Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Report

Trust Strategies in Long-term Management and Preservation of Digital Records

A Deliverable to the LongRec Research Project

Report no	1025
Authors	Arne-Kristian Groven
Date	2010-03-26
ISBN	978-82-539-0535-8

About the authors

Arne-Kristian Groven works as a senior research scientist at Norsk Regnesentral, where he has been working since 1996.

He is educated at the University of Oslo, Department of informatics, where he also worked as a research assistant for a shorter period of time. In addition he has been doing research at the Institute for Energy Technology, Halden Reactor Project.

Norsk Regnesentral

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses.
A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Trust Strategies in Long-term Management and Preservation of Digital Records**
Authors	**Arne-Kristian Groven**
Quality assurance	Trond Sirevåg (Riksarkivet/The National Archival Services of Norway), Inger-Mette Gustavsen (DNV)
Date	26. March
Year	2010
ISBN	978-82-539-0535-8
Publication number	1025

Abstract

Strategies for trust (enhancement) cover a broad field. This article has been investigating trust strategies addressing (collections of) digital records throughout the different phases of their life-cycle, from their creation as digital (business) records, their use and maintenance as part of electronic records management, followed by archiving when no longer useful in business. The long-term perspective of the archiving is an assumption in the discussion. The long-term perspective is also in some cases dominant when records are active, whenever it is required that records are active in business for several decades.

Different types of trust strategies are discussed, including: (A) Extensive descriptions and annotations, based on diplomatic analysis. (B) Documentation and storage of the provenance of the object, like the documentation of the chain of custody. (C) Relying on signatures and seals that are attached to the object or the claims that come with it. (D) Comparing the object in hand with other versions (copies) of the object that may be available. Since field is still immature, combinations of these strategies should be applied. Behind all these strategies are a set of security controls that also should be in place.

Keywords	Trust, Long-term Records Management, Long-term Preservation, Digital Signatures, Provenance, Diplomatics, Security
Target group	Project members, Others
Availability	Public
Project number	320369
Research field	Security
Number of pages	57
© Copyright	Norsk Regnesentral

Contents

List of figures

1 PROJECT BACKGROUND

This report is produced as a contribution to the LongRec (Long-Term Records Management) project[1] headed by Det Norske Veritas (DNV) in collaboration with a number of case partners, commercialization partners and research partners. The primary objective of LongRec is the persistent, reliable and trustworthy long-term archival of digital information records with emphasis on availability and use of the information.

LongRec is a three year project (2007-2009) partly funded by the Norwegian Research Council. The project constitutes the Norwegian team of the InterPARES 3 project[2],

LongRec addresses several research challenges, each of which is assigned a short name (in parentheses below): records transition survival (READ), long-term usage (FIND), preservation of semantic value (UNDERSTAND), preservation of evidential value (TRUST) and legal, social, and cultural framework (COMPLIANCE). Each research challenge is addressed by:

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

[1] The project's public web site is at http://research.dnv.com/longrec/

[2] The project's public web site is at http://www.interpares.org

2 INTRODUCTION

2.1 Electronic Records and Trust

In the case of electronic records management and long-term archiving of digital records, trust is related to whether we believe in the digital records presented to us, often years after they were created. Compiling available information, weighted by common sense and a sound scepticism towards the information, into rational trust decisions is a difficult task. To be presented some content in some presentation form (layout) and nothing else, years after the time of its creation is definitely not enough!

Whether something or someone are trusted or not is based on subjectivity. You either trust or distrust. A trust decision is always ultimately binary (trust or not) but the decision process is based on both knowledge and assumptions about the situation in case, i.e. unless one has complete knowledge about the situation, there is always a degree of uncertainty in the process. According to [31] trust can be defined as “perceived lack of vulnerability”. A trust decision implies a (human) judgment about the vulnerability implied by a certain action. [22] separates trust decisions into trusting “rational entities” that behave according to (programmed) instructions and logic, and on the other side “passionate entities” which may behave according to will.

In order gain trust, the digital records must be worthy of trust. They have to be trustworthy, with respect to the properties like *authenticity*. Authenticity is defined by the InterPARES research project [21] as: “The trustworthiness of a record as a record; i.e., the quality of a record that is what it purports to be and that is free from tampering or corruption. Authentic records are records that have maintained their identity and integrity over time.” This quality is attributed to an original or a true and faithful derivation of the original record. *Integrity* is the quality of being whole, complete and unaltered in all essential respects [21] [29]. Trustworthiness can also cover other qualities than authenticity and integrity, such as *reliability* (precision), i.e. consistency in outcome/behaviour (in our case consistent and complete creation of records), and *accuracy* (validity), i.e. true/correct reflections of the world.

In electronic records management, the digital material is the documentation of the transactions taking place between two or more parties, from the perspective of one of the participants like a company or a public office. The transactions and correspondence might include digital documents or paper documents being scanned into the records management system. Further it might include e-mails, audio or video, digitally created or scanned pictures, digitally created or scanned maps, drawings etc.

The time of creation and maintenance of digital records usually takes days, weeks, months, years, or in some rare cases even decades. The contributors are often many, both internal and external ones. The digital records are grouped and classified according to certain classification schemes, into fonds, subfonds, files etc. Records are interrelated according to classified themes where every new transaction related to a specific theme should be recorded and linked to already existing records of that theme. New descriptions and annotations to a record will most certainly appear during its life-time. Transformation (conversion) of bit streams from one data format to another is also needed in the long run, due to obsolescence over time of the data format in which the original digital material was represented. Also, systems may fail causing (partly) loss of information. All this complexity implies the risk of loosing trust.

2.2 The chosen Approach to Trust Strategies

The main focus of this document is to describe and discuss trust strategies, or more precisely rephrased, trust *enhancing* strategies that can be used in long-term records management to keep digital records trustworthy throughout their life-cycle. Taking a closer look at the (classical) life-cycle of digital records it consists of:

- An active phase;
- A passive phase;
- One or more transformations.

Trust strategies will in the following be discussed related to these three phases. In their active phase, also called the creation and maintenance phase, the records "live" inside a records management system, serving a business purpose. In the passive phase, also called the archival phase, the records "live" within an archive. The main transformation phase refers to the change into an archival representation of records from whatever pre-existing representation of records in some data format. One example of such is transformation from relational database structures into archival XML structures. In the OAIS terminology [27] the transformation phase includes construction of Submission Information Packages, SIPs, at the site of the records creator, the transmission of SIPs, and the SIP validation as part of the Ingest at the site of the archive.

Related to the different phases just explained, there are strategic options in every phase on how to enhance trust. Since evidence of authenticity, integrity etc. is needed in all phases of a record's life cycle, such evidence have to be collected and handled with care right from the time of record creation. There are only a few fundamental approaches for examining the authenticity of a digital object according to [24], and these are:

1. We examine the provenance of the object (for example, the documentation of the chain of custody) and the extent to which we trust and believe this documentation as well as the extent to which we trust the custodians themselves.

2. We perform a forensic and diplomatic examination of the object (both its content and its artifactual form) to ensure that its characteristics and content are consistent with the claims made about it and the record of its provenance.

3. We rely on signatures and seals that are attached to the object or the claims that come with it, or both, and evaluate their forensics and diplomatics and their consistency with claims and provenance.

4. For mass-produced and distributed (i.e., published) objects, we compare the object in hand with other versions (copies) of the object that may be available (which, in turn, means also assessing the integrity and provenance of these other versions or copies).

Much of the same view is reflected in [17], stating that the assessment of authenticity has historically always formed part of the traditional archival appraisal process:

- In the first instance, it has relied on confirming the existence of an unbroken chain of custody from the time of the records' creation to their transfer to the archival entity responsible for their long-term preservation. (*Provenance-based*)

- The assessment of authenticity has also depended on the archivist's knowledge of recordkeeping practices, both historically and in relation to the record types and administrative procedures of a specific creator. The general framework for this assessment was originally codified in diplomatics. (*Diplomatics-based*)

- A third, less frequently used method to confirm the identity and integrity of records is based on comparison. Records within a fonds are compared to copies forwarded to and held by external sources in the normal course of the creator's business. (*Redundancy based*)

Trust strategies are therefore in the following discussed in relation to their ability to support the fundamental approaches just mentioned. Before going further, these fundamentals will be briefly explained.

3 FUNDAMENTAL APPROACHES FOR ESTABLISHMENT OF TRUST

3.1 Digital Signatures

Digital signatures are used on the Internet to detect forgery and tampering, e.g. in financial transactions like withdrawal from your bank account using online/Internet banking. A digital signature is a mathematical scheme, a cryptographic tool for authenticating the source of the message and at the same time checking the integrity. It is usually based in challenges using private and public key pairs. A valid digital signature gives a recipient reason to believe that the message was created by a known sender and that it was not altered in transit.

Different purposes for the use of digital signatures exist:

- Identification, by creating a link between the document and the name of the signer (authentication);
- Authorization and data integrity; the signature implies that the signer accepts the content of the document or gives it a certain authority;
- Evidence, where a signed document provides a stronger proof of authenticity and integrity than a document without a signature;
- Symbolism, e.g. signing as a part of some ceremony;
- Fulfilment; denoting the end of a negotiation process

In the context of records management and long-term archival storage, digitally signed documents/material and also signed user sessions will increase the overall trustworthiness. Digital signatures can be used to demonstrate authenticity and bit integrity of a digital message or document, the latter by verifying the received bits using a cryptographic hash functions/check sums. This to check if the transmitted bits have not been corrupted.

In addition, authentication mechanisms based on public key cryptography offer better evidence than other mechanisms when it comes to controlling access to records management systems.

3.2 Redundancy

Redundancy is widely used in safety-critical engineering to increase reliability and fault-tolerance. In many safety-critical systems some parts of the control system may be triplicated. A fault caused by an error in one component may then be out-voted by the other two.

Redundancy is also widely used when talking about computer data storage. Data redundancy is a property of some disk arrays, most commonly in RAID systems, which provides fault tolerance so that all or part of the data stored in the array can be recovered in the case of disk failure.

To see how redundancy is used for digital preservation, LOCKSS (Lots of Copies Keep Stuff Safe), based at Stanford University Libraries can serve as an example[3].

During its operation, LOCKSS software compares the copies stored in (different) libraries' LOCKSS boxes to the content available on the publisher's website to establish the content's authoritative version. This version is used to repair lost or corrupted copies.

The LOCKSS Boxes then communicate over the Internet to continually audit the content they are preserving. If the content in one LOCKSS Box is damaged or incomplete, that LOCKSS Box will receives repairs from the content based on other LOCKSS Boxes in the distributed network of LOCKSS Boxes.

This cooperation between the LOCKSS Boxes avoids the need to back them up individually. It also provides unambiguous reassurance that the system is performing its function and that the correct content will be available to readers when they try to access it.

3.3 Establishing of Provenance

The word provenance means "to come from" in French (“provenir”). That is the origin, or the source, of something, but also the history of ownership or location of an object. The term was originally used in the field of art, but is now used in similar senses in a wide range of fields.

Provenance is a fundamental principle of archives, referring to the individual, group, or organization that created or received the items in a collection. According to archival theory and the principle of provenance, records of different provenance should be separated. Historically, in archival practice, proof of provenance is provided by the operation of control systems that document the history of records kept in archives, including details of amendments and annotations made to them.

In most fields the primary purpose of provenance is to confirm or gather evidence as to the time, place, and if appropriate the person responsible, for the creation, production or discovery of the object, but this will typically be accomplished by tracing the whole history of the object up to the present. Comparative techniques, expert opinions, and the results of various kinds of scientific tests may also be used to these ends, but establishing provenance is essentially a matter of documentation.

3.4 Diplomatic Analysis

A methodology for analysis of documentary forms aimed at understanding the administrative actions and generating functions of documents was established several hundred years ago. The methodology is called diplomatic analysis, or simply diplomatics. The same methodology has since been used for analysing digital records[4].

Diplomatics rests on the assumption that despite differences in nature, provenance or date, all documents present forms similar enough to make it possible to conceive one typical, ideal documentary form, the most regular and complete, for the purpose of examining all its elements.' Once the elements of this ideal form have been analysed and their specific function identified, their variations and presence or absence in existing documentary forms will reveal the administrative function of the documents manifesting those forms [7].

[3] The text presented in the rest of this subsection can be found at: http://www.lockss.org/lockss/How_It_Works

[4] In the InterPARES reseach project.

