. That is, trust outside OAIS describes trust in the process of submitting a record to the archive. Trust within OAIS describes trust after the records are archived.

2.5.1 Trust outside OAIS

Trust outside the OAIS is used to refer to trust related activities carried out before receiving the digital object for archiving. Submission and "Pre-Ingest" activities and Ingest are the OAIS functions which are treated for the discussion of trust outside OAIS.

Submission and "Pre-Ingest"

Submission and "Pre-Ingest" activities are carried out before the archive accepts responsibility for archiving the content. Most of the critical activities in the submission are management works like criteria of assessing submission, collection development strategy and procedures. Ways of addressing trust in those managerial activities are out of the scope of the study. However; the pre-ingest activities which includes ensuring unique object identifier, validating integrity of the digital object and assessing the significant properties of the digital object as described in RLG-OCLC (2002) are of interest to this study.

As described in the previous section, the mechanism of validating integrity is to calculate the message digest of the digital object to be ingested and compare it with the one that is already supplied with the object. If the two message digests are the same, then the chance of that object losing its integrity along the way is almost none.

The authenticity of the object is controlled by the details of the properties supplied along with the digital object. Establishing the properties requirement at this point is not the feasible interest of the study; however, if for instance the object claims to be written by a solitude monk in the medieval, the argument behind authenticity is "is there enough evidence to support that?". So, properties that have evidential value answering who, when, why, how, for how long and the like should be provided together with the object. However; in this thesis, authenticity is used to verify whether objects are actually ingested from where they claim to be or not; which can be verified using a private and public key combination as explained above.

Ingest

As depicted in Figure 2.3 ingest is a function which allows the objects submitted as Submission Information Package (SIP) to be prepared as Archival Information Package (AIP) for storage in the archive (RLG-OCLC, 2002). Receiving SIP is one of the functionalities of ingest(CCSDS, 2002). As it is shown in chapter five, authenticity, integrity and chain of custody are main-

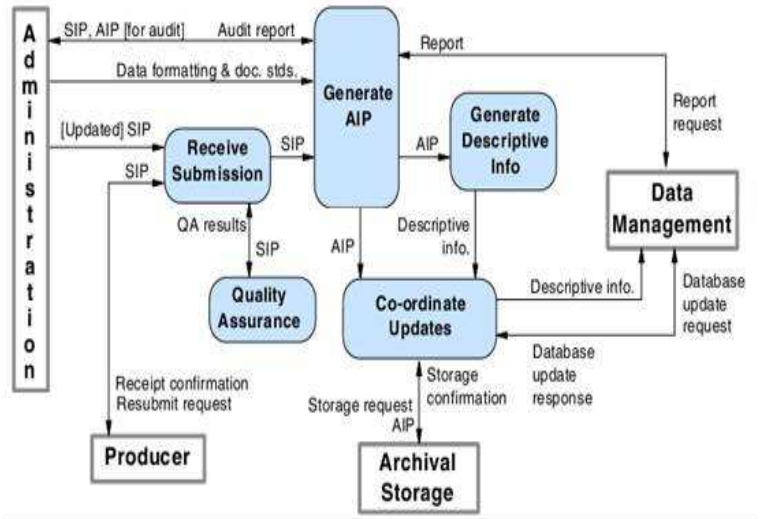

Figure 2.3: Ingest Function

tained using the public private key security mechanisms. The details of the how is also discussed in the same chapter.

2.5.2 Trust within OAIS

Trust within the OAIS is used to refer to the trust activities after the digital object is being archived and the archival body takes the responsibility of custodianship. The activities include moving the AIP after ingest into permanent storage, managing the storage hierarchy, refreshing the storage media, providing all necessary information to allow objects to be disseminated from the repository, and disaster recovery(RLG-OCLC, 2002).

Most of the trust issues after the object is ingested can be addressed by continuous integrity checking and keeping track of object access by users. Since the Dissemination Information Package (DIP) is out of the scope of this study, object access by users trust issues are not considered in this study. Cryptographic techniques changes because of security risks over time and transformations an object might undergo inside the archive; for instance format obsolescence, are the two major reasons that have an impact on the long term maintenance of integrity within the archive (Song & JaJa, 2007). In this study, Fedora Commons is used as an archiving software; and the process of integrity verification and maintenance is discussed in detail in the information modeling chapter.

As Müller, Fornaro, Rosenthaler, and Gschwind (2010) explained in their study maintaining authenticity and integrity in digital environment is a very

challenging task. Most of the studies undertaken in the preservation of digital objects integrity and authenticity uses the common cryptographic technique of hashing (Maniatis, Roussopoulos, Giuli, Rosenthal, & Baker, 2005; Shaw, 2000; Song & JaJa, 2007). These studies focus only on the maintenance of integrity and authenticity within the archive. But they do not address integrity and authenticity in the process of transferring records to archive. In addition there is no third party to validate the authenticity and integrity of the archived documents. As a result, these documents could lose their evidential value to court cases.

In addition; authenticity and integrity are not only archive data specific problems; for instance in other domains like electronic commerce, transactions between clients and service providers need to be trusted and maintain integrity and authenticity. In order to address these problems in electronic commerce, they use public key infrastructure (Wing & O'Higgins, 1999; Wang & Wulf, 1999). Public key infrastructure keeps the integrity and authenticity of transactions using a public trusted third party. The third party verifies the authenticity and integrity of the transactions; for example, the trusted third party verifies whether the amount a client transfers is unmodified throughout the whole series of transactions from the sender till the receiver end. In this work we used a public key infrastructure to improve the authenticity and integrity of archive data. Specifically we used public key infrastructure during the transfer of records to archive and after the documents are being archived. To the best of our knowledge our work would be the first one to use public key infrastructure to maintain the authenticity and integrity of archive files.

2.6 Summary

This chapter reviewed issues related to trust in the archiving of digital objects. Concepts related to trust are thoroughly analyzed for the sake establishing the focus of the study. Integrity and authenticity are used as a means of supporting evidence for a digital object to be trusted or while archived or long after it is being archived. The literature review also discussed the means of ensuring trust in the chain of digital objects custodianship using public key cryptography on the basis of which an archival framework is proposed and prototype is developed. Besides the evidential role of trust using integrity and authenticity, the OAIS model is discussed in relation to trust right before and after the object is archived. The discussion of the Norwegian e-signature act is included to analyze the admissibility of electronic documents as evidence for legal cases as Norwegian archive standard NOARK is the record

side context of the case study.

Chapter 3

METHODOLOGY

3.1 Purpose of the research

The research analyzes the functional requirements of archiving trusted digital objects having evidential values in court cases. The results of the study will identify the functional requirements and ways of ensuring authenticity, integrity and chain of custody of trusted digital objects archiving using NOARK-5 as a record management and Fedora Commons as an archive repository case. The study assumes that there exists a third party which is trusted and contents verified by it are legally acceptable.

3.2 Theoretical framework

Interpretive is the underlying assumption guiding this study. The researcher aims at constructing meaning in the handling of trusted records and archive from the context of the study; and as the result interpretive is one of the suitable underlying epistemological foundations for achieving such assumption(M. D. Myers, 2011). In addition, as Klein and Myers (1999) pointed interpretive has the potential to produce deep insight into information systems phenomena including the management of information systems and information systems development which is what this study seeks to realize by constructing meaning of trust in the context of the NOARK-5 and Fedora Commons. The functional requirements of authenticity, integrity and trusted chain of custody is constructed using the shared understanding of the concepts. The constructed meaning then will be applied to the context of the study.

3.3 Approach/Methods

The study analyzes the functional requirements of archiving trusted Norwegian electronic public administration records from shared contextual literature using an interpretive framework. This makes the study qualitative. Moreover, the data collected and the procedure followed to address the research questions are textual data which are used for the construction of the contextual meaning of the problem- trust. Furthermore, in a qualitative approach the data sources used include documents, texts and the researcher's impressions and reactions as M. Myers (2008) described it; so does document standards related to trusted archiving and maintenance and how that relate to the context of the study that is NOARK-5 and Fedora Commons are consulted to answer the research questions. Trust more than often tend to be the application of a constructed judgment of the object presented to an individual (Arne-Kristian Groven, 2008). Hence, a qualitative approach provides the benefit of analyzing the process of trusting a record or an object in the context of the study. As a result qualitative approach is chosen to be more suitable to the problem at hand.

3.4 Research Strategy

The process of transferring records from NOARK to the archive Fedora Commons and how authenticity, integrity and trusted chain of custody can be maintained is the focus of the study. NOARK is chosen because it is becoming a record keeping standard of the public sector in Norway; the country where the thesis is carried out. Proximity and accessibility other than the strong and clear records management tradition of Norway as (Florence, 2010) described are the main reason behind the choice of NOARK. Though getting real NOARK data is not as easy as it was initially anticipated, with the help of the NOARK standard specification sample records are generated for the purpose of validating our approach. The growing online user communities and more than ten years[1] of sustainable development and refinement of Fedora Commons to meet archival interests is the other reason for choosing Fedora Commons as the context of the study. Trust is studied using NOARK and Fedora Commons as the result an explanatory case study is employed as a strategy for undertaking this research.

[1] `http://www.fedora-commons.org/about/history`

3.5 Prototyping

NOARK produces records and marks them finalized once records are no more expected to undergo further transaction leading to modification. A MakeXMLFile extractor(see appendix) is applied to make XML files out of the NOARK records. Once the XML file is generated out of the records, XMLSec which is an XML security library(see section 5.2.1 for details) is used to sign the records which produces record export format. Record export is the proposed format for making records ready for transfer to the archive. The submission information package(SIP) then prepares the records and transfers them to the archive Fedora Commons. Figure 3.1 depicts the process of prototyping.