Diplomatics defines form as the set of rules of representation used to convey a message, that is, as the characteristics of a document which can be separated from the determination of the particular subjects, persons or places which it concerns. Documentary form is both physical (extrinsic), talking about the external make-up, and intellectual (intrinsic) talking about articulation.

According to [18] there are five necessary characteristics to determine for a digital object to be considered being a record. To be considered a record, a digital object must:

- possess a fixed form and stable content affixed to a stable medium;
- participate in an action;
- possess an archival bond, which is the relationship that links each record to the previous and subsequent record of the same action;
- involve at least three persons: the author, addressee and writer; in the digital environment, there are two more necessary persons: the creator and the originator; and
- possess an identifiable context (i.e., the framework in which the action in which the record participates takes place), including juridical-administrative, provenancial, procedural, documentary and technological contexts.

Non-records generally require a simpler preservation model because they exist autonomously from other documents and their purpose is, typically, limited to dissemination of information.

3.5 Forensic Analysis

Digital forensics is a branch of forensic science pertaining to legal evidence found in computers and digital storage media.

The goal of digital forensics is to explain the current state of a digital artefact. The term digital artefact can include a computer system, a storage medium (such as a hard disk or CD-ROM), an electronic document (e.g. an e-mail message or JPEG image), or even a sequence of packets moving over a computer network. The explanation can be as straightforward as "what information is here?" and as detailed as "what is the sequence of events responsible for the present situation?"

The field of digital forensics also has sub branches such as database forensics.

4 MANAGEMENT OF ACTIVE RECORDS

4.1 Records Management, a short Overview

The business activities going on in the private or public sector consist of different types of transactions. Digital documents (material), can serve as evidence of the transactions if they are properly captured, used and maintained as digital records.

The capture process is illustrated in Figure 1. Here records are created, based on the arrival of one or possibly several digital documents (material) from a source external to the organization creating the record or, alternatively, based on the internal production of such material. Various annotations (descriptions/metadata) are then added, or could be added at later stages. This information is either extracted from the digital documents themselves or from the environment in which the record was created.

To exemplify, a request of some kind is sent as e-mail to an organization. Based on procedures anchored in rules and regulations, the request can typically be put under records management. As a consequence various metadata are registered, like time and date of the registration and the person responsible for doing so. Further, the source of the e-mail and the person responsible for following up the request are other associated metadata together with many more. Another example might be the registration of a technical drawing. Metadata in this case will be time/date of arrival, who registered the drawing, the producer of the drawing, the person be responsible for following up the received drawing, and so on.

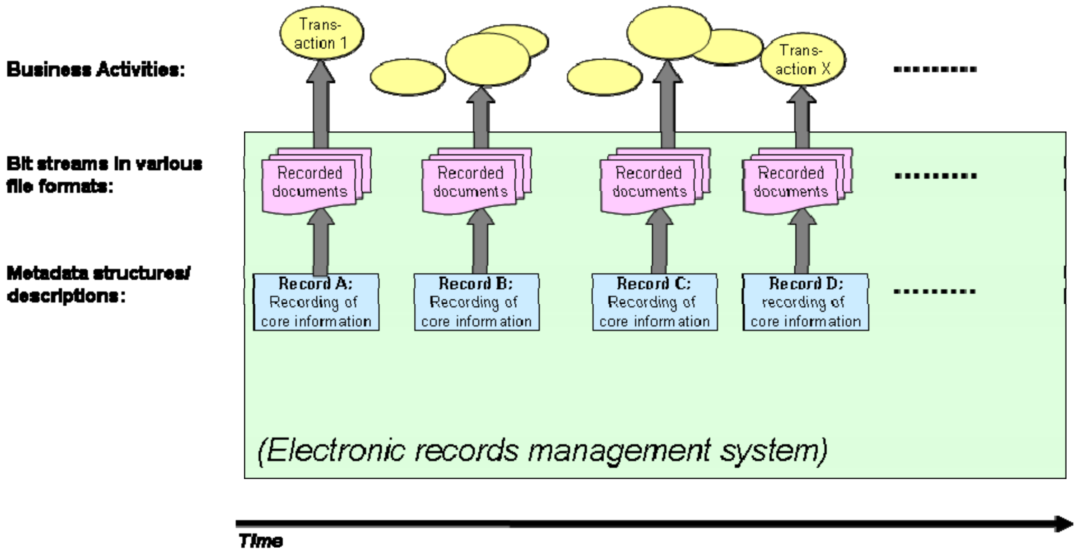
Business Activities:
Bit streams in various file formats:
Metadata structures/ descriptions:
Trans- action 1
Trans- action X
Recorded documents
Recorded documents
Recorded documents
Recorded documents
Record A: Recording of core information
Record B: Recording of core information
Record C: Recording of core information
Record D: recording of core information
(Electronic records management system)
Time

Figure 1 Capturing of records along the time line

Records management principles and automated records management systems are aids in the capture, classification, and ongoing management of records. Electronic records management systems typically uses databases for storage of records. Capturing the record within the electronic environment include interfaces between the records management system and applications such as word processors or e-mail clients, which are used to create or receive records. Systematic capture requires both a technical interface and a set of rules or procedures which govern its behaviour and successful application within the organization.

Records co-exist and interact with other records, solving common tasks together. Until the overall task is completely solved, the record will remain active. The active period of a records life-cycle is the period where it serves a function and is actively used within an organization. There is a huge difference between controlled records management, using dedicated records management systems, and documents floating around on servers and hard disks, documents that ideally should have been records. A main trait of records as opposed to pure documents or published information is the documentation of context information from various domains, in addition to the digital documents themselves.

4.2 Long-Term Perspectives

Having served their time as active records, it is either time for record destruction or for record archiving. The latter, due to juridical or other requirements. The period of archiving can be decades or centuries. In the archival phase, one challenge is to provide evidence of origin and authenticity of archived records during the whole archival period. Another challenge is to make records accessible for humans during the archival period.

Usually records are active for less than ten years. But in some cases records are active for decades, up to 50 years or more. Examples of such are found in [19]. This is specifically challenging since the records have to be active and full-functioning on the one hand and in addition they are caught by the aging problem which is the main focus of archival environments. In such cases, there will be a trade-off on whether to include archival functionality into the record keeping environment or, alternatively, to deposit the long-lived records into an archive multiple time over the years. One fact that has to be handled in such cases is the loss of bit integrity of the recorded document/material. Therefore preservation strategies have to be integrated, one way or another.

4.3 Standardization

ISO[5] defines records as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business". The International Council on Archives (ICA)[6] Committee on Electronic Records defines a record as "recorded information produced or received in the initiation, conduct or completion of an institutional or individual activity and that comprises content, context and structure sufficient to provide evidence of the activity." The key word in these definitions is evidence. Put simply, a record can be defined as "evidence of an event".

The ISO 15489: 2001 standard defines records management as "The field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including the processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records". ISO 15489:2001 states that records management includes:

- setting policies and standards;
- assigning responsibilities and authorities;

[5] See http://www.iso.org

[6] See http://www.ica.org

- establishing and promulgating procedures and guidelines;
- providing a range of services relating to the management and use of records;
- designing, implementing and administering specialized systems for managing records; and
- integrating records management into business systems and processes;

Public record keeping in Norway has to comply with a standard called Noark, an acronym for "Norsk arkivsystem", or Norwegian recordkeeping/archival system. For the time being the fourth revision of the standard, Noark-4 from 1999[7], is about to be succeeded by the fifth revision Noark-5[8].

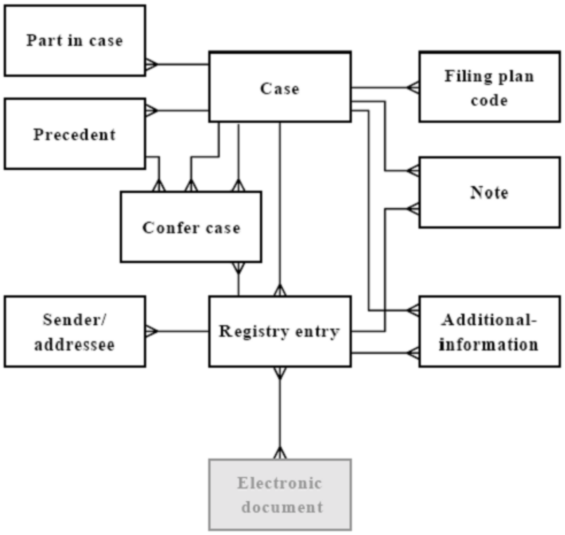

Figure 2 Simplified data model for the records management module in Noark-4

Noark-4 is a specification of functional requirements for electronic recordkeeping systems used in public administration in Norway. The specification lists requirements with regard to information content (what kind of information it should be possible to register and retrieve), data structure (design of each data element and the relationship between these elements) and functionality (the functions which the systems are to maintain). In Figure 2, part of the specified data model of Norak-4 presented. All boxes denote database tables associated with a set of requirements. Examples of such requirements are:

- It should be possible to register a document received or produced by an organization in a registry entry. As a minimum, it should be possible to register information which is obligatory in the table Registry entry.
- When created, a registry entry should always be associated with a case (file). One or more registry entries may be associated with a case (file).

[7] See http://www.arkivverket.no/arkivverket/lover/elarkiv/noark-4/english.html

[8] See http://www.arkivverket.no/arkivverket/Offentlig-forvaltning/Noark/Noark-5 (in Norwegian)

- As common information on a case (file), it should as a minimum be possible to register the information which is obligatory in the table Case.

Where Noark-4 described database models, containing specified elements and tables, Noark-5 is specifying XML-structures, leaving the representation to the records management system developer.

The European Moreq2 standard[9] is comparable to the Norwegian Noark-5 standard. Moreq is an acronym for “Model Requirements for the Management of Electronic Records.

[9] See http://www.moreq2.eu/

5 LONG-TERM PRESERVATION OF ARCHIVED RECORDS

5.1 Archiving, an Overview

In general, archives consist of records that have been selected for permanent or long-term preservation on grounds of their enduring cultural, historical, or evidentiary value.

Records, traditionally, go through an appraisal process prior to entering the archive. Appraisal is the process of distinguishing records of continuing value from those of no further value so that the latter may be eliminated. Effective appraisal, especially in the digital environment, is dependent on good systems of records creation and business scheduling of records.

After a while records loose their active business role for the record creator and become passive. Once the decisions are made regarding what type of records can enter an archive, these records are moved from the creating organization to an archive according to specified procedures. They are transformed into specified data formats probably differing from their data format. Next, archival descriptions are made. Traditional archival descriptions explain the nature and scope of the extant holdings, particularly for the benefit of users outside of the creating organization, who need to understand the organizational context of the materials to make best use of them. This access-oriented vision of description is confirmed by the international descriptive standard ISAD(G)[10], which argues that "the purpose of archival description is to identify and explain the context and content of archival material in order to promote its future accessibility."

In the public sector of different countries, national archives like the National Archives[11] in the UK or the National Archival Services of Norway[12] provide archival services. They are the main actors involved when it comes to defining national public archival policies. Such policies are, e.g., available in the area of acquisition and appraisal, defining what is worth archiving from public records management and what is not[13]. For private businesses, on the other hand, these questions might not be clarified, and the difference between the in-house archive and the in-house records management system might be fuzzy.

An international network of archives exist, called The International Council on Archives (ICA)[6]. It is an international non-governmental organization which exists to promote international cooperation in archiving having about 1,400 institutional members in 190 countries.

5.2 The OAIS Standard

Standardization provides both norms and transparency concerning how to organize archival objects, systems, and workflows. One of the most influential archival standards existing today is the ISO reference model for Open Archival Information System (OAIS), ISO 14721:2003.

Three different information package types are defined in OAIS [27] and these are:

- SIP - Submission Information Package

[10] See http://www.ica.org/en/node/30000

[11] See http://www.nationalarchives.gov.uk/

[12] See http://riksarkivet.no/

[13] Like the Norwegian Noark standard.

- AIP – Archival Information Package
- DIP – Dissemination Information Package

The OAIS reference model further defines six areas of concern and their interrelationship, as shown in Figure 3. The areas of concern are: 1) Ingest, the process that accepts submissions from producers and transforms this into a representation (AIP) suitable for the repository, 2) Data Management, 3) Archival Storage, 4) Administration, 5) Preservation Planning, and 6) Access.