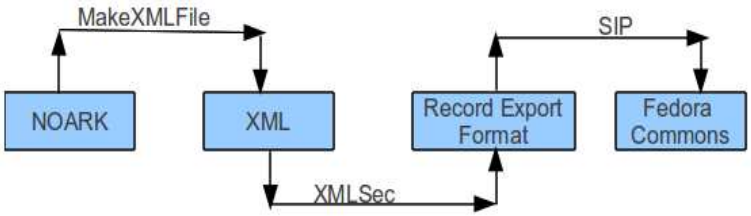

Figure 3.1: Prototyping- the process

3.6 Validation

As a proof of concept we implemented a prototype. We validated authenticity and integrity during and after transfer of records to archive. A transfer process is authentic if the signature of the records from NOARK and objects from Fedora Commons is verified authentic by the third party. If the message digest of the records and objects in the archive is verified unmodified and intact by the third party, the records and objects are proved to have maintained integrity. To do these, sample NOARK records are generated and a signature is appended into the individual records. After the records are signed and made ready for submission into the archive, a manual modification of records is done to see if the prototype can detect the failure in authenticity and integrity during and after transfer. As the result, the prototype detected the change and produced a failure message indicating the records compromise in integrity and authenticity (for more details, see Appendix).

3.7 Limitations of the Study

NOARK-5 data is not yet available for archival at the time of the research undertaking. As Florence (2010) described it will take four to five years before we see NOARK-5 documents being archived in archival institutions. Even if NOARK-4 and NOARK-3 data is available, it is not possible to get sample NOARK data to work with. This is primarily due to the fact that the NOARK data contains personal and sensitive information. However; sample NOARK data is generated on the bases of the NOARK specification to circumvent the limitation.

3.8 Ethical considerations

Initially it was anticipated that NOARK data will be obtained from the Norwegian national archive. On the bases of this assumption, the data was anticipated to be used merely for the purpose of doing the research maintaining confidentiality. Since the sample data is generated instead of working on the real data, there is no major ethical concern.

Chapter 4

INFORMATION MODELING

4.1 Introduction

This chapter deals with NOARK-5 and Fedora Commons in detail. Emphasis on NOARK-5 record structure is given to establish a framework for NOARK-5 compliant sample records generation in the next chapter. Fedora Commons object structure and its metadata support is also discussed in this chapter. A functional view of proposed architecture and details of ensuring trust in the chain of submitting records from NOARK to Fedora Commons is the main focus of the chapter leading to the conclusion.

4.2 NOARK-5

The use of the term NOARK from here on refers NOARK-5 and reference to other versions of NOARK will use explicit version number followed by NOARK like NOARK-3 to refer to version 3 of NOARK. Though NOARK is both a recordkeeping and an archive specification, it is referred in this thesis as a recordkeeping system. A recordkeeping system is used to support the day to day activities of a business organization(Duranti & MacNeil, 1996).

NOARK is organized into three distinguished layers of abstraction named inner core, outer core and complete. The inner core is the must meet requirement for a solution to be a NOARK compliant (MoReq, 2009). The inner core requirement includes functionality for archiving, archiving based on the requirements of the law and regulations, functions for administration and operation of the core. The outer core is an optional component used for integrating the inner core with other per-existing systems implemented with individual organizations requirement. The NOARK complete functionality is for integrating external solutions with the NOARK core functionality. Figure

4.1 depicts the organization of the NOARK functionality taken directly from the NOARK standard specification. The inner core comprises functionality

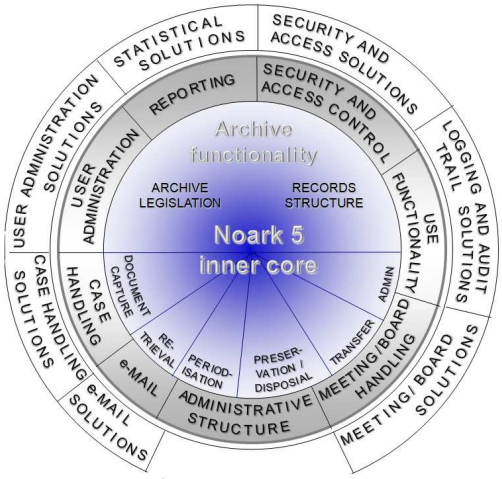

Figure 4.1: NOARK 5

including document capture, retrieval, peridoisation, preservation/validation, transfer and administration where as the outer core includes e-mail, reporting and user administration. Solutions like e-mail solution and user administration solutions are part of the NOARK complete. The difference between e-mail in the outer core and e-mail solutions in the complete core is, the e-mail in the outer core refers to a preexisting e-mail service within an organiztion implementing NOARK whereas the e-mail solution in the complete refers the integration of the organization e-mail system with external e-mail solutions.

Since the study uses NOARK as a recordkeeping system to specify its record structure, the next section discusses the record structure of NOARK on the bases of which sample records are generated.

4.2.1 Record: Metadata and Structure

A record can be a result of a transaction between two parties. History of legal proceedings, transcript of a trial or all the testimonies and items introduced into evidence and lead to a particular decision can all be an example of a record. Figure 4.2 demonstrates a typical transaction leading to the creation of a record. A particular office receives and processes an application and dispatches a reply on the basis of approval from the manager. A typical

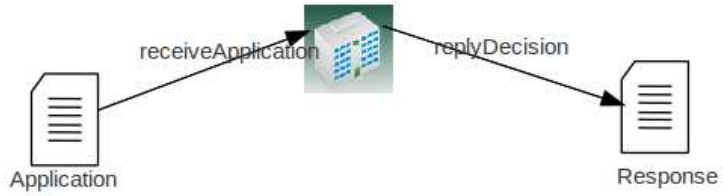

Figure 4.2: Transaction

record about this transaction comprises the reception of the application, the reasoning leading to a decision on the application and the final response to the application received. The registry of this particular context is an instance for the creation of record. Records are then organized into files.

To delve into the structural requirements of records, it is necessary to identify the types of records in NOARK context. NOARK identified two types of records- simplified and basic records. Simplified records are records that fit into the existing record structure and containing all the necessary metadata in order to link the record into the existing record structure. Basic record is a record to keep track of records involving transaction. The metadata requirement for simplified record and basic record is shown in the following xml formatted structures shown in Figure 4.3 and Figure 4.4. Those are the metadata requirements for NOARK records. In addition to the metadata requirements, the standard lists a set of structural requirements for record. The obligatory structural requirement for records include:

1. It is obligatory for a simplified record to be divided into different types

2. It is obligatory if the file level is used, a simplified record must belong to (only) one basic file and a basic file can contain no, one or several simplified records

3. It must be possible for a simplified record to contain no, one or several document descriptions and a document description must be included in one or more simplified records

4. It must be possible to expand a simplified record to a basic record

The structural requirements are further illustrated using the conceptual diagram of a record shown in Figure 4.5. The idea behind reviewing the record structure and metadata is to assess the overview of integrity and authenticity

```
<srecord>
<!--  metadata for simplified record -->
      <systemID>
      <!--  it is obligatory and must be included during transfer to archive -->
      </systemID>
      <recordtype>
      <!--  it is obligatory and must be included during transfer to archive -->
      </recordtype>
      <createdDate>
      <!--  it is obligatory and must be included during transfer to archive -->
      </createdDate>
       <createdBy>
      <!--  it is obligatory and must be included during transfer to archive -->
      </createdBy>
      <archivedDate>
      <!--  it is obligatory when the document is archived -->
      </archivedDate>
      <archivedBy>
      <!--  it is obligatory when the document is archived -->
      </archivedBy>
      <referenceParent>
      <!--  it is obligatory reference to a file class  -->
      </referenceParent>
      <referenceRecordssection>
      <!--  it is an optional and included if a series is used  -->
      </referenceRecordssection>
      <referenceDocumentdescription>
      <!--  it is obligatory and can occur more than once -->
      </referenceDocumentdescription>
      <referenceDocumentobject>
      <!--  it is optional and can occur more than once -->
      </referenceDocumentobject>
</srecord>
```

Figure 4.3: metadata for simplified record

at a record level in NOARK. Most of the activities at record level are related to authenticity, activities like timestamping and logging records creator as shown in the metadata structure. The standard also imposes integrity requirements of metadata for a record once it is marked final. The standard explicitly specifies an obligatory requirement for a record marked finalized, it must not be possible to add more description to it. However; it does not further specify how to ensure this in the implementation.

4.3 Fedora Commons

Fedora[1] is an acronym for Flexible Extensible Digital Object Repository Architecture developed by the Digital Library Research Group at Cornell University as an architecture for storing, managing and accessing digital contents in the form of digital objects. The licensing philosophy behind fedora's development is free and open source under the apache version 2.0 licensing creating an open pool of development and testing platform from

[1] `http://www.fedora-commons.org/about`

```
<brecord>
      <!--  metadata for basic record -->
      <recordID>
      <!--  it is obligatory and is recommended yy/nnnnnn-nnnn format-->
      </recordID>
      <title>
      <!--  it is obligatory  -->
      </title>
      <officialTitle>
      <!--  it is obligatory if words in the title are screened during transfer  -->
      </officialTitle>
      <description>
      <!--  it is optional and occurs only once -->
      </description>
      <keyword>
      <!--  it is optional and can be repeated -->
      </keyword>
      <author>
      <!--  it is obligatory and can be repeated -->
      </author>
      <documentmedium>
      <!--  it is obligatory for mixed physical and electronic record -->
      </documentmedium>
      <storagelocation>
      <!--  it is optional and is used for physical records -->
      </storagelocation>
</brecord>
```

Figure 4.4: metadata for basic record

developers all over the world giving fixes and solutions to bugs from users and developer community. The current version of fedora is version 3 that is the version used in the study which can be used to operate as a standalone server, which is designed to be used with other software for designing a complete solution(TheFedoraDevelopmentTeam, 2008). The key research questions leading the development of fedora are interaction with heterogeneous collections of complex digital objects, the design of generic and genre-specific objects, association of services and tools with objects, control policies on individual and group objects, long term management and preservation of digital objects. As can be understood from the key areas of questions, the heart of the questions circle around digital objects. The next section discusses fedora objects in detail.

4.3.1 Fedora Objects

Digital objects are the central building blocks of fedora architecture. Objects in fedora are defined generically to provide essential characteristics for many types of digital contents like documents, images, multimedia, metadata and more. Fedora objects are aggregates of one or more content items into an object. Figure 4.6 shows basic model of a fedora object taken from fedora documentation[2].

[2]`http://fedora-commons.org/documentation/3.0b1/userdocs/digitalobjects/objectModel.html`

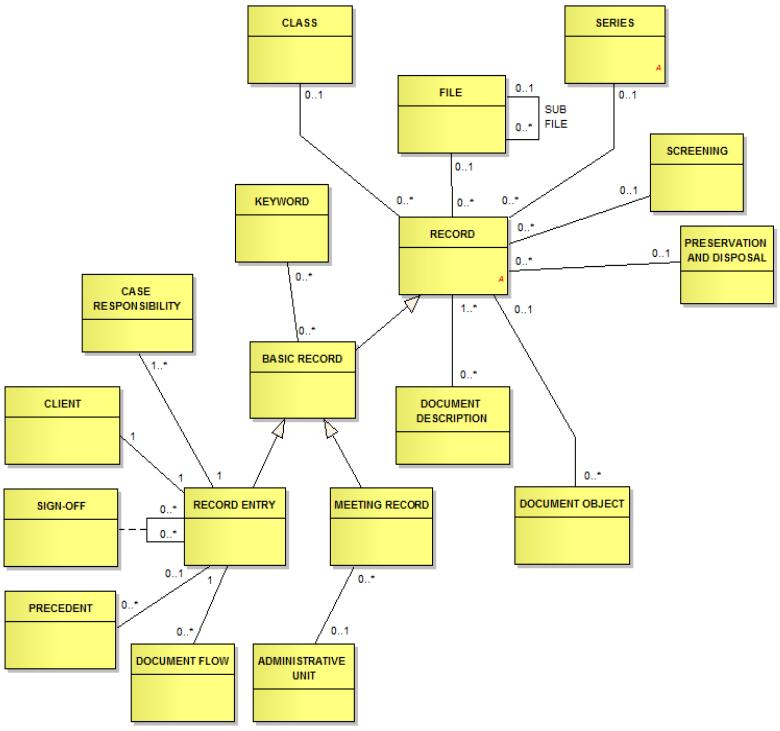

Figure 4.5: Conceptual model at record level

An object has a persistent unique identifier which can either be defined by users or automatically assigned by the repository. With the restriction of not more than being 64 characters, a valid object identifier can be “demo:1” or “demo:myrepository” which are also case-sensitive. The datastream represents the actual content item and metadata of the content item. The object properties are system defined to manage and track the object in a repository. These include datastream identifier for uniquely identifying the datastream within an object; datastream state- one of the options from active, inactive or deleted; created date; modified date; versionable which is set either true or false so as to keep the original along with modified versions; descriptive label for datastream; MIME type of datastream; format identifier; alternate identifier; checksums which is used to keep track of the integrity of the datastream using standard algorithms; byte stream content and control group which could be internal xml if the content is stored internally as xml document, managed content if both datastream and metadata stored internally in the repository, externally referenced if the content is stored outside the repository and the repository is middling in streaming the content access, and redirect reference content which the repository is used to store the universal

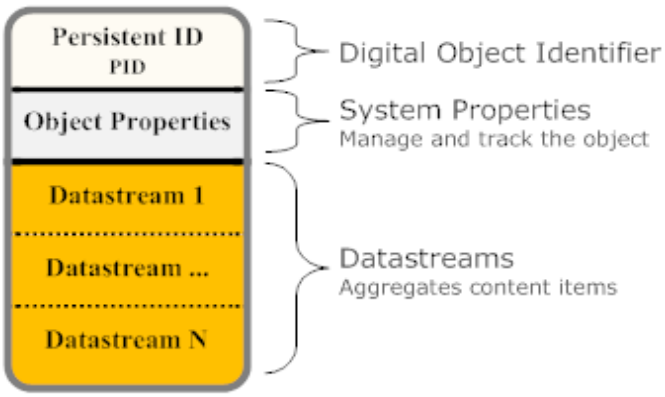

Figure 4.6: Fedora object model

resource locator(URL) and access to the digital object happens without the repository middling between.

Object Types

Fedora identifies four types of digital objects. All the different object types conforms to the basic fedora object discussed above. Data object, service definition object, service deployment object and content object model.

Electronic texts, images and books are some of the representations that fit in data objects. Data objects are the actual content in the repository. They can freely be shared between fedora repositories. Service definition object is a control object used to store an interface listing the operations that are supported by the data object which helps to add customized functionality to data objects. In other words by using service definition objects, it is possible to define a contract of supported operations on data objects. A collection of service definition object constitutes a registry of service definitions in the repository. Service deployment object describes how the operations defined in the service definition object are delivered by the repository. The content model object is a formal model characterizing class of digital objects. Model of permitted relationships, excluded or required between models of digital object can be provided with content model objects.

The rationality behind highlighting the different types of fedora objects is to understand object types and on the basis of which handling authenticity and integrity of records transfered into objects of fedora archive can be made. As can be seen from the different types of objects, they all share the common object properties which include checksum as one of the properties.

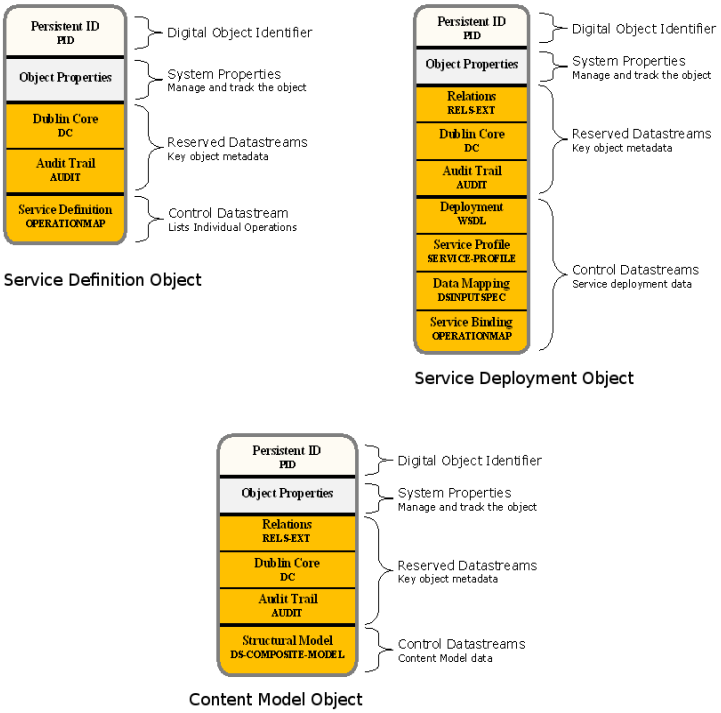

Figure 4.7: Fedora Object Types

As a result, placing an integrity checksum on fedora objects can be used to ensure objects integrity during ingest or submit. The next section discuss the common metadata supported in fedora commons.

4.3.2 Fedora commons: metadata

This section describes the two common metadata fedora supports for describing the objects discussed above – FOXML and METS. Both describes objects in fedora repository using the standard xml format.

FOXML: Fedora Object XML

FOXML is a simple xml format used to ingest and export objects to and from fedora repository. In addition, it is also used to store objects internally into the fedora repository. FOXML xml schema defines elements that correspond to fedora digital objects.

At the highest level FOXML digital object comprises object properties and data streams as described above. The skeleton xml representation of

fedora object looks like the figure shown in Figure 4.8.

```
<digitalObject PID="uniqueID">
    <!--core object properties -->
        <objectProperties>
            <property/>
            <property/>
            ....
        </objectProperties>
    <!-- datastream section; there can be zero or more datastreams -->
        <datastream>
            <datastreamVersion/>
            <datastreamVersion/>
            ....
        </datastream>
 </digitalObject>
```

Figure 4.8: FOXML schema

A particular interest of this study is trust. FOXML address integrity of objects while ingesting or after the objects are stored within the repository using checksums. The checksum is specified using contentDigest tag within the datastreamVersion tag.

Figure 4.9 shows how to place an md5 integrity checksum on the object to be ingested in to Fedora Commons archive using FOXML.

```
    .....
<datastreamVersion ID= "IntegrityDemo.1" Label = "Integrity" MIMETYPE = "application/pdf">
    <contentDigest TYPE= "MD5" DIGEST= "044f098f7dec57d2cea457e1f0198cf9"/>
    ....
</datastreamVersion>
```

Figure 4.9: FOXML Integrity support

METS: Metadata Encoding and Transmission Standard

METS provide an xml document for encoding metadata necessary for the management of digital objects within a repository and exchange of such objects between repositories. METS document could be used in the role of submission information package (SIP), archival information package (AIP) or dissemination information package (DIP) within the OAIS reference model[3]. The standard METS document comprises of seven subsections – METS Header describing itself, Descriptive Metadata, Administrative Metadata, File Section, Structural Map, Structural Links and Behavior.

Figure 4.10 shows the standard elements comprising a METS document. However; fedora commons tailored METS to meet the specific requirements

[3]http://www.loc.gov/standards/mets/METSOverview.v2.html

```
<mets>
        <metsHdr>
    <!--mets header records minimal metadata regarding the mets document itself-->
    <!--created date, modified date, names of agents played role creating the mets -->
        </metsHdr>
        <dmdSec>
    <!--descriptive metadata section may contain to pointer to external metadata -<mdref> -->
    <!--or pointer to internally embedded metadata - <mdWrap>  -->
        </dmdSec>
        <amdSec>
    <!-- administrative metadata to the files comprising the digital object   -->
        </amdSec>
        <fileSec>
    <!-- contains one or more <fileGrp> elements used to group together related files -->
        </fileSec>
        <structMap>
    <!--presents to users of the digital library a hierarchical structure for navigation -->
        </structMap>
        <structLink>
  <!-- used to record the existence of hyperlinks between items within the structural map -->
        </structLink>
        <behaviorSec>
    <!-- used to associate executable behaviors with content in the METS object -->
        </behaviorSec>
</mets>
```

Figure 4.10: METS skeleton

of fedora commons. Enumerating how the standard mets mapped into fedora METS is not the focus of this study. For the purpose of embarking on to one of the objectives of the study, integrity of an object is specified in the file section of the metadata for individual files constituting the object. Figure 4.11 is an excerpt of an object encoded in mets for fedora ingest.

The above section discussed FOXML and METS in the context of fedora. The discussion of the metadata described how to encode integrity in the process of ingesting digital objects into the fedora repository. The next section describes the proposed architecture of our study.