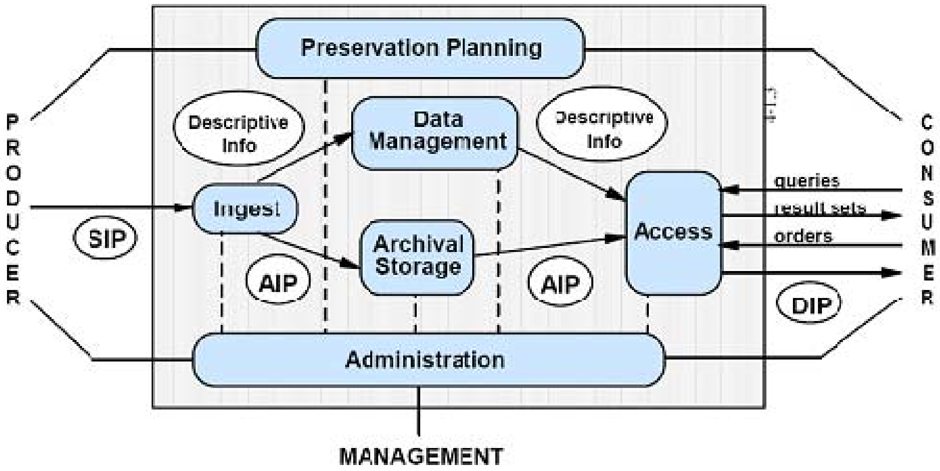

Figure 3 An Overview of OAIS

According to OAIS an AIP can be of two subtypes, either i) AIU, an Archival Information Unit, or ii) AIC, an Archival Information Collection. The AIUs are the atomic units of an archive and the AICs are aggregations of multiple AIUs (AIPs), where aggregation criteria are determined by the archival organization.

Each AIU contains exactly one Content Information object, which in turn may consist of multiple files, and exactly one set of Preservation Description Information, PDI. This can be illustrated in the following:

- Content Information (CI):
  - the digital encoding (bit stream) of , e.g., a movie in a specific format
  - the Representation Information needed to understand this format
- Preservation Description Information (PDI):
  - e.g., date of creation, featured actors, movie studio, etc, and a checksum for integrity.

Each AIC is itself a complete AIP containing PDI that e.g. describe date and motivation for creation, context to other AICs, level of security etc. A single AIP might belong to multiple AICs.

Each PDI is classified as follows:

- provenance: custody, history, processing history
- context: why the CI was produced, how it relates to other CI objects
- reference: reference code or ISBN
- fixity: a checksum, digital signature, authenticity indicator or similar

OAIS specifies the overall structure of archival packages, focusing on the AIPs, and most of the standard describes how to generate, access, and manipulate AIPs in different work processes. The SIPs and the DIPs are given minor roles in the standard, compared to the AIPs.

5.3 Metadata Standards

Metadata schemas (standards) of relevance include;

- the PREMIS[14] dictionary for preservation metadata; and
- the Library of Congress's Metadata Encoding and Transmission Standard (METS)[15].

An extensive list of metadata standards can be found at Hardvard University Library, Office for Information Systems[16].

One or only a few unified metadata schemes are probably unlikely, as stated in the following: "Much can be said about metadata. But for the time being, any set of metadata should identify its own schema because settling on a single worldwide scheme seems unlikely in the foreseeable future [6]."

5.4 Digital Preservation Strategies

Long-term preservation of digital artefacts is often described using the metaphor "transmission in time" as an analogy to the "transmission in space" that goes on in telecommunication and on the Internet. But for "transmissions in time", if the time span is long enough, it is impossible to base trust on bit integrity. It is impossible because at some point in time the original software and hardware will no longer be available for interpreting the original bit stream.

The original bit streams/strings might still be available, decades and centuries from now, but the messages they represent will eventually not be readable (accessible) in their original environment any more. So in the long-term scenario, one of the main goals is to avoid loosing readability, when software and hardware environments are lost. This can be achieved by performing continuous monitoring and risk assessment of data formats and then by applying

[14] http://www.loc.gov/standards/premis/

[15] http://www.loc.gov/standards/mets/

[16] http://hul.harvard.edu/ois/digproj/metadata-standards.html

suitable preservation strategies when necessary to avoid loosing readability. Below, in Table 1, are listed the different types of preservation strategies available today[17].

	Computer Museum	Emulation	Encapsulation	UVC	Batch Conversion	Conversion on Access
Change in bits	no change	no change	no change	change once	change many times	change once
Hardware specification	no	yes	no	no	no	no
Hardware components	yes	no	no	no	no	no
Format specification	n/a	n/a	the original format's specifications	the mediatory format's specifications	the latest format's specifications	the original format's specifications
Format converter	n/a	n/a	n/a	from the original format to a mediatory format	from the format currently used to a new format	from the original format to a new format
Format interpreter	the original interpreter	the original interpreter	the new interpreters	the new interpreter	the current interpreter	the current interpreter

Table 1 Overview of the different digital preservation strategies.

Included here is the so called museum strategy, that is keeping the old files, the software accessing the files, and the hardware until it breaks down. Encapsulation means in most cases XML-encapsulations, of digital documents/material and all the associated metadata. We take a closer look at the other strategies:

- Emulation: Keep original data formats unchanged but develop and maintain emulation software to process and interpret these formats on top of new platforms

- Virtual machines, e.g. UVC: Addressing the main criticism of the emulation strategy[18]. Involves one conversion from the original data format to a virtual machine, the Univeral Virtual Computer, with a limited set of instructions. Then acts as an emulated platform

[17] Source: Longrec project, the READ activity.

[18] The main argument against the emulation strategy is the risk of making errors when when developing an emulator, due to the complexity of modern hardware and software platforms.

- Conversion/transformation[19]: Convert the a digital artefact/digital document to a new data format when regarded necessary in order to be able keep readability and to discard old technologies[20]

There is a recursive nature in all these three alternatives above, using the conversion strategy as an example: Starting with the original document represented as a bit stream in an original data format, the result of a conversion is a derived copy of the original document represented in a new data format. Sometimes in the future this new data format might also be obsolescent. Therefore yet a new conversion has to take place.

The same applies for emulations, software/hardware platforms becomes obsolescent after a while and have to be substituted by yet a new platform. Hence, new emulator back-ends have to be developed.

5.5 TRAC

There also exist quality management initiatives addressing trustworthiness of archives. The most well-known developed criteria for establishment of trust in digital repositories, resulting in the publication of a RLG-OCLC report entitled *Trusted Digital Repositories: Attributes and Responsibilities* [28]. The report defined: the characteristics of a trusted digital repository; listed relevant attributes of such a repository; called for compliance with the OAIS as well as administrative responsibility, organizational viability, financial sustainability, technological and procedural suitability, system security, and procedural accountability. It also recommended that a process should be developed for the certification of digital repositories. A new document, version 1.0 of the *Trustworthy Repositories Audit & Certification: Criteria & Checklist (*TRAC) was published in February 2007 [30] presenting criteria for audit and certification.

The idea that repositories have to pass various audit and certification criteria to call themselves "trustworthy digital repositories" has gained support among larger archival institutions worldwide.

The TRAC checklist is divided into three sections:

- Organizational infrastructure
- Digital object management
- Technologies, technical infrastructure, and security.

Opponents to this approach claim that cost and effort needed to be certified excludes the vast amount of smaller digital repositories and archives. Moreover, the "medicine" might not be sufficient to provide trustworthiness of the digital records that resides within a certified trustworthy digital repository [15].

TRAC will not be followed up any further in this article.

[19] The word migration is sometimes used instead of conversion or transformation. The word conversion, or alternatively transformation, is used as one specific type of migration, with refreshment, replication, and repackageing as other types of migration.

[20] There exist two variants, batch conversion or conversion on demand.

6 LONG TERM RECORDS MANAGEMENT VS. LONG TERM PRESERVATION OF DIGITAL RECORDS

6.1 Trust across the Life-Cycle of Records

Digital records go through different phases during their life cycle and a generalized view of these phases is presented below in Table 2. Here, the column under the heading “Action type” generalizes what goes on in each phase following a traditional life-cycle view.

If we look at the table, the active phases of a record is covered in phase 1 and 2. In a traditional life-cycle of records, the often complicated transformation to archival submission format is covered in phase 3 and 4. Finally, the creation of archival format together with a records life as (passive) archival object is covered in phase 5 and 6.

Life-Cycle Phase		Action type
1	Record Creation	▪ Capturing Digital Material ▪ Verifying/validating the Digital Material ▪ Adding new Descriptions & Metadata
2	Records Use & Maintenance	▪ Reformatting Digital Material ▪ Adding/modifying Descriptions & Metadata
3	SIP Creation	▪ Capturing all relevant Records (Digital Entities, Descriptions & Metadata) in a SIP (Submission Information Package) ▪ Reformatting all relevant Records into SIP format ▪ Adding new Descriptions & Metadata into the SIP
4	SIP Transmission & Validation	▪ Transmitting the SIP ▪ Verifying/validating the transmitted SIP
5	AIP Creation	▪ Capturing all relevant Records (Digital Entities, Descriptions & Metadata) in a AIP (Archival Information Package) ▪ Reformatting all relevant Records into AIP format ▪ Adding new Descriptions & Metadata into the AIP
6	AIP Use & Maintanance	▪ Reformatting Digital Material ▪ Adding/modifying Descriptions & Metadata

Table 2 Overview of the different phases in a records life-cycle

Roughly speaking the actions are of the following types, regardless if they appear within records management of active records or archival management of passive records:

- Capturing digital material;
- Verifying and validating digital material;
- Reformatting digital material; and

- Adding new descriptions and metadata.

To make records trustworthy over time, evidence of all the above-mentioned actions has to be properly collected and documented within reasonable security controls, through all the phases. Included here is the identification of the actors for the purpose of making them accountable for their actions. Another challenge lies in identifying the right type of evidence and the suitable level of documentation. Yet another challenge lies in the securing that reformatting does not compromise trust.

6.2 Short-term Records Management, Long-term Preservation of Records

The traditional view of a record’s life-cycle is that the period as active record is relatively short, from days or months up to some years, while the archival period is long. In such a scenario the transformation into archival formats, including the SIP (Submission Information Package) generation, transmission, and validation according to OAIS, often takes place as a kind of batch process. The Norwegian Noark-4 standard (and the newer Norak 5 standard) follows these lines. Here, the record creator (archive creator) has the juridical responsibility for a record for 25 years, from the creation of the record. But records are deposited into an archive, like the Norwegian Archival Services, every five years, according to Noark.

Norak-4 covers records management in the context of public case handling. Figure 4 is an attempt to illustrate certain aspects of what the Noark-4 standard instruct the owner of record collections, the archive creator, to do as part of archival deposit.