4.4 Proposed Architecture

This section describes a high level view of the proposed architecture. In addition to the high level view architecture, a functional view of the proposed architecture addressing trust requirements are also discussed.

```
                    ....
<METS:fileSec>
  <METS:fileGrp ID="DATASTREAMS">
    <METS:fileGrp ID="IMAGE" STATUS="A">
      <METS:file ID="IMAGE.0" MIMETYPE="image/x-mrsid-image" OWNERID="E"
CHECKSUM="c07f516d77d8a5ca452775d489ffe78c"
                          CHECKSUMTYPE="MD5">
        <METS:FLocat LOCTYPE="URL"
          xlink:href="http://demo.prototype.edu/test/demoImage.sid"
          xlink:title="Sample Demo Image, Fedora Mets"/>
      </METS:file>
    </METS:fileGrp>
    <METS:fileGrp ID="DRAWING-ICON" STATUS="A">
      <METS:file ID="DRAWING-ICON.0" MIMETYPE="image/jpeg" OWNERID="M"
          CHECKSUM="6497ceb7a8477fbe9ba4ff9e6e57999f"
                          CHECKSUMTYPE="MD5">
        <METS:FLocat LOCTYPE="URL"
          xlink:href="http://demo.prototype.edu/test/demoImage.jpg"
          xlink:title="Sample Demo Image, Fedora Mets JPG"/>
      </METS:file>
    </METS:fileGrp>
</METS:fileSec>
                .....
```

Figure 4.11: Object Integrity encoding fedora METS

4.4.1 High level view

At a very high level, records are submitted to the archive. As discussed above, records are handled by NOARK. The NOARK records are then submitted to Fedora Commons archive repository. The assumption is both NOARK and Fedora Commons communicate via the public Internet infrastructure.

Figure 4.12 shows the process of submitting records into the archive without addressing trust requirements. Records submitted this way have very little evidential value for court cases. As discussed in the literature section, a trusted third party verifies trust requirements – integrity and authenticity which adds evidential value to the content being archived. The trusted third party verifies the authenticity of the records and the archive. Detailed discussion about the trusted third party is treated in the functional view section of this chapter.

The high level view with trust(Figure 4.13) is the proposed architecture of this study. The trusted third party is a public body which is trusted by both the NOARK solutions and the Fedora Commons archival solutions for securing their interaction. In addition, since the trusted third party is a public body, the contents authenticated by it will be assumed to have legal acceptance. The next section discusses the functional view of the proposed

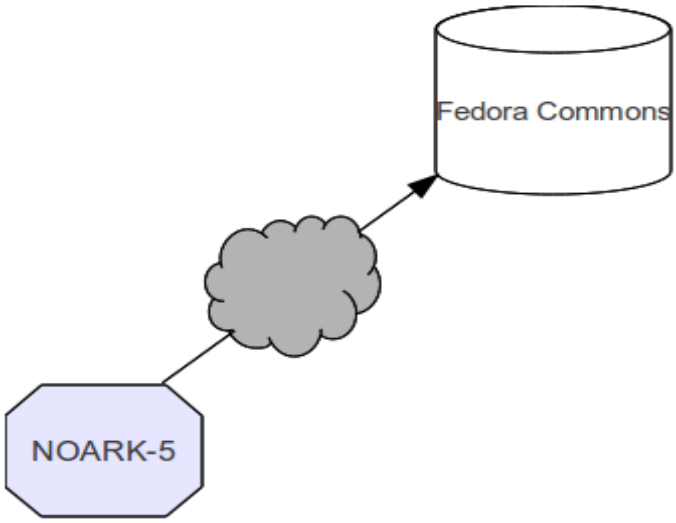

Figure 4.12: High level view without trust

high level view.

4.4.2 Functional view

The following section describes the functional view of the high level architecture depicted in the above section.

The NOARK record specification specifies two obligatory functional requirements for a record and one optional. The obligatory requirements are:

- There must be a service/function for updating an Administrative unit and Executive officer on a Record (Registry entry)
- There must be a service/function for updating a Client on a Registry entry.

The optional requirement specifies a must need for a service/function for updating a Registry management unit on a Record (Registry entry).

On the bases of the above requirement specifications, the functional view of the NOARK record is expressed(see Figure 4.14).

Record export is proposed to be the format of records when they are ready to be transfered to the archive; which can either be ingested by fedora commons or prepared by the submission information package for submission. As a result, trust parameters – integrity and authenticity check are placed with respect to the record export.

Figure 4.15 shows the structure of the record export.

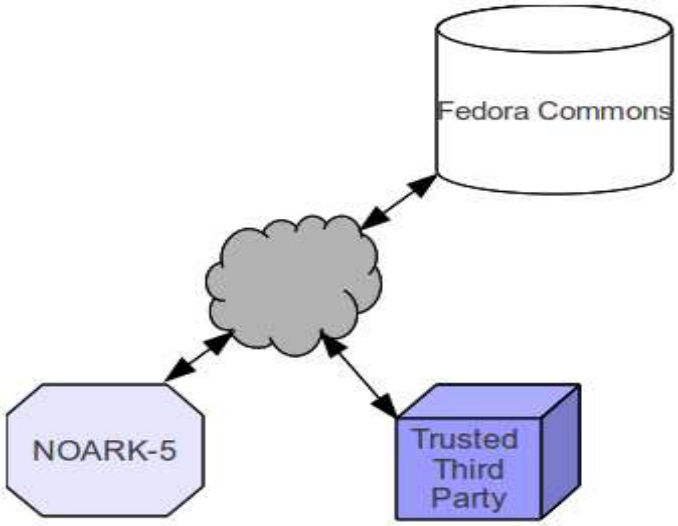

Figure 4.13: High level view with trust

The trusted public third party keeps the public keys of NOARK and Fedora Commons initially. For the purpose of distinguishing the keys of NOARK and Fedora Commons, a label of DS-NO and DS-FC are used respectively. During the process of verifying the submission process, the third party is also proposed to keep its own copy of the contents transferred. Having a copy of records transfered to the archive by the third party will have the following advantages:

- it enables further verification of objects in the archive before moving the objects to permanent archive
- redundancy increases the chance of records availability
- evidential value will be improved if records are also deposited with the third party.

The numbers in Figure 4.17 indicates the order of messages exchanged between NOARK, Fedora Commons and Trusted Third Party. Message labeled as <1> takes place first then <2> takes place and so forth. A request to the submission information package (SIP) from the archive via the public infrastructure is made by NOARK as shown in the figure. A reply to the request is verified by the trusted party using the public key of archive. After the verification of the SIP, the authenticated SIP is replied back. After the records are prepared for submission using the authenticated SIP, a signature of the NOARK will be appended and a submit operation will be carried

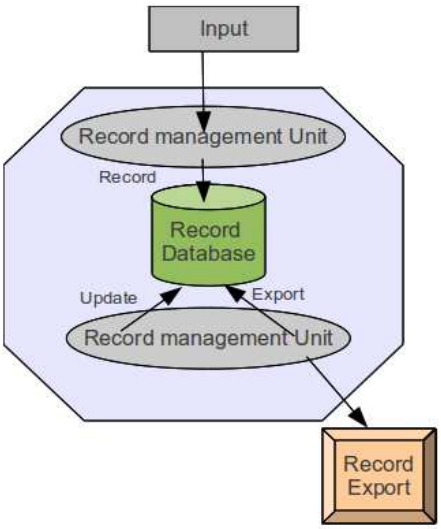

Figure 4.14: Functional view of NOARK record

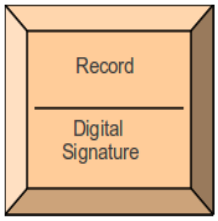

Figure 4.15: Record export format

out. The trusted party again verifies the signature of NOARK and copies the whole content and sends it to the archive and itself. Finally the content will be deposited in a temporary storage before moved into the archive for permanent archiving. This helps the archive to verify the authenticity and integrity once again against the trusted third party and move the content to permanent storage. Figure 4.18 shows the signature validation strategy used to verify authenticity and integrity of records.

4.5 Public Key Cryptography

Cryptography is used to convert a document or message into a scrambled or disguised data. The disguised or scrambled data is assumed to be safe from being read by anyone except those the document or message is meant for. The process of converting a document into scrambled data is often referred

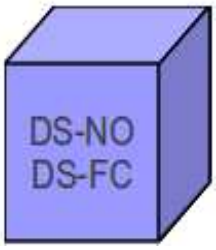

Figure 4.16: Trusted third party

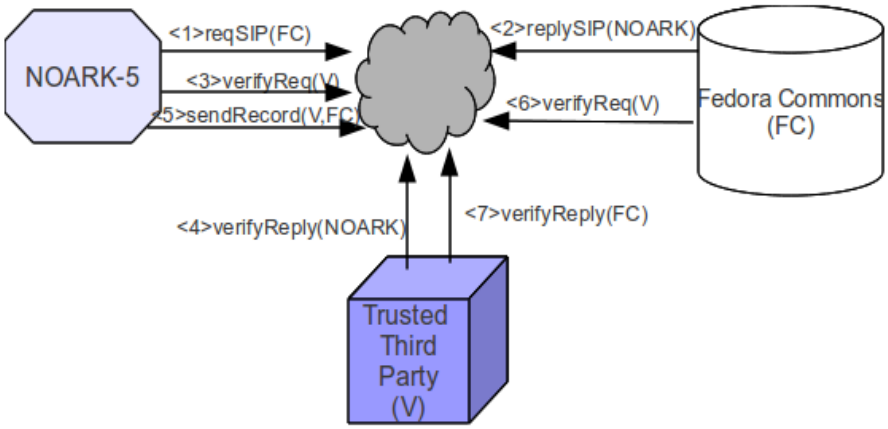

Figure 4.17: The proposed architecture: functional view

in cryptographic terms encryption. The reverse process, that is converting the scrambled data back into original message is called decrypting. The two common forms of encrypting a message are symmetric encryption and asymmetric encryption.

Symmetric encryption technique uses a shared secrete key to encrypt and decrypt the message. To ensure the safety of the message, the shared secrete key must be kept secrete all the time. This type of encryption is very easy to use and consumes less computing power. Since the secrete key is shared, authenticity cannot be proved. In addition to this, when the number of participants exchanging the message increases, distributing the shared secrete key securely become problematic. However; one of the objectives of our study is proving authenticity of records. As the result symmetric encryption technique is not our preferred choice.

Asymmetric encryption on the other hand uses pairs of keys. As described in chapter two, the pairs of keys are public key and private key. The private key as the name implies is private and kept secrete. The public key can

```
READ record, recordSignature, DS-NO, DS-FC
    SET hash TO decrypt(recordSignature)
    SET hashRecord TO digest(record)
    IF hash equals hashRecord
        signature is valid
    ELSE
        signature is invalid
```

Figure 4.18: Validate signature

either be distributed publicly or kept in a public place. Unlike symmetric encryption technique, the exchange of keys in asymmetric encryption can take place openly. One of the pairs of keys is used to encrypt the document and the other to decrypt the document. There is no need to share keys. It is demonstrated in chapter two how this method proves authenticity and integrity of a document. Due to this capability, public key cryptography uses asymmetric encryption technique; which is the preferred choice for the framework we proposed in the study.

One of the basic questions to ask in relation to asymmetric encryption technique is how to distribute public keys? The answer to this as shown in our proposed framework is the trusted third party. The trusted third party also serve as a key distribution center which is commonly known as certificate authority. A certificate authority is a public organization who registers and provides assurances that a particular public key belongs to a particular organization or person(Mel & Baker, 2001). In our case, the trusted third party registers the public keys of NOARK and Fedora Commons in its registry and publicly distributes the keys. In addition to distributing the keys, the trusted third party also provides assurances that the keys belong to NOARK and Fedora Commons as shown in Figure 4.17.

4.6 Summary

The chapter discussed the structure of NOARK records and their metadata requirement. In addition, discussion of fedora objects structure and the common supported metadata types of fedora commons is given. The discussion of NOARK records is given in order to generate a sample NOARK data for testing the proposed framework. Furthermore, it also helps to match the records to the archival metadata requirement for permanent storage maintaining trust requirements. Finally the functional view of the proposed framework and signature validation strategy is discussed.

Chapter 5

EXPERIMENTS AND FINDINGS

5.1 Introduction

This chapter describes the experimental settings and findings of the study. Details of the hardware and software used, the sample data and findings answering the research questions are the sections constituting the chapter.

5.2 Experimental Settings

5.2.1 Hardware and Software

A personal computer 4GB of RAM, Intel Pentium Dual-Core cpu of 2.00GHz is the hardware used to carry out most of the development work.Ubuntu Linux, Microsoft XP, XMLSec, MySQL and Java are the software tools used to carry out the study. Ubuntu is used as a server operating system hosting the archive repository Fedora Commons. Microsoft XP is used as a client requesting the submission information package for submitting records in to the archive or as client where Fedora Commons ingests records for archiving. XMLSec[1] is a library software which supports major security standards like XML signature and XML encryption. MySQL is used to create NOARK records in a database. Java is the programming language used to generate sample NOARK records and extract them in XML format from the database.

[1] `http://www.aleksey.com/xmlsec/`

5.2.2 Data

Sample records

The process of creating sample records is carried out on the popular open source database software MySQL. A database called noark is created with two flat tables named srecord and brecord signifying simplified record and basic record respectively. The following tables summarized the data types of the fields in the srecord and brecord table according to the NOARK metadata catalog.

Table 5.1: Simplified record

Catalog number	Name	Data type	Remark
M001	systemID	TextString	Unique
M081	recordtype	TextString	Undefined
M600	createdDate	DateTime	
M601	createdBy	TextString	
M604	archivedDate	DateTime	
M605	archivedBy	TextString	
M200	referenceParent	TextString	Undefined
M208	referenceRecordssection	TextString	arkivdel.systemID
M207	referenceDocumentdescription	TextString	Unique
M216	referenceDocumentObject	TextString	Undefined

Table 5.2: Basic record

Catalog number	Name	Data type	Remark
M004	recordID	TextString	Auto, e.g 2011/3869-8
M020	title	TextString	
M025	officialTitle	TextString	Public name of archive unit
M021	description	TextString	Textual description
M022	keyword	TextString	Keywords describing contents
M024	author	TextString	Name created or authored
M300	documentmedium	TextString	Optional for electronic
M301	storagelocation	TextString	Optional for electronic

Once the database tables are created, sample records are added into the tables. Now at this stage we have NOARK data into our database. Since

scalability is not an issue in our study, we took two records one from each to address the integrity and authenticity requirements. The Java MakeXMLFile code (see Appendix) is applied to extract records from the database and convert them into XML. The records at this point are in standard XML format. The next step is to prepare the XML files for submission. To do that, the records in the standard XML format should be prepared in the record export format discussed in the next section.

Record export

As discussed in the information modeling section, the format of the record export is the records with digital signature appended on.

To prepare the record export format, additional XML tags are added to hold the signature information. This is necessary to separate the record information from the signature information. Figure 5.1 and Figure 5.2 shows the tags added to append the signature information.

After signature tags are added, the record is signed with the private key of NOARK using XMLSec resulting the record export format(see Figure 5.3 and Figure 5.4) which is ready to be submitted into the archive using the submission information package(SIP). A SIP; whose authenticity is verified by the third party, is requested from Fedora Commons to submit the signed records into the archive. As shown in Figure 5.5 a request to SIP is made by using the ip address of the SIP as we do not have a domain name registered for it. The third party verifies the identity of the SIP using its public key which is registered and found with the third party. As can be seen from Figure 5.6 the message "The identity of this website has not been verified" is displayed as the key is not publicly registered. To address this public registration problem for the purpose of testing the prototype, the public key is manually verified as shown in Figure 5.7.

Finally the records are signed using the private key giving the record export format(see Figure 5.3 and Figure 5.4).

5.3 Findings

* *Functional requirements of Authenticity and Integrity*
  Based on the literature studies done in this work, the functional requirements of authenticity and integrity for archival objects could be appending digital signature of the record management system to the records to be archived. The digital signature as discussed in chapter two, is made by signing the message digest value of the record with the

```
<?xml version="1.0"?>
<noark5>
    <srecord>
        <systemID>890</systemID>
        <recordtype>Internal Document without follow-up</recordtype>
        <createdDate>2011-02-04 10:30:55</createdDate>
        <createdBy>Immegiration Office</createdBy>
        <archivedDate>2011-05-06 20:44:44</archivedDate>

        <archivedBy>Immegration Office</archivedBy>
        <referenceParent> </referenceParent>
        <referenceRecordsection>2011/3869-1.980</referenceRecordsection>
        <referenceDocumentdescription>A student registration
                day</referenceDocumentdescription>
        <referenceDocumentObject> </referenceDocumentObject>
    </srecord>
  <Signature xmlns="http://www.w3.org/2000/09/xmldsig#">
  <SignedInfo>
   <CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315" />
   <SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-sha1" />
   <Reference URI="">
    <Transforms>
     <Transform Algorithm="http://www.w3.org/2000/09/xmldsig#enveloped-signature" />
    </Transforms>
    <DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1" />
    <DigestValue></DigestValue>
   </Reference>
  </SignedInfo>
  <SignatureValue/>
  <KeyInfo>
    <KeyName/>
  </KeyInfo>
 </Signature>
</noark5>
```

Figure 5.1: Simplified record signing template

private key of the record keeping system – NOARK. Our experimental result shows that digital signature is sufficient condition to ensure authenticity and integrity of archival objects.

* *Functional requirements of chain of custody*
In order to make the process of transferring records to archive we have introduced a trusted third party. Documents verified by the trusted third party as it is described in chapter three, are legally acceptable. For that to happen, the third party needs to verify both the record management system (NOARK) and the archive repository (Fedora Commons). In addition to this, the third party also stores the documents transfered to the archive in its own database. By introducing third party as it was stated in chapter four who can manage records transfer to the archive by verifying and copying archive data could improve

```
<?xml version="1.0"?>
<noark5>
    <brecord>
        <recordID>2011/3869-1</recordID>
        <title>Registration</title>
        <officialTitle>International Students Registration</officialTitle>
        <description>A student registration day</description>
        <keyword>'student' 'International' '2011'</keyword>

        <author>Immegration Office</author>
        <documentmedium>electronic</documentmedium>
        <storagelocation> </storagelocation>
    </brecord>
  <Signature xmlns="http://www.w3.org/2000/09/xmldsig#">
  <SignedInfo>
   <CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315" />
   <SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-sha1" />
   <Reference URI="">
    <Transforms>
     <Transform Algorithm="http://www.w3.org/2000/09/xmldsig#enveloped-signature" />
    </Transforms>
    <DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1" />
    <DigestValue></DigestValue>
   </Reference>
  </SignedInfo>
  <SignatureValue/>
  <KeyInfo>
    <KeyName/>
  </KeyInfo>
 </Signature>
</noark5>
```

Figure 5.2: Basic record signing template

the evidential value of the archival objects. For instance, proofing the archive objects against the third party could improve the evidential value of the object for court cases.