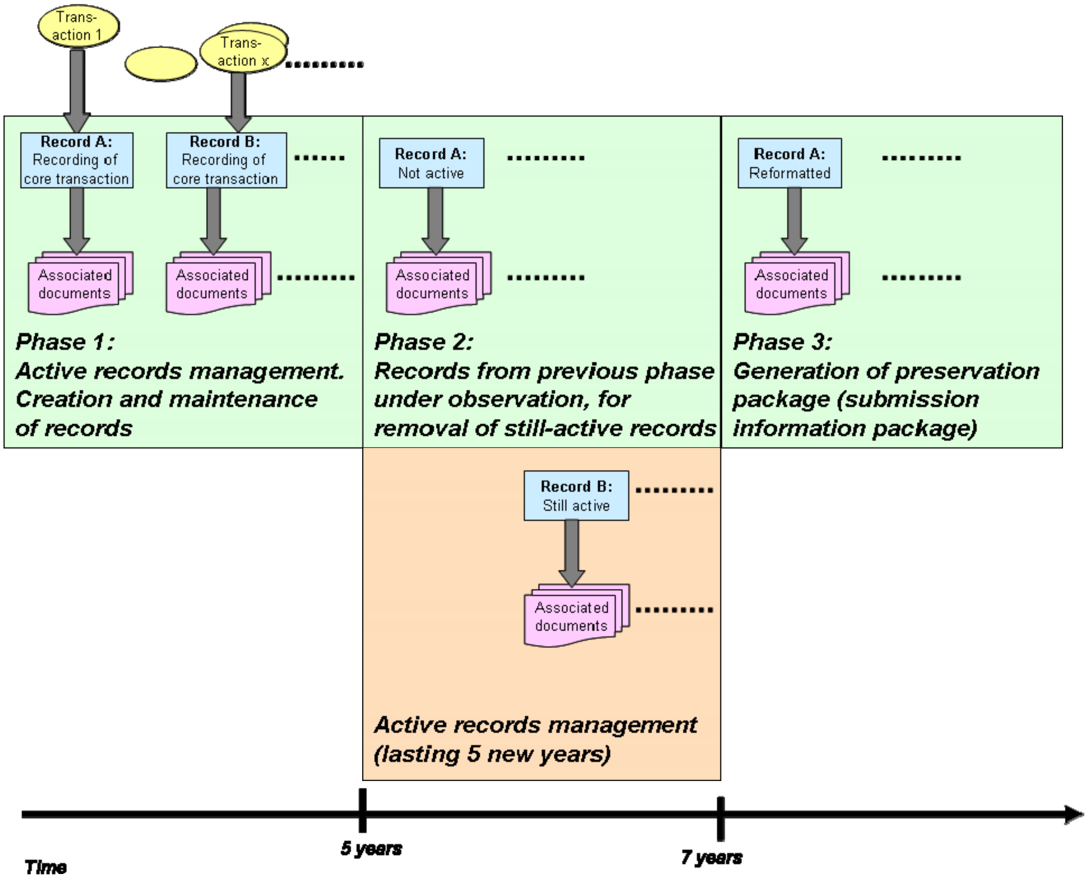
Trans-action 1
Trans-action x
Record A: Recording of core transaction
Record B: Recording of core transaction
Associated documents
Associated documents
Phase 1: Active records management. Creation and maintenance of records
Record A: Not active
Associated documents
Phase 2: Records from previous phase under observation, for removal of still-active records
Record A: Reformatted
Associated documents
Phase 3: Generation of preservation package (submission information package)
Record B: Still active
Associated documents
Active records management (lasting 5 new years)
Time
5 years
7 years

Figure 4 Phases prior to submission into archival depot

Since each period is defined to be lasting for five years, new databases serving the electronic records management system have to be created every five years. All new records (record metadata) are put here, what about the records created, used, and maintained in the previous period that still are active? To solve this dilemma the databases of the previous period are kept for two more years, and every time a record is needed in the course of business that record and the whole collection to which it belongs are removed from the old database and inserted into the new one.

After the two year period is over a huge SIP is created based on the (old) databases containing the collections of records that where not activated during the two year waiting period. The created SIP is an XML encapsulation, ideally, containing the same information as the old databases. Although in practice this completeness/integrity of information is not always that easy to achieve, sometimes resulting in re-generation of SIPs.

The exceptions among the records are worth a comment. In the scenario described above where the vast majority of records are active only for a few years, there might be collections of records that are reloaded into the records management database, not only for one additional five year period, but for several. In these cases, long-term preservation challenges might occur when the status of the records are active.

6.3 Long-term Records Management

The traditional view of a records life-cycle might need some re-thinking when it comes to digital records that are active in business for many decades, like 10-50 years or more.

One solution is to generate archival deposits, SIPs, every 5-10 years or so. But this approach generates several challenges in the archive regarding synchronizing the deposited information.

Additionally, preservation strategies have to be moved down to the active business records management arena, since old digital material faces the risk of becoming obsolete.

An alternative, which will be referred to later in this report, is to create archival packages right from record creation. As a requirement for succeeding here, the amount and type of descriptions and annotations developed when creating SIPs in the traditional scenario can no longer be expected. Instead, atomic SIPs/AIPs containing only one, or a few, records have to be created. These must be able to refer to each other within the active records management system.

Without clear strategies on how to archive and preserve, long-term records management seem to be a good candidate for what is called the “museum strategy” in the long-term preservation literature. In the museum strategy the main goal is to keep the old systems running until they break down. For critical long-lived records this is a high risk project. Therefore the archival dimension has to be taken care of along the lines discussed in this section.

7 THE SYSTEM SECURITY PREREQUISITE

A few security controls are presented and discussed in the following. These security controls should be used in order to enhance trustworthiness of either electronic records management systems or electronic archival systems.

7.1 Authorization and Access Control

Authorization means setting the rules for access to records and to the systems holding them. Confidentiality means preserving authorized restrictions on access and disclosure, including means for protecting personal privacy and proprietary information. It is the quality of restricting access to authorized users and ensures that record data remains private data that will not be disclosed without permission.

Access control means enforcing access according to the authorizations. Two main access control models exist: Mandatory Access Control or Discretionary Access Control. In the latter case, the information owner or any other authorized user can determine the authorizations that other users obtain. This model is assumed to apply to most repositories. Furthermore, authorization and access control may be applied as follows: i) Identity based access control where authorizations are set for individual users and enforced accordingly; ii) Role based access control where authorizations are granted to defined roles and users are assigned their relevant roles; iii) Task based access control where authorizations are assigned to the users or roles involved in a particular instance of a process. Role based access control is the most used model today in the world of record management.

User and role management are more and more taking place in a separate identity management system offering standard interfaces. This way, user and role management is common across all IT-systems, and a records management or archival system may be replaced with the new system reusing the same interface and the same user and role information.

7.2 Auditing and Audit Logs

Accountability is the ability to provide a report, explanation, or justification of decisions, events, actions, conditions, or understandings. By auditing records, organizations are allowed to maintain accountability with regard to the use of documents, because they know precisely; i) How someone has used or modified a record; ii) How often each type of usage occurred; and iii) When that usage occurred.

The goal of auditing is usually to verify the effectiveness and correct implementation of existing technical and organizational security measures on the one hand, and uncover any previously unidentified weaknesses on the other. The results of an audit are increased overall security through elimination of vulnerabilities, demonstrable security level in case of disputes and recourses, competitive advantage, and optimization of security management.

Trustworthiness in relation to auditing of records depends on;

- The ability to capture each use and modification session of a record in audit logs/audit trails;

- The accessibility, including search and retrieval, of log information over time;

- The security protection of the audit logs over time;

7.3 Database security

Database security is the system, processes, and procedures that protect a database from unintended activity. Unintended activity can be categorized as i) Authenticated misuse, ii) Malicious attacks, or iii) Inadvertent mistakes made by authorized individuals or processes.

Database security is more critical as networks have become more open. Traditionally databases have been protected from external connections by firewalls or routers on the network perimeter with the database environment existing on the internal network opposed to being located within a demilitarized zone. And in additional network security devices that detect and alert on malicious database protocol traffic.

Databases provide different kinds and layers of information security, typically specified in the data dictionary, including:

- Access control;
- Auditing;
- Authentication;
- Encryption;
- Integrity controls;

Trustworthiness increases by performing various security measures on the databases, including: Risk and vulnerability assessments, based on various security scans of both the database and its environment, to discover and correct misconfiguration of controls; Continuous monitoring for database security standard compliance, specifically patch management and permission management/review; Risk and vulnerability assessments of application software/sever and their database dependencies;

There should also be regular reviews of permissions granted to individually owned accounts and accounts used by automated processes. The accounts used by automated processes should have appropriate security controls around password storage such as sufficient encryption and access controls to reduce the risk of compromise.

Trustworthy database auditing is worth some elaboration. Database auditing involves observing the database user activity, in order to ensure that unauthorized users are not accessing the database. This provides the auditor with assurance that the policies, procedures, and safeguards that management has enacted are working as intended and also helping the auditor to identify any violations that may have occurred.

The ideal approach to effectively capture and analyze database activity, is through non-trigger audit agents associated with each database server. Non-trigger audit agents capture all relevant activity, regardless of the application used. The non-trigger database audit agents gather information through three means:

- Database transaction log – Each database maintains a database transaction log through the normal course of its operation, which gathers data modifications and other activity. This approach is not practical however as these logs are used for forward recovery and their formats are largely undocumented. Additionally, SQL SELECT access to database objects is not logged.

- Database's built-in native logging– Obtains additional information, such as permission changes and data viewing activities. Each database management system has some type of audit trace capability such as Oracle's Fine Grain Auditing (FGA) capability.

- Third-party tools that monitor network and/or system activity in real-time looking for database access. Some solutions use agents which enable both local access as well as network access to be monitored; while others are restricted to just monitoring network traffic. These solutions are typically called Database Activity Monitoring (DAM) solutions.

Use of agents or native logging is required to capture activities executed on the database server, which typically include the activities of the database administrator. Agents allow this information to be captured in a fashion that can not be disabled by the database administrator, who has the ability to disable or modify native audit logs. This separation of duty is a typical auditor requirement, stating that audit trails should be securely stored in a separate system not administered by the database administration group.

8 TRUST STRATEGIES BASED ON DIPLOMATICS

8.1 InterPARES

The InterPARES[21] 1 international research project, 1999-2001, focused on how to preserve authenticity and reliability of administrative and legal records generated within databases and document management systems. These were records intended to approximate the physical documents generated in the course of established business procedures in well-understood juridical contexts. For this investigation, concepts from contemporary archival diplomatics were used for identifying features of documents that make them records- fixed, reliable, complete representations of transactions.

The case studies of InterPARES 1 found that few systems contained entities satisfying the diplomatic definition of a record. Even systems that did contain records did not retain enough information about identity and integrity; so, by definition, the records could not be preserved authentically according to the InterPARES' requirements for authenticity. This illustrates the fact that electronic systems are often being designed to manage data rather than records. The studies also encountered types of information interfaces that did not fulfil the fixity requirements expected of records; like interfaces assembling information from various, continuously updated sources. Like records, such interfaces present the decisions and actions of organizations, but they are not stored or fixed, which raises the question of whether they could be preserved in a sensible way.

The InterPARES 2 project was carried out between January 2002 and December 2006 based on the experience from InterPARES 1. Its goal was to develop a theoretical understanding of for records generated by experiential, interactive and dynamic systems, of their process of creation, and of their present and potential use in the broadly defined areas of artistic, scientific and governmental sectors, together with methodologies addressing trustworthiness, archival selection, preservation and technology analysis. One of the primary findings of InterPARES 1 was that the chain of preservation of digital records has to begin with records creation, resulting in dedicated research efforts of the InterPARES 2 to the analysis of the creation of the records that it aims to preserve. InterPARES 2 also sought to avoid the problems incurred in the course of InterPARES 1 that resulted from that project's preestablished epistemological perspective on the concept of record. Hence InterPARES 2 decided not to define at the outset the concept of record, instead leaving it open to any possibility as presented by the research findings and, consistent with this stance, to accompany the deductive approach with an inductive one.

The InterPARES 1's Authenticity Task Force developed two sets of practical guidelines for ensuring the authenticity of digital records over time, addressing different phases in the lifecycle of records. A distinction is here made between active records that are maintained by the creator for current and future reference, and inactive records that have been transferred to the custody of an archive for long-term preservation. The guidelines consist of two sets of requirements;

- The benchmark requirements forming a basis for presuming or verifying the authenticity of the creator's digital records; and

[21] InterPARES is an acronym for "International Research on Permanent Authentic Records in Electronic Systems".

- The baseline requirements support the production of authentic copies of digital records after they have been transferred to the preserver's custody;

Both sets of requirements define and give a basis for assessing the records' identity and integrity, which must be preserved for the copies to be authentic. In addition, InterPARES 2 has developed a set of "Creators Guidelines"[22], to be used by record creators as a tool for pro-actively being able to support reliability and authenticity right from the time of record creation.

Below is a complete listing of the different themes (theme headings) covered in the "Creators Guidelines" and likewise for the Benchmark Requirements "supporting the presumption of authenticity", found in the "Preservers Guidelines" of InterPARES. This is done to give the reader a notion of the coverage of these guidelines. Then this strategy towards trust is discussed.

8.2 The Creator Guidelines

If we take a look at the recommendations presented in the "Creator Guidelines" [23] appendix of the InterPARES 2 book [8] it is an extensive list covering the following themes:

1. Select hardware, software and data formats that offer the best hope for ensuring that digital materials will remain easily accessible over time.

    a. Choose software that presents materials as they originally appeared.

    b. Choose software and hardware that allow you to share digital material easily.

    c. Use software that adheres to standards.

    d. Keep specifications of software.

    e. If you customize software, make sure you document the changes you make.

    f. Document the construction of your system as a whole to help ensure its accessibility.

    g. Choose widely used non-proprietary, platform-independent, uncompressed formats with freely available specifications where possible.