* *How to ensure authenticity and integrity during and after transfer*
In order to ensure authenticity and integrity during and after transfer we used the signature validation strategy mentioned in chapter four(see Figure 4.18) and based on our finding considering the signature validation strategy ensures authenticity and integrity during and after records are transferred into the archive. That is the records transferred to the archive using our strategy are found to be authentic and the content is intact. We have deliberately changed the content and signature of records to be transferred to the archive and our proposed solution properly detected the failure in authenticity and integrity of

```
<?xml version="1.0"?>
<noark5>
      <srecord>
            <systemID>890</systemID>
            <recordtype>Internal Document without follow-up</recordtype>
            <createdDate>2011-02-04 10:30:55</createdDate>
            <createdBy>Immegiration Office</createdBy>
            <archivedDate>2011-05-06 20:44:44</archivedDate>

            <archivedBy>Immegration Office</archivedBy>
            <referenceParent> </referenceParent>
            <referenceRecordsection>2011/3869-1.980</referenceRecordsection>
            <referenceDocumentdescription>A student registration day</
referenceDocumentdescription>
            <referenceDocumentObject> </referenceDocumentObject>
      </srecord>
<Signature xmlns="http://www.w3.org/2000/09/xmldsig#">
   <SignedInfo>
    <CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315"/>
    <SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-sha1"/>
    <Reference URI="">
     <Transforms>
      <Transform Algorithm="http://www.w3.org/2000/09/xmldsig#enveloped-signature"/>
     </Transforms>
     <DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1"/>
     <DigestValue>JDU+BRpitK8ZBN3FNhJ6AOCYEeQ=</DigestValue>
    </Reference>
   </SignedInfo>
   <SignatureValue>ZQLVZYoNCvAEQ7sIgiR1DlnntxEBD+SqYFAH5BsLSRXc8MyGf8UeKyHdjYea5CfY
af+5XiMxgkoydXs0iyhc8vlKosciZF4m05eIP2IikELRFKKiL57LnPtkxvEJzlJG
RaiFjlKKuG9DlXLYxvgKO5kFN8qxm1iLmlGSEOqZK12vaFKxs3Ol90BteHjldeY0
SOg34nrWGFBiwj+0hbioFuWh1g/+RfbM6mjzeiLR58hBVcecyoTOvA/UbqrDL8jl
SFp2/KaPlf3BQh8Uclrd9n1hOKwsq+ThMexh0oiJT3epmAOtl4TDYMgzUwhVEr/s
h55wkUymu0ADB9DkCpt4eeRjo5+iDumSvl1bVxfSxa8DrAnp0g++9QhCrOr3rEiC
gx2Fp1GMd5dpC9OQLK1f2aDR/pXdWa6rywm1o/dhmBLKFtxWEJKGZDxOZQBrQ7++
ltXwzmi7DsFNWysgyzaaSpq3xEJOllLi/SNOAIEvlkZ1rhygV2PV7kGmoCydxWxF
+3y1ykPrfvfhQy9QaaH30CckUWhr1TiDS7CQQaY7lfB464PsiwlN10Ls6Jp/H3Z9
aAAVnO/TCN+GnNjHD2NTteWoxBQvqy07PXVE8vF/A/jvTFpizt3eKgW+TiVCPKYs
RmRY00PHlCbotCeAq10UrCq4QOtx79l5AcsCw91NVSc=</SignatureValue>
   <KeyInfo>
      <KeyName/>
   </KeyInfo>
  </Signature>
</noark5>
```

Figure 5.3: Record Export: simplified record

the records during transfer. In addition, we made changes to objects after they have been archived and the proposed strategy identified the failure in authenticity and integrity of the changed objects. Our experiment shows that records archived using signature validation strategy maintain authenticity and integrity during and after transfer into the archive(see Appendix).

5.4 Summary

This chapter described the experiments and findings of the study. Sample NOARK records are generated and transferred into the archive. The functional requirements of authenticity and integrity and ensuring them is experimented. The result of the study shows that records archived in our

```
<?xml version="1.0"?>
<noark5>
      <brecord>
            <recordID>2011/3869-1</recordID>
            <title>Registration</title>
            <officialTitle>International Students Registration</officialTitle>
            <description>A student registration day</description>
            <keyword>'student' 'International' '2011'</keyword>

            <author>Immigration Office</author>
            <documentmedium>electronic</documentmedium>
            <storagelocation> </storagelocation>
      </brecord>
  <Signature xmlns="http://www.w3.org/2000/09/xmldsig#">
   <SignedInfo>
    <CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xml-c14n-20010315"/>
    <SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-sha1"/>
    <Reference URI="">
     <Transforms>
      <Transform Algorithm="http://www.w3.org/2000/09/xmldsig#enveloped-signature"/>
     </Transforms>
     <DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1"/>
     <DigestValue>AMJSIjdpaNkfQu3bGplEgUvnuss=</DigestValue>
    </Reference>
   </SignedInfo>
   <SignatureValue>lViT3YDK0qj9zuxRXtgVfWOK4PeBUgWZm7uahPg4U61TK0MW4ypJbUJYVMSrZ6es
dsvw7rOGOoADLVBBE2oGnK4BFznaEM+NjdUPapl1NLWaUSL4vL+s7NOnsMacZYy9
i1nxPDboSZeCOsHcpy9Kedyf0v0Z+zek3uyVmNXJCJj1Yd90ViGMUL0wuDns7mNP
ya2Dl3vPJbe1qRnwlVzOXTJkcvE0/Evi3K9hflU4hphCJHxF1gAZAdOE45RGcdcS
UEcZVgFdHaaQJF+P+MAkwQX/8fSTVh8NdoifDagzaSy3cT3zFYTxyK/hm8HBN64J
8rf6WZZ26QeEisY/yaVWhh9FHfpbt15YrR9nCUDzR0u++yv4+SZ+WdGYHt3+waTV
4jtUjvwzbOynVQZevmu7129l6285UojQ6ClQblvU7TufailGlyGZU3ab1Pz9RY/k
awUYMH+lx1HSbHG0ggfvfPW16lhHglGsAHyaqFtkl5d7iLtjypO73jYvBkWyDjFn
8ydAfcuxe7Wl4DS2Apvokc+ntnu8fF+gdybrBgnwri6h9dzVXl8Z/1Pp1qeoskFF
dQ7DtnBYTzCBekMgXQl72Dilay3QearaYC2sLU3aISGBgVkOVBPsLHU3RLR6fpgA
AVkrqNrzkZukdmGuZjLsRsXddW/9SeicRhguHCATBe0=</SignatureValue>
   <KeyInfo>
     <KeyName/>
   </KeyInfo>
  </Signature>
</noark5>
```

Figure 5.4: Record export: basic record

framework are authentic, intact and have evidential value for court cases.

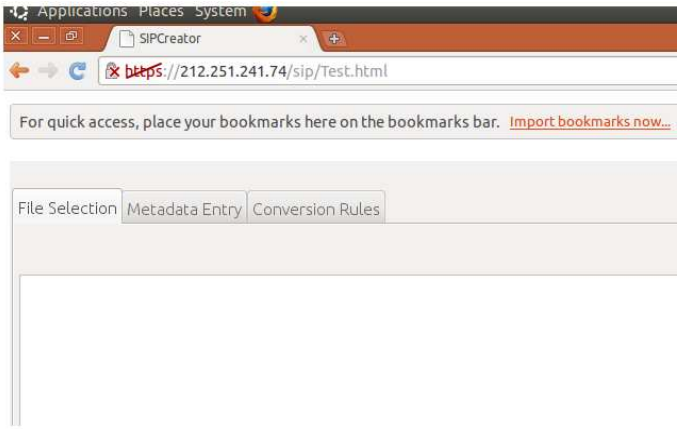

Figure 5.5: SIP request

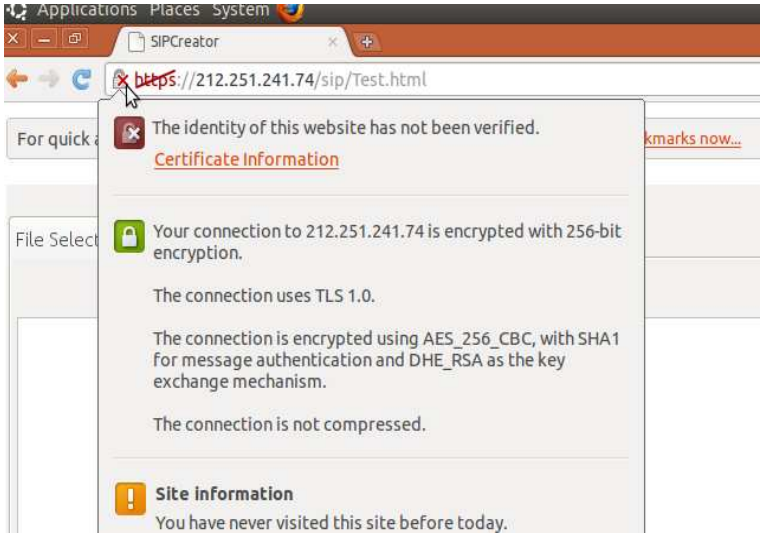

Figure 5.6: SIP authenticity

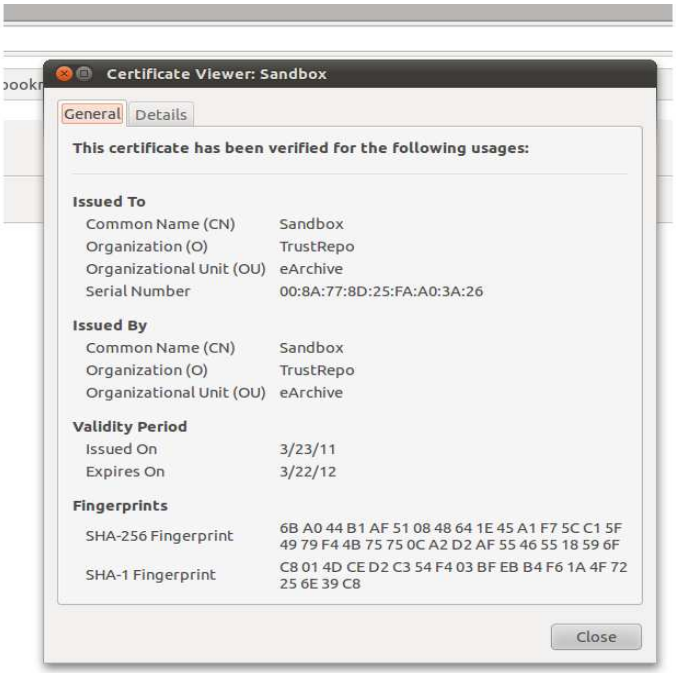

Figure 5.7: SIP certificate

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

The thesis sets out to analyze the functional requirements of trust in electronic records management and archiving using authenticity and integrity as the two elements of trust. To achieve the objectives, a framework is proposed and prototyped for validation. This chapter concludes the thesis by outlining its main contributions and directions for future work.

6.1 Main contributions

Interpretive is the underlying theoretical framework used to undertake this study focusing on the explanatory case study of trust in recordkeeping and archiving. The research has shown how authenticity and integrity can be maintained which in turn increases the evidential value of the objects archived.