2. Ensure that digital materials maintained as records are stable and fixed both in their content and in their form.

3. Ensure that digital materials are properly identified.

    a. Names of the persons involved in the creation of the digital materials

    b. Name of the action or matter

[22] All three guidelines are available from the following address: http://www.interpares.org/ip2/ip2_documents.cfm?cat=pg

[23] http://www.interpares.org/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf

c. Documentary form
d. Digital presentation
e. Date(s) of creation and transmission
f. Expression of documentary context
g. Indication of attachments
h. Indication of copyright or other intellectual rights
i. Indication of the presence or removal of a digital signature
j. Indication of other forms of authentication
k. Indication of the draft or version number
l. Existence and location of duplicate materials outside the digital system.

4. Ensure that digital materials carry information that will help verify their integrity.

   a. Names of handling person/office
   b. Name of person or office with primary responsibility for keeping the materials
   c. Indication of annotations added to the materials
   d. Indication of any technical changes to the materials or to the application(s) responsible for managing and providing access to the materials
   e. Access restriction code
   f. Access privileges code
   g. Vital record code
   h. Planned disposition

5. Organize digital materials into logical groupings

6. Use authentication techniques that foster the maintenance and preservation of digital materials.

7. Protect digital materials from unauthorized action.

8. Protect digital materials from accidental loss and corruption.

9. Take steps against hardware and software obsolescence.

10. Consider (other) issues surrounding long-term preservation.

8.3 The Preserver Guidelines' Benchmark Requirements

The basic premise of the diplomatic approach is that recordkeeping functions and processes endure even if the physical manifestation of the record changes because of technological implementations [10].

Based on diplomatic analysis the InterPARES project has developed benchmark requirements [17] supporting the presumption of authenticity of electronic records[24]. To support a presumption of authenticity the preserver must obtain evidence that:

- A.1.a) The value of the following *identity attributes* are explicitly expressed and inextricably linked to every record:
    - Names of the persons (author, writer, originator, addressee) concurring in the formation of the record;
    - Name of action or matter;
    - Date(s) of creation and transmission (chronological date, received date, archival date, transmission date);
    - Expression of archival bond (classification code, file identifier)
    - Indication of attachments;
- A.1.b) The value of the following *integrity attributes* are explicitly expressed and inextricably linked to every record:
    - Name of handling office;
    - Name of office of primary responsibility;
    - Indication of types of annotations added to the record;
    - Indication of technical modifications;
- A.2) The creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records
- A.3) The creator has established and effectively implemented procedures to prevent, discover, and correct loss or corruption of records.
- A.4) The creator has established and effectively implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change.

[24] http://www.interpares.org/display_file.cfm?doc=ip2(pub)preserver_guidelines_booklet.pdf

- A.5) the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator;

- A.6) If authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication.;

- A.7) If multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative;

- A.8) If there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

8.4 Creation of Trust in InterPARES

The benchmark requirements supporting the presumption of authenticity of electronic records, is meant to as a tool for preservers. The assessment of the authenticity of the creator's records should take place as part of the appraisal process, when decisions are made on what to archive. This should happen as early as possible, long before transfer to archives. This assessment should then be verified when the records are transferred to the preserver's custody. The benchmark requirements are the conditions that serve as a basis for the preserver's assessment of the authenticity of the creator's electronic records. Satisfaction of these benchmark requirements will enable the preserver to infer a record's authenticity on the basis of the manner in which the records have been created, handled and maintained by the creator.

Historically, in the history of InterPARES, the benchmark requirements were created in InterPARES 1 while the Creators Guidelines were developed in InterPARES 2. They are very much interrelated. So, if the Creators Guidelines are followed, the presumption of authenticity is hence stronger. Not surprisingly, InterPARES values technology-independent authentication through presumption of authenticity higher than technology dependent cryptographic techniques like digital signatures. This due to the fact that digital signatures, according to InterPARES, are subject to obsolescence themselves and, by virtue of their purpose and inherent functionality, cannot be migrated to new or updated software applications together with the documents to which they are attached.

A presumption of authenticity is an inference that is drawn from known facts about the manner in which a document has been created and maintained. Adoption and consistent application of the recommendations presented in both the Creators Guidelines and the Preservers Guidelines, in which the benchmark requirements are a par, provide, according to InterPARES, the best evidence to support such a presumption. The recommendations are cumulative: the higher the number of satisfied recommendations and the greater the degree to which an individual recommendation has been satisfied, the stronger the presumption of authenticity.

Both the Creators Guidelines, in theme 3 and 4, and the benchmark requirements, A1, are concerned with capturing the right type of metadata to be able to establish the identity of the digital material and (some sort of) integrity. The concept of *reliability* is discussed and described in InterPARES as the combination of completeness at time of creation (capture) and the strength of the creation process. This has to be established at time of creation, in order to be able to say something about the reliability of records. Therefore, the above mentioned points serve a major role in achieving reliability. Reliability is about having more trust in, e.g., a signed, stamped, sealed letter of well-known origin compared to an unsigned note of dubious origin. The same applies in the digital world, e.g. an e-mail arriving by the use of state-of-art cryptographic support is more reliable than a plain e-mail.

Both the Creator Guidelines and the benchmark requirements cover security measures, without that being the main topic. But the importance of these measures is highlighted.

Success in following the InterPARES approach seem to be very much dependent of ones ability to identify the digital material according to the Diplomatic and InterPARES definitions (criteria) presented in Table 3. The backbone is to establish integrity alternative to the traditional bit integrity by extracting the "right" metadata and by understanding various qualities, and variations, regarding the form and content of the digital material at hand.

CONCEPT		***DESCRIPTION***
Fixity:		The quality of a record that ensures *fixed form* and *stable content*
	Stable content	The quality of a record that makes the information and data contained in it immutable, and requires changes to be made by appending an update or creating a new version
	Fixed form	The quality of a record that ensures the documentary appearance or presentation is the same each time the record is retrieved.
Bound variability		The quality of a record that ensures that its documentary presentations are limited and controlled by fixed rules and a stable store of content data, form data and composition data, so that the same user activity, query, request, or interaction always generates the same result.
Documentary form:		The rules of representation according to which the content of a record, its administrative and documentary context, and its authority are communicated. Documentary form possesses both *extrinsic* and *intrinsic* elements.
	Extrinsic elements	The elements of a record that constitute its external appearance, including presentation features such as font, graphics, images, sounds, layouts, hyperlinks, image resolutions, etc., as well as digital signatures, seals, and time stamps, and special signs (digital watermarks, logos, crests, etc.).
	Intrinsic elements	The elements of a record that convey the action in which the record participates and its immediate context, including the names of the persons involved in its creation, the name and description

		of the action or matter to which it pertains, the date(s) of creation and transmission, etc.

Table 3 Concepts from Diplomatics, extended and defined for digital records in InterPARES

As a final comment concluding this section, to anyone involved in the design of electronic records management systems. They should think about the options one have regarding making not only information content (documents etc.) but also metadata structures, more immutable.

9 TRUST STRATEGIES SUPPORTING REDUNDANCY

9.1 About Redundancy

Redundancy in this context is distribution of multiple copies of digital material, either information content or associated metadata or annotations. Redundancy can be applied in all phases of records management or archival process.

The reasons for wanting redundancy in records management and archival processes, is the option of being able to compare different copies (instances) of a piece of digital material at a later stage, if needed. It should be emphasized that the digital representations (data formats) of these instances/copies should be the same. By comparing multiple copies of the same format, the comparison process can be highly automated.

Comparison of different representations (data formats) of the same material, is much more complicated. Either visual traits have to be compared using digital picture analysis, or both representations have to be supported with enough evidence (metadata and/or free text descriptions) so that the fixity concerning both stable content and fixed form can be established by this additional evidence.

In the rest of this section we will take a closer look at one method and associated tools that enable some kind of redundancy in the non-trivial case of relational SQL databases. In addition, it also solves other problems concerning preservation of databases.

9.2 SIARD- Preservation of Relational Databases

The "Software Independent Archiving of Relational Databases", SIARD, was developed in a research project funded by the Swiss Federal Archives. It has produced a preservation workflow and a set of tools, described in Figure 5. The status as of ultimo 2009 is that the Swiss Federal Archives[25] offers the SIARD format which is an open, published, standardised archiving format for relational databases. Secondly, it provides a set of software tools, the SIARD Suite, to convert databases into the SIARD format. This is also available for free.

The SIARD format as well as the SIARD Suite software are based on internationally accepted standards such as XML, Unicode and SQL:1999. The SIARD Suite software converts databases into a collection of easy-to-handle XML files, thereby preserving content, relations and metadata. It allows to view the primary data and to enhance the metadata according to each organization’s individual policy. SIARD Suite consists of three main components:

- SiardFromDb (called A0 in earlier versions) is a migration tool. It converts databases from three of the most wide-spread database formats (Oracle, Microsoft SQL Server and Microsoft Access) into the new, archivable SIARD format. The new file carries the file-name extension “.siard” and consists of several XML files for the primary data (content.xml) and the metadata (metadata.xml). All of these XML files are stored in a compressionless zip-folder (PKZip/Zip64).

- SIARDEdit (called A1 in earlier versions) allows users to document, to update and to enhance the metadata. Users can also search the metadata based on

[25] For more information take a loook at http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en

various criteria as well as compare and match metadata against those of other archived data.

- SIARDToDb (called A2 in earlier versions) enables users to load SIARD files into any of the supported database systems: Oracle, Microsoft SQL Server or Microsoft Access. For instance, it is possible to convert an Oracle database into a SIARD file and then upload the SIARD file into a new Microsoft SQL Server database. Thus, more complex searches within the archived primary database content can be conducted.

SIARD seek to exploit standard SQL for long-term preservation purposes by actively extracting data and data logic from relational database management systems, RDBMS, by using the specialized ingest tool, SiardFromDB, which map different "SQL flavours" to generic SQL, and transparently trace and document those parts which cannot be mapped [20].

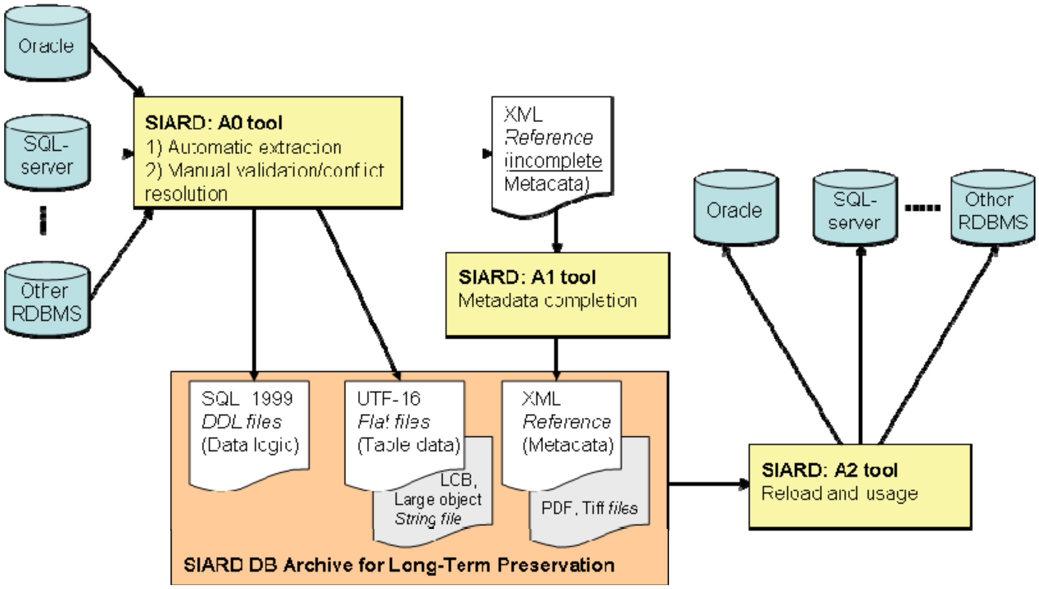
Oracle
SQL-server
Other RDBMS
SIARD: A0 tool
1) Automatic extraction
2) Manual validation/conflict resolution
XML Reference (incomplete Metadata)
SIARD: A1 tool
Metadata completion
Oracle
SQL-server
Other RDBMS
SQL 1999 DDL files (Data logic)
UTF-16 Flat files (Table data)
LOB, Large object String file
XML Reference (Metadata)
PDF, Tiff files
SIARD DB Archive for Long-Term Preservation
SIARD: A2 tool
Reload and usage

Figure 5 Overview of the SIARD approach

The SQL standard itself is far from self-contained and represents a major challenge when it comes to long-term preservation of relational databases. This because SQL:1999 explicitly identifies 381 so-called implementation-defined items and 137 so-called implementation-dependent items. Their implementation is left open for any manufacturer of RDBMS products. As long as a manufacturer completely documents all implementation-defined items, the product can rightly claim to comply with the SQL standard, though it differs from all competing products. Another reason is that most of today's RDBMS implement only (and sometimes faultily) the core and the entry level of the standard completely, but add plenty of non-standard, product-specific enhancements, leading to different "flavours of SQL". Additionally, almost all RDBMS products use their own procedural programming languages for stored routines (Oracle PL/SQL, Postgres PL/pqSQL, Microsoft T-SQL, PL/Perl etc.) rather than implementing the standard's procedural language SQL/PSM. Finally, it is an almost trivial remark that modern RDBMS move the physical storage of the data from the operating system (file level) to the application level (the internal storage of the RDBMS).

In times where the rapidly growing size and complexity of relational databases in modern relational database management systems seems to out space the ability of archives to ingest, manage, and preserve them, SIARD seems to represent an efficient solution worth trying.

SIARDs contribution to increased trust is firstly to provide a transformation to a clean subset of SQL which enables export across platforms. Secondly the transparency regarding trace and documentation of those parts which cannot be mapped into archival format.

10 TRUST STRATEGIES INVOLVING DIGITAL SIGNATURES

A digital signature refers to an electronic signature (usually) based on a Public Key Infrastructure, or PKI. Trust is based on a public key certificate issued by a certification authority together with the application of cryptographic hash functions. Digital signatures are used to secure the authenticity of digital material like records or documents.

Several challenges are present when discussing digital signatures in the context of long-term records management (long-term preservation) and trust. Two of these challenges are:

- How to demonstrate former validity of digital signatures?
- Can digital signatures, with a modified infrastructure, be useful within archiving or records management?

Both aspects will be discussed in the following sections.

10.1 Problems related to demonstration of former validity of Digital Signatures

A goal in long-term preservation is to be able to demonstrate former validity of a digital signature.

The ETSI standard TS 101 733 on electronic signature formats [32] states: "It would be quite unacceptable, to consider a signature as invalid even if the keys or certificates were later compromised".

But a signature could be forged due to leakage of the signature key or vulnerability of algorithms if the public key certificate were allowed to exceed the validity period. This is also true for revocation because a signature could be forged using the leaked key.

Thus there is a need to be able to demonstrate that the signature keys were valid around the time that the signature was created to provide long term evidence of the validity of a signature."

10.2 Approaches for Solving the Validity Problem

According to the guidelines from the National Archives and Records Administration [26] there are two main approaches to demonstrate former validity of a digital signature:

- Documentation on e-signatures validity, or
- Ability to revalidate e-signature.

Independent of which approach that is chosen one must determine what information needs to be retained to maintain a valid, authentic, and reliable signed record [26], and to preserve the link or association between the various components of a signed record over time [9].

The ETSI standard TS 101 733 suggests a solution based on time stamping by a trusted service. Further they discuss nested time stamping by a trusted service with stronger cryptographic algorithms and keys than the user as a technique for protecting against degeneration of keys and algorithms.

The Fraunhofer Gesellschaft has a project on transformation of signed electronic documents called TransiDoc[26]. They discuss two main problems, namely weakening of electronic signatures and changes in data formats that break the signature of signed documents. They follow up with an analysis of the state of the art to resolve these problems [23].

A modular framework for concrete application of electronic signatures is described in [5]. They propose an architecture consisting of different layers: infrastructure for digital signatures, application of electronic signature (middleware-layer), and application-layer. They also distinguish between three different applications of electronic documents: static documents (no workflow), dynamic documents with state variations (workflow), and finally documents with state variations (workflow) and external data exchange, with particular focus on the static documents.

To preserve the long-term authenticity of electronic records EVERSIGN [25] proposes a solution they call Signature Validity Extension. Their solution makes use of the long-term signature format in the standard RFC3126[27]. We take a closer look at the EVERSIGN initiative in the next section.

10.3 EVERSIGN, a Mechanism for Signature Validity Extension

As already mentioned, EVERSIGN [25] proposes a solution they call Signature Validity Extension, which is a technology for overcoming the validity period and revocation of public key certificates and vulnerability of cryptographic technology used for digital signatures. The purpose is to maintain the long-term validity of digital signatures. To be able to do so the four requirements for signature validity extension described in Figure 6, has to be fulfilled.

If the requirements are satisfied, then the following confirmation steps 1-3 can be used to distinguish whether the original signature is true or false.

By using the long-term signature format, RFC 3126, the requirements above can be satisfied as follows:

- Requirement (1): Assign a standard time stamp to a signature value (ES-T signature time-stamp).

- Requirement (2): Store verification information items such as the set of public key certificates and CRL and OCSP responses (ES-C and ES-X verification information references and verification information).

- Requirement (3): Assign a time stamp to electronic records with signature (ES), signature time stamp, verification information reference, and the entire verification information (ES-A archive time stamp).

[26] TransiDoc: http://www.transidoc.de/website-transidoc/index-en.html (14.nov 2007)

[27] RFC3126: http://www3.tools.ietf.org/html/rfc3126 (Obsoleted by RFC5126 http://www3.tools.ietf.org/html/rfc5126). Technically equivalent to ETSI TS 101 733.

- Requirement (4): Overlap a time stamp on the entirety in order to maintain long-term tamper-resistance (outside archive time stamp).

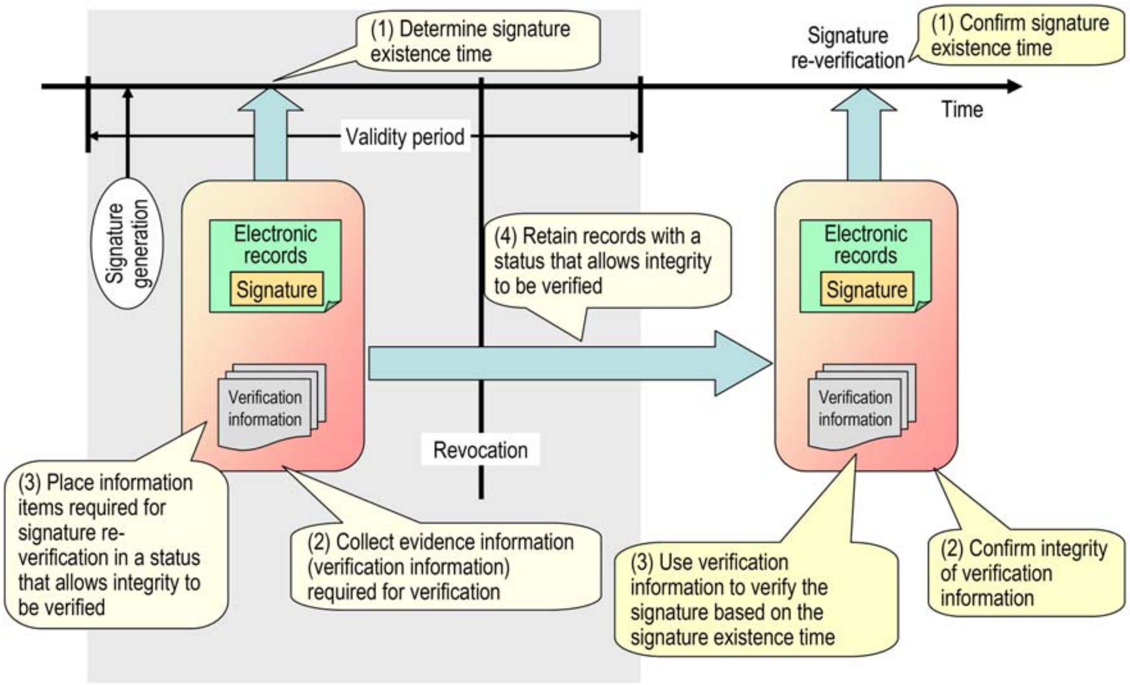
(1) Determine signature existence time
Signature re-verification
(1) Confirm signature existence time
Time
Validity period
Signature generation
Electronic records
Signature
(4) Retain records with a status that allows integrity to be verified
Electronic records
Signature
Verification information
Verification information
Revocation
(3) Place information items required for signature re-verification in a status that allows integrity to be verified
(2) Collect evidence information (verification information) required for verification
(3) Use verification information to verify the signature based on the signature existence time
(2) Confirm integrity of verification information

Figure 6 Requirements, 1-4, for Signature Validity Extension, together with Confirmation Steps, 1-3

It is claimed by the creators of this method that by using long-term signature format with a time stamp to meet requirements 3 and 4, the method has features that makes it superior to other proposed methods in the following ways:

- Standard PKI technology allows anyone to verify the validity.

- Processing to construct and extend a long-term signature can be performed by anyone and can be taken over by others in the middle of processing.

- Trust is based on only the trust point in standard PKI without needing to consider the safety of the system and operation, which are currently difficult to confirm.

10.4 Trust Strategies at Archival Entrance

From the point of view of archival institutions confronted with the need to develop policies relative to the preservation of digitally signed documents, three possible solutions have emerged according to [2]:

- Preserve the digital signatures: This solution supposes the deployment of considerable means to preserve the necessary mechanisms for validating the signatures, and does not address the need to simultaneously preserve the intelligibility of documents;

- Eliminate the signatures: This option requires the least adaptation from archival institution, but impoverishes the description of the document, as it eliminates the signature as one technical element used to ensure the authenticity of the documents;

- Record the trace of the signatures as metadata: This solution requires little technical means, and records both the existence of the signature and the result of its verification. However, digital signatures lose their special status as the primary form of evidence from which to infer the authenticity of the document.

While the first solution has often been implicitly codified in evidence law reforms (perhaps without realizing its full practical implications), it is the last solution which is most congruent with both archival practice and theory: "the findings of InterPARES indicate that integrity assurance and continuing accessibility are the key outputs of the archival recordkeeping function and that these are primarily assured through procedural and descriptive metadata. [...] Archival metadata must support the continued authenticity of records by describing the records as they were received from the records' creators and thoroughly documenting the entire process of preservation" [11].

Approaches like the EVERSIGN, or similar approaches, will in the future contribute to more refined and trustworthy solutions.

Regardless of the strategy chosen, digital signatures should be verified at Ingest if OAIS is followed. That is when SIPs are received for archival deposit. The alternatives discussed in this section could also easily be applied when signed digital materials are entering a records management system.

10.5 Can Digital Signatures be used during Records Management or Long-term Preservation?

Today's PKI based digital signatures need to be modified or relaxed one way or another to be applicable as evidence of authenticity over periods longer than a few years. Until satisfactory solutions are found and agreed upon between key players one should probably not go for digital signature regimes within archives or long-term records management systems. But for records management with records having their active status lasting only a few years, digital signatures might be used during records management.

Until suitable relaxations of digital signature infrastructures are found one has to settle for the second-best which here means using cryptographic hash functions on Archival Packages. The next section takes a closer look at one such approach, called the Digital Container approach, combining XML structures and cryptographic check sums. This combination is about to become more wide-spread in the archival world.

Then we conclude the digital signature strategy part, by presenting an example where XML structures actually are combined with digital signatures within an archival context. This approach follows the UVC preservation approach, forcing one initial conversion of formats, eliminating the need for more conversions. Once the conversions are made the archival XML structures can be signed, following a web of trust approach contrary to the traditional PKI approach.

10.6 Digital Containers: XML-Formatted Archival Packages with associated Hash Checksums

Following the OAIS standard, long-term archiving implies standardized archival packages, AIPs. But the standardized archival packages can be implemented differently. The information and meta information can be distributed across various database tables and systems. From the perspective of accessibility and efficiency this might be wise, but at the risk of loosing information, e.g., due to broken links between the different data sources. Loss of completeness, integrity, authenticity, and other properties might be the consequence.

As an alternative to distributed representation, sealed XML encapsulation is about to become state-of-art. The encapsulation should ideally include all of the content information, both original and derived, together with all the associated metadata. As an example of sealed XML encapsulations, the City of Antwerp in Belgium has been deploying a solution named Digital Container, developed by Expertisecentrum David, eDAVID,[3] [4]. Certain aspects of this solution are described in the following.