The main contribution of the study lies in its attempt to propose an archival framework and apply the concepts of public key infrastructure into the context of trusted archiving using NOARK and Fedora Commons. We have seen how authenticity and integrity can be achieved using digital signatures and how to detect a breach in authenticity and integrity. By focusing on trust in electronic documents, we have also come to a closer understanding of the problems in trusted archiving of digital objects having evidential value. The proposed framework helped in establishing of trust in electronic records archival.

Maintaining trust in records is a very complex matter as records can be in transaction anytime before transferred to the archive. NOARK has a mechanism of maintaining provenance of records by logging the activities and the responsible party initiating the activity in every phase of the trans-

action. In addition, the study narrows down its focus on transfer of records labeled finalized to the archive; and analysis of the functional requirements of authenticity and integrity during transfer to the archive.

The study demonstrated a method of appending digital signature on records and a mechanism of verifying authenticity and integrity during transfer of records to archive.

In the first chapter we asked five main questions repeated below:

1. what are the functional requirements of authenticity for long term archiving of trusted objects?

2. what are the functional requirements of integrity for long term archiving of trusted objects?

3. what are the functional requirements of trusted chain of custody for archiving of trusted objects?

4. how do we ensure authenticity and integrity during the process of transferring records to archive?

5. how do we ensure authenticity and integrity after objects are being archived?

We have shown why addressing these questions is interesting and relevant. We have also answered the questions emphasizing on authenticity, integrity and chain of custody in records archiving.

Table 6.1 summarizes the research questions and the activities done to answer the questions.

Table 6.1: Summary of activities

Question	Pre Ingest	Transfer	Archive
1	Sign objects	Key exchange	Verify signature
2	Object hash	Send hash	Verify hash
3	Trusted third party(TTP)	TTP logs process	TTP verifies process
4	Verify public key	Compare hash	Validate signature
5			Validate signature

The authenticity of a digital object is a verification of a set of accompanying claims associated with the digital object . For a paper based letter, the authenticity of the letter can be verified by expert analysis of a given signers signature. The equivalent for digital objects has been discussed in Section (2.3 and 4.5) where we see how public key cryptography is used to

ensure authenticity as it provides for automated signing and verification of signatures.

Question 1, the functional requirements of authenticity for the long term archiving of trusted objects are with today's technology given through the use of public key cryptography. Public key cryptography can and should be used to sign every digital object to be transferred to the archive. The integrity of a digital object relates to whether or not that object has been altered in anyway during the transfer to the archive. To achieve this the use of hashing functions like MD5 can be employed. In fact these kind of functions are all that is required to determine the integrity of a digital object. Interestingly public key cryptography can also be used to ensure integrity. When a digital object is signed with a public/private key any change to the digital object will result in the failure of a verification process on the digital object. Public key cryptography also opens up for privacy in that it allows the digital content to be encrypted in a manner so only the archive can decrypt the digital object.

Question 2, the functional requirements of integrity for long term archiving of trusted objects are also covered by public key cryptography.

Question 3, the functional requirements of trusted chain of custody for archiving of trusted objects are covered by the presence of a trusted third party. One can not achieve trust unless the integrity and authenticity of the objects are ensured. Trust as discussed in Section 2.2 is something that is open to interpretation, but it is critical. Our study has shown that authenticity and integrity can be built into the system to help the process of trusting an object but it really is a result of the use of public key cryptography to ensure them.

Question 4, to ensure authenticity and integrity during the process of transferring records to archive can be achieved by using the signature validation strategy shown in Figure 4.18. Signature validation strategy allows for the transfer of digital objects over the Internet. No special requirements are needed but using the signature validation strategy to ensure the integrity and authenticity of the digital objects is recommended, especially when records contain sensitive information.

Question 5, to ensure authenticity and integrity after objects are being archived is achieved by using the signature validation strategy mentioned above. The validation strategy ensures the ingested objects maintain their authenticity and integrity. To achieve this digital objects must be stored in a digital library that prevents objects from tampering, accidental or deliberate.

The underlying assumption guiding this study is interpretive as it has the potential to produce deep insight into information systems phenomena including the management of information systems and information systems development which is what this study seeks to realize by constructing mean-

ing of trust in the context of the NOARK-5 and Fedora Commons. Documents, texts, standards and the researchers impression regarding trust gives a qualitative approach to the study. In addition, the use of NOARK-5 and Fedora Commons as context of the study makes the research strategy an explanatory case study. We proposed a framework for archiving records as described in section 4.4. We also developed a prototype as a proof of the concepts in the proposed framework and validated the prototype(see section 3.6 for detail on validation).

6.2 Future work

The study geared towards the submission side of archiving and how to maintain trust in the process of submitting records to archive. An extension of the study may approach the problem from the dissemination side and address trust issues from the users of the archival objects side. The mechanisms of arriving at the decision of trusting or not trusting an electronic content from the users side and functional expectation of users for trusting a content presented to them is an interesting problem which can further be investigated.

In the study emphasis is given from records to archive submission and on how to ensure trust after records are being archived. A natural extension of this could be "what" and "how" to recover authenticity and integrity of objects losing their authenticity and integrity after they are permanently archived.

Dealing with legal matter specially in an electronic communication systems involves a complex matter of handling issues at a national and international context. This might call for coordination and understanding of legal issues in the involved parties which even sometimes worsen the situation as what is legally acceptable in certain boundary is considered not acceptable in another geographical boundary. Mechanisms of addressing such issues could be proposed, coordinated and studied over time.

This study is an explanatory case study of trust in NOARK and Fedora Commons. Studying the applicability of the proposed trust handling in other archival context may be a problem worth pursing for further expansion and validation of the work described above.

References

Allinson, J. (2006, June). *Oais as a reference model for repositories: An evaluation.* Available from http://www.ukoln.ac.uk/repositories/publications/oais-evaluation-200607/Drs-OAIS-evaluation-0.5.pdf

Arne-Kristian Groven, H. A. T. F., Jon Ølnes. (2008). *Preservation of trust in long-term records management systems a state of art overview for the longrec project* (Tech. Rep.). Norwegian Computing Center.

Bearman, D., & Trant, J. (1998, June). *Authenticity of digital resources: Towards a statement of requirements in the research proces.* Available from http://www.dlib.org/dlib/june98/06bearman.html

Buneman, P., Chapman, A., & Cheney, J. (2006). Provenance management in curated databases. In *Proceedings of the 2006 acm sigmod international conference on management of data* (pp. 539–550). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1142473.1142534

CCSDS. (2002, January). *Reference model for an open archival information system (oais).* Available from public.ccsds.org/publications/archive/650x0b1.PDF

Chapman, A. P., Jagadish, H. V., & Ramanan, P. (2008). Efficient provenance storage. In *Proceedings of the 2008 acm sigmod international conference on management of data* (pp. 993–1006). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1376616.1376715

Cheney, J. (2009, October). Workshop on theory and practice of provenance event report. *SIGMOD Rec.*, *38*, 57–60. Available from http://doi.acm.org/10.1145/1815918.1815932

Cheney, J., Chong, S., Foster, N., Seltzer, M., & Vansummeren, S. (2009). Provenance: a future history. In *Proceeding of the 24th acm sigplan conference companion on object oriented programming systems languages and applications* (pp. 957–964). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1639950.1640064

Denning, D. E. (1993). A new paradigm for trusted systems. In *Proceedings on the 1992-1993 workshop on new security paradigms* (pp. 36–41). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/283751.283772

Duranti, L., & MacNeil, H. (1996). The protection of the integrity of electronic records: An overview of the ubc-mas research project. *Archivaria*, *42*, 46-67.

Florence, M. (2010). *Digital preservation in norway's record keeping and archiving traditions: an exploration of authenticity practices using mixed methods research.* Unpublished master's thesis, Oslo University College.

Gladney, H. M. (2004, July). Trustworthy 100-year digital objects: Evidence after every witness is dead. *ACM Trans. Inf. Syst.*, *22*, 406–436. Available from http://doi.acm.org/10.1145/1010614.1010617

Hart, P. E., & Liu, Z. (2003, June). Trust in the preservation of digital information. *Commun. ACM*, *46*, 93–97. Available from http://doi.acm.org/10.1145/777313.777319

Hoeven, H., Albada, J., Información, U. P. G. de, & UNISIST. (1996). *Memory of the world: lost memory, libraries and archives destroyed in the twentieth century.* UNESCO. Available from http://books.google.co.uk/books?id=f0fZGwAACAAJ

Irons, A. (2006). Computer forensics and records management – compatible disciplines. *Records Management Journal*, *16*, 102 - 112.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly, Special Issue on Intensive Research*, *23:1*, 67-93.

Lekkas, D. (2003). Establishing and managing trust within the public key infrastructure. *Computer Communications*, 1815-1825.

Lynch, C. (2000). Authenticity and integrity in the digital environment: An exploratory analysis of the central role of trust. *Authenticity in a Digital Environment*.

Maniatis, P., Roussopoulos, M., Giuli, T. J., Rosenthal, D. S. H., & Baker, M. (2005, February). The lockss peer-to-peer digital preservation system. *ACM Trans. Comput. Syst.*, *23*, 2–50. Available from http://doi.acm.org/10.1145/1047915.1047917

Mel, H., & Baker, D. (2001). *Cryptography decrypted.* Addison-Wesley.

Milakovich, M., & Gordon, G. (2008). *Public administration in america.* Cengage Learning. Available from http://books.google.co.uk/books?id=v4U0dImzX7YC

MoReq. (2009, April). *Noark 5 standard for records management.* Available from http://www.moreq2.eu

Muir, A. (2001). Legal deposit of digital publications: a review of research and development activity. In *Proceedings of the 1st acm/ieee-cs joint conference on digital libraries* (pp. 165–173). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/379437.379475

Myers, M. (2008). *Qualitative research in business management*. SAGE. Available from http://books.google.co.uk/books?id=AMXll6Cd-LAC

Myers, M. D. (2011). Qualitative research in information systems. *MIS Quarterly (21:2), June 1997, pp. 241-242. MISQ Discovery, archival version, June 1997*. Available from www.qual.auckland.ac.nzwww.qual.auckland.ac.nzwww.qual.auckland.ac.nz

Müller, F., Fornaro, P., Rosenthaler, L., & Gschwind, R. (2010, July). Peviar: Digital originals. *J. Comput. Cult. Herit.*, *3*, 2:1–2:12. Available from http://doi.acm.org/10.1145/1805961.1805963

Peiris, H., Soysa, L., & Palliyaguru, R. (2008). Non-repudiation framework for e-government applications. In *Information and automation for sustainability, 2008. iciafs 2008. 4th international conference on* (p. 307 -313).