The main structure of an AIP in eDAVID, as illustrated in Figure 7, consists of three parts: i) the identifier for the AIP; ii) all representations and the essential metadata of the record; iii) the checksum. The identifier and the checksum serve mainly for the management of the AIPs. The identifier contains the unique and preferably a permanent ID of the computer file with the AIP as content and is the reference to the AIP. The checksum acts as 'fixity information' and can also be used as (part of) the AIP identifier. With a checksum, the validity of the AIPs can be thoroughly checked afterwards by comparing the embedded and the recalculated hash values with each other. This check can be carried out completely automatically and randomly. If the embedded hash value is not equal to the recalculated hash value, an alarm function can be activated (for example, to retrieve a backup). For the checksum, not only the hash value is preserved, but also identification of the applied hash algorithm[28].

It is worth noticing that cryptographic hash functions are used in this example, by the archive as checksums to detect accidental (or intentional) data corruption. Cryptographic hash functions have many information security applications, notably in digital signatures and other forms of authentication.

[28] A cryptographic hash algorithm is a deterministic procedure that takes an arbitrary block of data and returns a fixed-size bit string, the (cryptographic) hash value, such that an accidental or intentional change to the data will change the hash value.

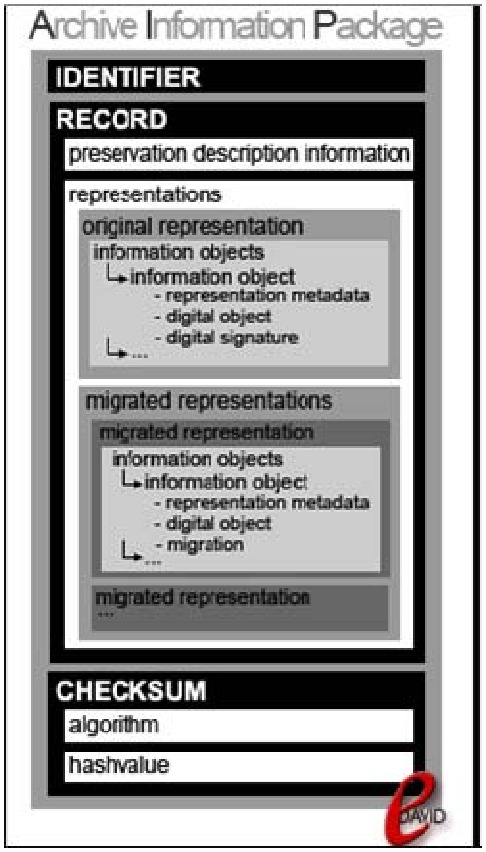

Figure 7 Encapsulated representation of AIPs[29]

10.7 Trustworthy Digital Objects: XML-Formatted Archival Packages with associated Digital Signatures

[15] [16], suggests a similar encapsulation XML-structure, but with one major difference. In addition to cryptographic message authentication, like the digital container approach just presented, a digital signature regime based on the web of trust principle is suggested. A web of trust structure is unaffected by such things as company failures and differs from PKI in that way.

The solution is called "trustworthy digital objects" (TDO), i.e., . These objects can, e.g., durable digital objects functioning both as active objects within the frames of a records management system and at the same time complying with archival standards like the OAIS. The TDO is meant to take both the role of an OAIS SIP and AIP. The obvious place for a change to information about any object is in some TDO version of that object, hence avoiding the problem with conflicting updates of multiple copies. Derivative information, such as that in library and archival catalogues, can be created and synchronized, often if necessary, by machine processes.

[29] Source: The eDAVID approach, Antwerp, Belgium.

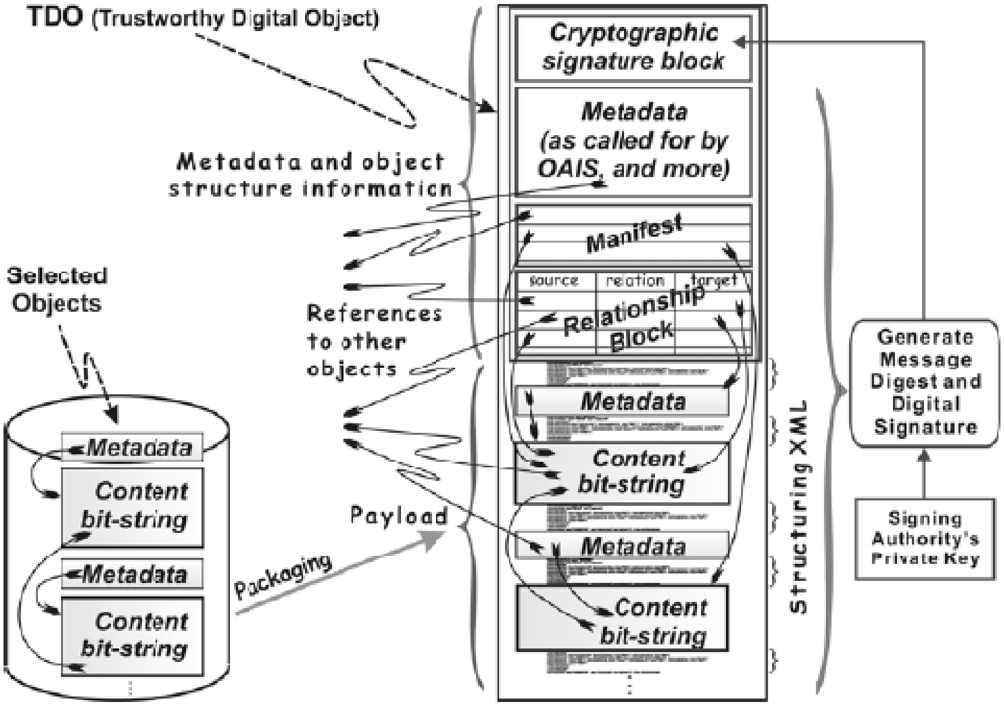
TDO (Trustworthy Digital Object)
Metadata and object structure information
Selected Objects
References to other objects
Payload
Packaging
Metadata
Content bit-string
Metadata
Content bit-string
Cryptographic signature block
Metadata (as called for by OAIS, and more)
Manifest
source
relation
target
Relationship Block
Metadata
Content bit-string
Metadata
Content bit-string
Structuring XML
Generate Message Digest and Digital Signature
Signing Authority's Private Key

Figure 8 Encapsulated SIPs/AIPs, another example[30]

The brief description of the TDO architecture is described as follows in [16]: For a record to be suitable for preservation, it must contain provenance information or be linked to such information, and the whole record must be protected against undetected tampering. We intuitively want to convey something along the lines of, "At such and such a time, John Doe communicated a document X to the distribution list Y." A TDO thesis is that authoritative metadata should be bound tightly to each preserved object. Derivatives can be extracted from full records to create repository catalogues with only modest software additions to repository implementations. The idea is to package source information collections so that:

- The bit-string set that represents a work is XML-packaged with registered schema.
- Each bit-string that represents part of the work is encoded in a computing-platform-independent representation or is accompanied by a bit-string encoded for everlasting intelligibility.
- Integrity is assured by cryptographic message authentication.
- The package includes provenance evidence, technical metadata, and one or more identifiers of the object itself.
- Links to contextual information are secured by cryptographic message authentication codes of the linked entities.

[30] Source Gladney: http://home.pacbell.net/hgladney/hmgpubs.htm#pubs

- Information loss is minimized by replication in mutually independent repositories.

- Cryptographic signatures are grounded in keys that widely trusted institutions publish periodically.

The TDO structure offers options for evidence of authenticity, contextual information, and whatever might make an object self-describing. Its objective is to enable all reasonable creators' and custodians' choices, rather than to prescribe what choices information creators should make.

11 TRUST STRATEGIES FOR DOCUMENTING PROVENANCE

The Principle of Provenance distinguishes the archival profession from other information professions in its focus on a document's context, use and meaning[31]. This Principle, generally concerned with the origin of records, has three distinct meanings [1].

1. Generally, it refers to the "office of origin" of records, or that office, administrative entity, person, family, firm, from which records, personal papers or manuscripts originate.

2. It refers to collecting information on successive transfers of ownership or custody of a particular record, paper, or manuscript; and

3. It refers to the idea that an archival collection of a given records creator must not be intermingled with those of other records creators.

In the discussion to follow we are specifically discussing the second meaning of provenance.

11.1 Provenance Documentation

Establishing provenance, i.e., the true history of ownership for digital records is comparable to establishing provenance in the world of fine art[32]: "Art historians have always sought to know the identity of previous owners, but such information is often difficult to establish. When a painting has been owned […] for generations there may be no record of sale. Frequently, private collectors prefer to buy and sell works anonymously […]. Moreover, many dealers and auction houses […] are no longer in business and their records may have been lost or destroyed. Thus it is rare to find works of art having a complete history of ownership. […] Reconstruction of a complete history of ownership for a given work can be difficult and sometimes impossible. Many records of ownership have been destroyed as a result of […] disasters […], and neglect. Information is sometimes withheld […]. Much archival information remains undiscovered or difficult of access."

In the world digital records management the responsibility of a digital record is moved from one person to another is similar to the change of ownership of a painting. Over the years an old painting has to be restored, involving either cleaning or repair of the painting. A digital record is likewise subject to new descriptions and annotations, and also to data format conversion or migration. The important provenance information needed is the data/time of an activity, the activity description at some level of detail, and identification of the actors, both who they are and their role at the time. Additionally, the motives behind an action should also ideally be documented. All this represents evidence regarding the action, its impact, and the accountability of the actors.

Similar to the world of art, reconstruction of the history of ownership will be difficult to achieve in digital record keeping. Having information spread around in databases for several years

[31] http://www.unc.edu/~winget/research/provenance.pdf

[32] http://vanweyenbergh.com/art_provenance.htm

without any intermediate processing will probably result in loss of information about ownership. As a result there will be gaps in provenance.

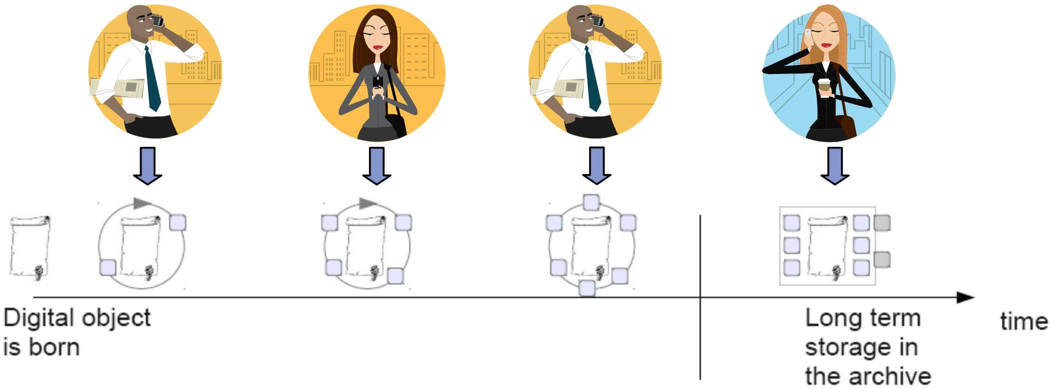

Figure 9 Successive transfers of custody combined with modifications of a particular record

Given the general difficulties in reconstructing provenance, a provenance-based trust strategy should be proactive instead of reactive in its nature. As a consequence, the record creator has to be responsible for such a strategy, not the archive.

As important as the documentation of provenance is the associated security controls. The evidence of provenance has to be secured.

11.2 Provenance Documentation in InterPARES

Establishment of provenance is covered through the baseline requirements of the InterPARES' "Preservers Guidelines". These requirements outline the minimum conditions necessary to enable the preserver to attest to the authenticity of copies of inactive electronic records. The word "copy" refers here both to the result of either "lossless" or "lossy" copy functions.

The requirements regarding what to document in case of copying, including migration and conversion, are listed below. All of the requirements must be met before the preserver can attest to the authenticity of the electronic copies in its custody. Satisfaction of these baseline requirements will enable the preserver to certify that copies of electronic records are authentic:

- B.1) The procedures and system(s) used to transfer records to the archival institution or program; maintain them; and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that:
  - Unbroken custody of the records is maintained;
  - Security and control procedures are implemented and monitored; and;
  - The content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.
- B.2) The activity of reproduction has been documented, and this documentation includes:

  - The date of the records' reproduction and the name of the responsible person;
  - The relationship between the records acquired from the creator and the copies produced by the preserver;
  - The impact of the reproduction process on their form, content, accessibility and use; and
  - In those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user.

- B.3) The archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created.

11.3 Provenance Documentation in the Durable Objects Initiative

As part of the Trustworthy Durable Object work an in-depth discussion of the concept of authenticity was presented in [12]. Here, the authors find it helpful to be explicit about criteria for workable definitions of authentic. One may need similarly careful definitions of the words for other quality measures- words that might include useful, essential, secure, legal, and so on. With this in mind, the following criteria are presented:

- Distinguish as clearly as possible between objective facts and subjective judgments.
- Within the work represented by [13], any word denoting a quality shall allow for objective evaluation of technical solutions relative to explicit requirements statements.
- Authentic should be binary- either true or false (for any entity compared to some prior entity).
- The meaning of authentic should depend as little as possible on the kind of entity in question.
- The definition for digital objects should exhibit minimal discontinuity with existing tradition.
- Whether an entity instance is or is not authentic should not depend on the intention of any human being- not its producer, not any custodian, and not its eventual users.
- The meanings of words used within a single conversation about qualities should not intersect.

One goal here is separate "objective" statements form "subjective", in order to make the former suitable for automation. The context here is an approach called Trustworthy Durable Objects, aimed at addressing long-term records management where SIPs (AIPs) are created when the active record is created.

Following up the work on defining words and meanings a formal definition of authentic is given, in which provenance plays a major part:

- Given a derivation statement R, "V is a copy of Y (V=C(Y))"
- Given a provenance statement S, "X said or created Y as part of event Z"
- Given a copy function C:, "C(y) = Tn(... (T2(T1(y))))"
- Then, if V is related to Y according to R
    - We say that V is a *derivative* of Y
- if R and S are true
    - We say that "by X as part of event Z" is a *true provenance* of V
- if C conforms to social conventions for the genre and for the circumstances at hand
    - We say that V is sufficiently faithful to Y
- if V is a sufficiently *faithful derivative* of Y with true provenance
    - We say that V is an *authentic copy* of Y

Here "copy" means either "later instance in a timeline" or "conforming to a specific conceptual object". Each transformation Tk potentially adds, removes, or alters the information carried by its input signal. To preserve authenticity, the metadata accompanying the input in each transmission step should be embedded in the corresponding output by including a description of the transformation in Tk. This is strictly necessary only for steps that alter the information content in a meaningful way.

These metadata should identify who is responsible for each Tk choice and all other circumstances important to judgments of authenticity. Suitable metadata schema are being discussed widely, e.g., in the METS initiative. Trustworthy packaging for objects and metadata is described in [14] and [15]. An object's accumulated metadata are the digital equivalent of a traditional audit trail for a physical archival holding.

Whether or not the consumer accepts a transmission as authentic will be his/her subjective decision based on weighing the evidence inherent in and accompanying the object — evidence that often extends to context provided by other objects. The provenance definitions above convey minimal requirements. The producers of provenance information might include more information, such as identification of or links to documents providing evidentiary context. Doing so is often prudent or customary.

In particular, the choice of the copy function C(y) is a subjective decision. Particularly for an object that cannot be transmitted perfectly, the producer who hopes that the eventual consumer will judge what he receives to be authentic should consider including evidence in C(y). For an object whose history includes several transmission steps this might be done in each transformation Tk(y).

12 CONCLUDING REMARKS

Strategies for trust (enhancement) cover a broad field. We have in this report been investigating trust strategies addressing and supporting trust decisions along the lines described in [24]:

1. Examining the provenance of the object (for example, the documentation of the chain of custody) and the extent to which we trust and believe this documentation as well as the extent to which we trust the custodians themselves.

2. Performing a forensic and diplomatic examination of the object (both its content and its artifactual form) to ensure that its characteristics and content are consistent with the claims made about it and the record of its provenance.

3. Relying on signatures and seals that are attached to the object or the claims that come with it, or both, and evaluate their forensics and diplomatics and their consistency with claims and provenance.

4. For mass-produced and distributed (i.e., published) objects, comparing the object in hand with other versions (copies) of the object that may be available (which, in turn, means also assessing the integrity and provenance of these other versions or copies).

Trust (enhancing) strategies along all these lines have been described, using state-of-art research initiatives or best practice approaches as examples.

The trust context in this article has been that of digital records, from their birth and active role as part of electronic records management, ending up in archival environments for long-term preservation and access. The distinction between (i) Records management and long-term archiving on one side, and (ii) Long-term records management is discussed. The latter option is whenever records are expected to live over decades in a business context as active records.

The actions taking place in both records management and archiving are of the following types and trust strategies are discussed related to these actions:

- Capturing digital material;

- Validating digital material;

- Reformatting digital material; and

- Adding new descriptions and metadata.

The most critical phases related to trust is the initial capture of what is going to become a digital record, further the transferring of collections of digital records from active records management to archives, and finally the data format transformations/conversions that has to go on from time to time to be able to read/access the digital material also in the future. In all these critical phases obligations are put on the actors, obligations regarding extensive documentation of that action. This documentation is needed to assure trust across transformations and transmissions, and it becomes the primary evidence when, eventually, the bit integrity is broken and cryptographic tools are no longer valid after a transformation. For this reason verification and validation of digitally signed material is also very important to provide evidence of prior validity of digital signatures.

Regarding the use of digital signatures, in the case of (shorter-term) records management followed by long-term preservation of passive records, there is no problem in using them for authentication during records management. In the long-term digital records management case, modified signature infrastructures are needed.

Hopefully, this article will be useful as a starting point for those who are going to define their own long-term records management trust strategies. The field of trust establishment and maintenance is still immature and will in the future be subject to more research. This report can also be seen as an overview of the different research directions.

REFERENCES

[1] Bellardo, L. J., Bellardo, L. L., *A glossary for archivists, manuscript curators, and records managers*. Chicago: Society of American Archivists, 1992.

[2] Blanchette, J.-F., *The digital signature dilemma*, Annals of telecommunications / Annales des télécommunications, Vol. 61, no. 7-8, July-August, 2006.

[3] Boudrez, F., *Digital Signatures and Electronic Records*, retrieved from <http://www.expertisecentrumdavid.be/docs/digitalsignatures.pdf>, 2005.

[4] Boudrez, F., Digital *Containers for Shipment into the Future*, retrieved from <http://www.expertisecentrumdavid.be/docs/digital_containers.pdf>, 2005.

[5] Berbecaru, D., Lioy, A., Maino, F., Mazzocchi, D., Ramunno, G., *Towards concrete application of electronic signature,* AICA symposium 2000, Taormina (CT), Italy, September 27-30, pp. 543-561, 2000,

[6] Chan, M., Zeng, M. L., *Metadata Interoperability and Standardization—A Study of Methodology*, D-Lib Magazine, Volume 12 Number 6, 2006.

[7] Duranti, L., *Diplomatics: New Uses for an Old Science (Part V)*, Archivia, Volume 32, 1991.

[8] Eppard, P:, Domain 2 Task Force, "Appendix 20: Creator Guidelines – Making and Maintaining Digital Materials: Guidelines for Individuals," [electronic version] in International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records, Luciana Duranti and Randy Preston, eds. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008). <http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_20.pdf>, 2008.

[9] *Electronic Signatures and Records Act (ESRA) guidelines*, New York State Office for Technology, May 26, 2004, Revised: September 28, 2007.

[10] Gilliland-Swetland, A.J., Eppard, P. B., *Preserving the Authenticity of Contingent Digital Object*, D-Lib Magazine, Volume 6 Number 7/8, ISSN 1082-9873, 2000.

[11] Gilliland-Swetland, A.J., *Electronic Records Management*. ARIST 39, pp. 219-25, 2005.

[12] Gladney, H. M., Bennett, J. L., *What Do We Mean by Authentic? -What's the Real McCoy?*, D-Lib Magazine, Volume 9 Number 7/8, ISSN 1082-9873, retrieved from <http://www.dlib.org/dlib/july03/gladney/07gladney.html>, 2003.

[13] Gladney, H. M., *Trustworthy 100-Year Digital Documents: Executive Summary of a Digital Preservation Proposal*, 2003.

[14] Gladney, H. M., *Trustworthy 100-Year Digital Documents: Evidence Even After Every Witness is Dead*, 2003.

[15] Gladney, H. M., *Durable Digital Objects -Rather Than Digital Preservation*, preprint, 2008, retrieved from <http://eprints.erpanet.org/146/01/Durable.pdf>.

[16] Gladney, H. M., *Long-Term Preservation of Digital Records: Trustworthy Digital Objects*, DRAFT, June 2009, retrieved from <http://home.pacbell.net/hgladney/LDPreview2.pdf >.

[17] Hackett, Y., Domain 3 Task Force, "Appendix 21: Preserver Guidelines – Preserving Digital Records: Guidelines for Organizations," [electronic version] in International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records, Luciana Duranti and Randy Preston, eds. (Padova, Italy:

Associazione Nazionale Archivistica Italiana, 2008).
<http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_21.pdf>, 2008.

[18] Hackett, Y., Underwood, W., Eppard, P., *"Part One—Case and General Studies in the Artistic, Scientific and Governmental Sectors: Focus Task Force Report,"* [electronic version] in International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records, Luciana Duranti and Randy Preston, eds. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008).
<http://www.interpares.org/display_file.cfm?doc=ip2_book_part_1_focus_task_force.pdf>, 2008.

[19] Hawkins, K., "Diplomatic Analysis, Case Study 19: Preservation and Authentication of Electronic Engineering and Manufacturing Records," [electronic version] in International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2:.
<http://www.interpares.org/display_file.cfm?doc=ip2_cs19_diplomatic_analysis.pdf>, 2006.

[20] Heuscher, S., Järmann, S., Keller-Marxer, P., Möhle, F., *Providing Authentic Long-term Archival Access to Complex Relational Data*, European Space Agency Symposium "Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data", Frascati, Italy, October 7, 2004.

[21] *InterPARES 2 Project, The InterPARES 2 Project Glossary*, [electronic version] in International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records, Luciana Duranti and Randy Preston, eds. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008).
<http://www.interpares.org/display_file.cfm?doc=ip2_book_glossary.pdf>, 2008.

[22] Jøsang, A., *The Right Type of Trust for Distributed Systems*, Proceedings of the 1996 New Security Paradigms Workshop, 1996.

[23] Kunz, T., Okunick, S., Viebeg, U., *Long-term security for signed documents: services, protocols, and data structures*, 2006.

[24] Lynch, C., *Authenticity and Integrity in the Digital Environment: An Exploratory Analysis on the Central Role of Trust*, presented at Authenticity in a Digital Environment, Washington, D.C., CLIR report 92, May 2000.

[25] Miyazaki, K. and Tanaka, M. EVERSIGN*: Preserving the Long-Term Authenticity of Electronic records*, Mitsubishi Electric Advance, Vol. 118, ISSN 1345-3041, June 2007.

[26] *Records Management Guidance for Agencies Implementing Electronic Signature Technologies,* 2000, retrieved from http://www.archives.gov/records-mgmt/faqs/pdf/electronic-signiture-technology.pdf.

[27] Consultative Committee for Space Data Systems*: Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1, Blue Book, 2002, retrieved from public.ccsds.org/publications/archive/650x0b1.pdf.

[28] RLG/OCLC Working Group on Digital Archive Attributes, *Trusted Digital Repositories: Attributes and Responsibilities*, RLG-OCLC Report, 2002, retrieved from <http:www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>.

[29] *Integrity*, [electronic version] in Stanford Ensyclopedia of Philosophy <http://plato.stanford.edu/entries/integrity/>.

[30] *Trustworthy Repositories Audit & Certification: Criteria & Checklist* (TRAC), Research Libraries Group (RLG) and the National Archives and Records Administration (NARA) , 2007, retrieved from: http://www.crl.edu/PDF/trac.pdf.

[31] Ølnes, J., *A Taxonomy for Trusted Services*, First IFIP Conference on e-Commerce, e-Business, e-Government (I3E), Zurich, 2001.

[32] ETSI TS 101 733 V1.5.1 (2003-12), *Electronic Signatures and Infrastructures (ESI); Electronic Signature Formats*, RTS/ESI-000017, <http://docbox.etsi.org//EC_Files/EC_Files/ts_101733v010501p.pdf>, 2003.