RLG-OCLC. (2002, May). *Trusted digital repositories: Attributes and responsibilities.* Mountain View, California 94041 USA. Available from www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf

Shaw, G. (2000). Digital document integrity. In *Proceedings of the 2000 acm workshops on multimedia* (pp. 143–144). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/357744.357912

Song, S., & JaJa, J. (2007). New techniques for ensuring the long term integrity of digital archives. In *Proceedings of the 8th annual international conference on digital government research: bridging disciplines & domains* (pp. 57–65). Digital Government Society of North America. Available from http://portal.acm.org/citation.cfm?id=1248460.1248470

TheFedoraDevelopmentTeam. (2008). Introduction to fedora [Computer software manual].

Toma, C. L. (2010). Perceptions of trustworthiness online: the role of visual and textual information. In *Proceedings of the 2010 acm conference on computer supported cooperative work* (pp. 13–22). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1718918.1718923

Tullberg, J. (2008, October). Trust–the importance of trustfulness versus trustworthiness. *The Journal of Socio-Economics*, *37*(5),

2059-2071. Available from http://ideas.repec.org/a/eee/soceco/v37y2008i5p2059-2071.html

Wang, C., & Wulf, W. (1999, aug). Towards a scalable pki for electronic commerce systems. In *Advance issues of e-commerce and web-based information systems, wecwis, 1999. international conference on* (p. 132 -136).

Waugh, A., Wilkinson, R., Hills, B., & Dell'oro, J. (2000). Preserving digital information forever. In *Proceedings of the fifth acm conference on digital libraries* (pp. 175–184). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/336597.336659

Wing, P., & O'Higgins, B. (1999, sep). Using public-key infrastructures for security and risk management. *Communications Magazine, IEEE*, *37*(9), 71 -73.

Appendices

```
import java.sql.*;

public class CreateTable {
	public static void main(String[] args){
		System.out.println("Creating NOARK Tables!");
		Connection con = null;
		String dbname = "noark";
		try{
			String url = "jdbc:mysql://localhost/";
			String user = "noark";  //database username
			String pass = "sandbox"; // database password
			Class.forName("com.mysql.jdbc.Driver");//.newInstance();
			con = DriverManager.getConnection(url+dbname, user, pass);
			System.out.println("Connection Established");
			Statement stmt = con.createStatement();
		   stmt.executeUpdate(
					"CREATE TABLE srecord("
				+"systemID INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,"
				+"recordtype VARCHAR(150)  , createdDate TIMESTAMP DEFAULT NOW(),"
				+"createdBy VARCHAR(150), archivedDate DATETIME  , archivedBy VARCHAR(150),"
				+"referenceparent VARCHAR(150),  referenceDocumentDescription VARCHAR(150)  ,"
				+"referenceRecordSection VARCHAR(150), referenceDocumentObject VARCHAR(200) )");
			System.out.println("Table created successfully");
		}
		catch (Exception e)
		{
			System.err.println("Unable to connect to the database");
		}
		finally
		{
			if(con!=null)
			{
				try
				{
					con.close();
					System.out.println("Database connection terminated");
				}
				catch (Exception e){}
			}
		}
	}

}
```

Figure 1: Sample NOARK Table Generator

```
import java.sql.*;

public class CreateTable {
	public static void main(String[] args){
		System.out.println("Creating NOARK Tables!");
		Connection con = null;
		String dbname = "noark";
		try{
			String url = "jdbc:mysql://localhost/";
			String user = "noark";  //database username
			String pass = "sandbox"; // database password
			Class.forName("com.mysql.jdbc.Driver");//.newInstance();
			con = DriverManager.getConnection(url+dbname, user, pass);
			System.out.println("Connection Established");
			Statement stmt = con.createStatement();
		   stmt.executeUpdate(
					"CREATE TABLE srecord("
				+"systemID INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,"
				+"recordtype VARCHAR(150)  , createdDate TIMESTAMP DEFAULT NOW(),"
				+"createdBy VARCHAR(150), archivedDate DATETIME  , archivedBy VARCHAR(150),"
				+"referenceparent VARCHAR(150),  referenceDocumentDescription VARCHAR(150)  ,"
				+"referenceRecordSection VARCHAR(150), referenceDocumentObject VARCHAR(200) )");
			System.out.println("Table created successfully");
		}
		catch (Exception e)
		{
			System.err.println("Unable to connect to the database");
		}
		finally
		{
			if(con!=null)
			{
				try
				{
					con.close();
					System.out.println("Database connection terminated");
				}
				catch (Exception e){}
			}
		}
	}
```

Figure 2: Sample NOARK srecord Generator

```
import java.io.StringWriter;
import java.io.*;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.ResultSetMetaData;

import javax.xml.parsers.DocumentBuilder;
import javax.xml.parsers.DocumentBuilderFactory;
import javax.xml.transform.OutputKeys;
import javax.xml.transform.Transformer;
import javax.xml.transform.TransformerFactory;
import javax.xml.transform.dom.DOMSource;
import javax.xml.transform.stream.StreamResult;

import org.w3c.dom.Document;
import org.w3c.dom.Element;

public class MakeXMLFile {
      public static void main(String args[]) throws Exception{

            PrintStream printScreen    = null;
            PrintStream fileStream     = null;
            printScreen = System.out;
            fileStream = new PrintStream(new FileOutputStream("srecord.xml"));
            System.setOut(fileStream);

            String user = "noark";  //database username
            String pass = "sandbox"; // database password
            String dbname = "noark";
            String url = "jdbc:mysql://localhost/";
            DocumentBuilderFactory factory = DocumentBuilderFactory.newInstance();
            DocumentBuilder builder = factory.newDocumentBuilder();
            Document doc = builder.newDocument();
            Element results = doc.createElement("noark5");
            doc.appendChild(results);

            Class.forName("com.mysql.jdbc.Driver");
            Connection con = DriverManager.getConnection(url+dbname, user, pass);
            ResultSet rs = con.createStatement().executeQuery("select * from srecord");
            ResultSetMetaData rsmd = rs.getMetaData();
            int colCount = rsmd.getColumnCount();

            while (rs.next()) {
                Element row = doc.createElement("srecord");
                results.appendChild(row);
                for (int i = 1; i <= colCount; i++) {
                  String columnName = rsmd.getColumnName(i);
                  Object value = rs.getObject(i);
                  Element node = doc.createElement(columnName);
                  node.appendChild(doc.createTextNode(value.toString()));
                  row.appendChild(node);
                 }
               }

            DOMSource domSource = new DOMSource(doc);
          TransformerFactory tf = TransformerFactory.newInstance();
          Transformer transformer = tf.newTransformer();
          transformer.setOutputProperty(OutputKeys.OMIT_XML_DECLARATION, "no");
          transformer.setOutputProperty(OutputKeys.METHOD, "xml");
          transformer.setOutputProperty(OutputKeys.ENCODING, "ISO-8859-1");
          StringWriter sw = new StringWriter();
          StreamResult sr = new StreamResult(sw);
          transformer.transform(domSource, sr);

          System.out.println(sw.toString());
          System.setOut(printScreen);
          System.out.println(sw.toString());

          con.close();
          rs.close();
      }
```

Figure 3: Sample XML extractor

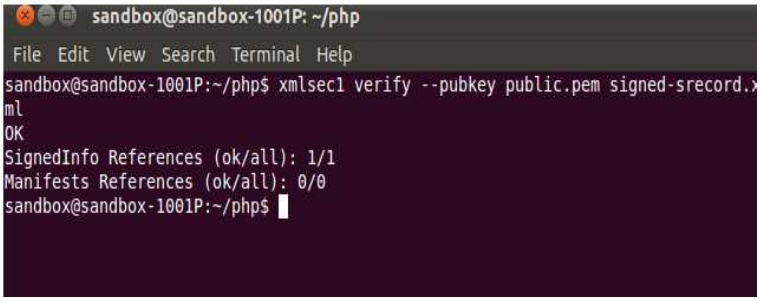

Figure 4: verifying the record export: accepted integrity

Figure 5: verifying the record export: failed integrity

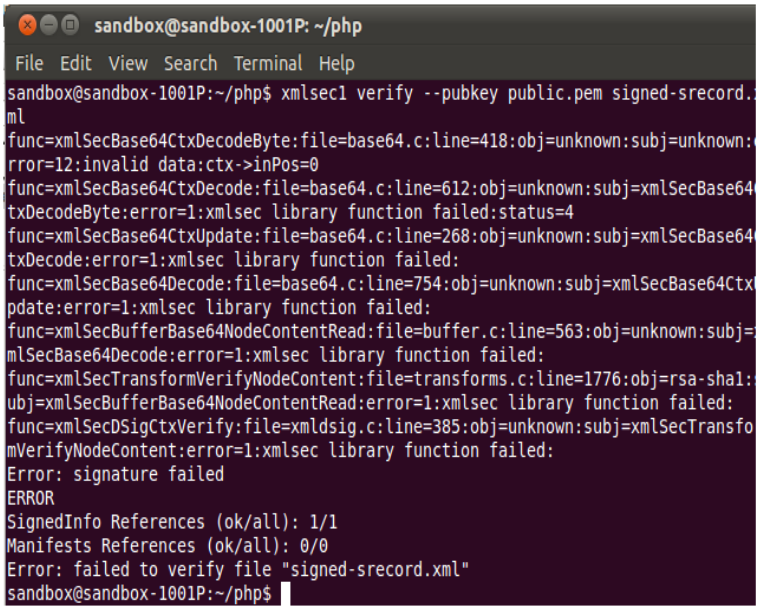

Figure 6: Signature failure: authenticity of the recordkeeping system