Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Report

Preservation of Trust in Long-Term Records Management Systems

A State of Art Overview for the LongRec Project

Report no	1017
Authors	Arne-Kristian Groven, Jon Ølnes, Habtamu Abie, Truls Fretland
Date	23. April 2008
ISBN	978-82-539-0527-3

© Copyright:
Norsk Regnesentral

About the authors

Arne-Kristian Groven and Habtamu Abie are Senior Research Scientists in the DART department at Norsk Regnesentral. Truls Fretland is a Sesearch Scientist at the same place.

Jon Ølnes is a Senior Researcher at Det Norske Veritas (DNV) Research & Innovation.

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.

DNV- Det Norske Veritas

DNV is an independent foundation established in 1864. DNV's objective is "safeguarding life, property, and the environment" and the vision is "global impact for a safe and sustainable future". DNV's core competence is to identify, assess and advise on how to manage risk. The main focus industries are maritime (ship classification and other services) and energy (mainly oil and gas) but DNV offers services to a broad range of industry sectors. Many services are cross-sector, such as management system certification (e.g. ISO9001, ISO14001, ISO27001) and corporate responsibility. DNV Research & Innovation is a separate department in DNV. The IPT (Information Processes and Technology) programme in DNV R&I investigates future IT-related services from DNV. DNV R&I initiated the LongRec project and leads the project consortium.

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	**Preservation of Trust in Long-Term Records Management Systems. A State of Art Overview for the LongRec Project**
Authors	**Arne-Kristian Groven, Jon Ølnes, Habtamu Abie, Truls Fretland**
Quality assurance	Inger-Mette Gustavsen, DNV R&I
Date	23. April
Year	2008
ISBN	978-82-539-0527-3
Publication number	1017

Abstract

This report is produced as part of the work done in LongRec project which is partly funded by the Norwegian Research Council, project number 176818/I40. The primary objective of this joint-industry project is the persistent, reliable, and trustworthy long-term archival of digital information records with a lifespan of tens or hundreds of years.

The main topic in this report is to give a state of art overview on how to preserve the trust in digital records over decades? The answer is not a simple one and the report goes through different problem areas using using authenticity as criterion for trustworthiness and evidential value.

Keywords	trust, evidential value, security, long-term records management, digital preservation, authenticity
Target group	Long-term records management audience
Availability	Open
Project number	320369
Research field	Security
Number of pages	55

© Copyright	Norsk Regnesentral

Summary

This report is written as part of the LongRec-project, and is one of several subtopics related to long-term preservation of digital records. It gives a state-of-the art overview regarding preservation of trust.

At first the concepts of trust and trusworthiness is discussed within the context of a digital repository and associated processes.

TRAC, a quality management approach, is briefly presented, focusing on organizational infrastructure, digital object management, in addition to technologies, infrastructure, and security within digital repositories. 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. Opponents claim that cost and effort needed to be certified excludes the vast majority of smaller digital repositories and that being certified as a trustworthy digital repository is not sufficient to provide trustworthiness of the digital records that resides inside the repositories. Encapsulated, durable encoded objects should instead be in focus.

A "best practice" example is then presented. This approach is based on encapsulation, using XML, of the original digital content (bitstreams) and associated metadata (ingested into the repository) and the content and associated metadata for all derivations of the original bitstream. In this way the risk of loosing crucial information over time is minimized. The encapsulation also includes authentication mechanisms to be used in the archival context and allows freedom on how to handle digital signatures.

Digital signatures are then discussed. Signatures are unable to testify the identity and integrity of a digital document over time. The main value of a digital signature is lost after the first change in the bitstream of the digital object/document. Among the topics discussed are strategies for testifying the existence, in the first place, of the the identy and integrity of the signatures themselves.

In order for a digital record to be a "competent witness" of a juridical fact, a digital object/-document must be accompanied by traces of all of the operations which it is susceptible to incur: creation, modifications, annotations, signature, conversion, transmission, etc. One of the main challenges is to express authenticity through metadata, and make it last through changes in the content-bitstream. In addition security services that can last over longer periods of time has to be designed and implemented

Contents

List of figures

1 Introduction

1.1 The LongRec Project

This report is produced as a contribution to the LongRec (Long-Term Records Management) project 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*. The project's public web site is at http://research.dnv.com/longrec/

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, http://www.interpares.org

LongRec addresses several research challenges[1], 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.

1.2 This Report

This report is the state of the art report for the TRUST (preservation of evidential value) research area). The report describes the common ground for further research in this area in LongRec and also addresses topics of particular interest to the project partners.

There are many different reasons for preserving documents. For society in general, historical and scientific research are two good reasons for preserving documents or other artefacts. In the business community, documents are mainly preserved for legal reasons. Documents are kept because of legal requirements or because of obligations to do so by virtue of a contract, or for the sake of their value as evidence.

A person can prove his trustworthiness by fulfilling an assigned responsibility - and as an extension of that, to not let down our expectations. The responsibility can be either material, such as delivering a mail package on time, or it can be a non-material such as keeping an important secret to oneself.

Preserving an electronic document as evidence makes little sense unless one can trust that the document and its content are authentic. This implies that the document must at all times be properly protected from unauthorized events (accidental or deliberate), and that all events that occur must be properly recorded.

[1] We refer to the project's web site http://research.dnv.com/longrec for a description of the research challenges.

2 Trustworthiness in Long-term Preservation Systems

2.1 Timeline of a Preserved Digital Object

Figure 1 illustrates the travel in time (and space) of a digital object/entity, from creation, through becoming a digital record, until it is finally being read (accessed) years, decades, or even centuries after the time of creation. The arrows indicate the timeline.

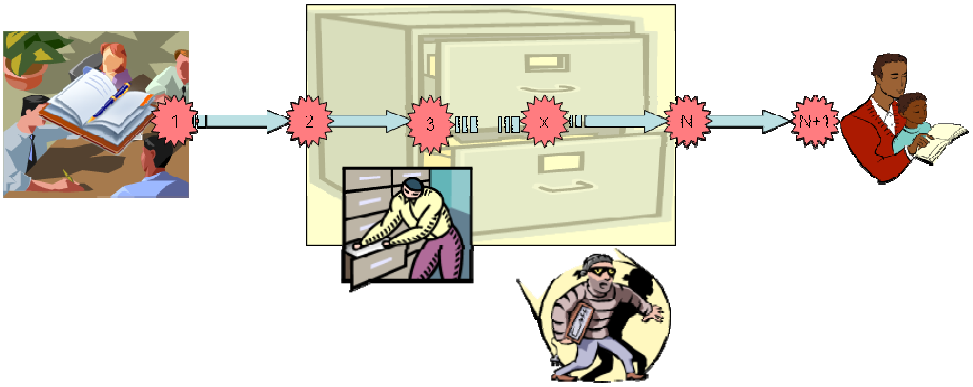

Figure 1: The time travel of a digital entity/digital record, where 1 through n+1 indicates critical points in the lifecycle.

The figure also illustrates the different actors involved, from the (semi-)stable document, or other type of digital entity/record, is produced by one or more individuals in a context. This context is e.g. a group of people or an organisation altogether defining the origin of the digital entity.

The last group of actors involved along the timeline are the ones reading or using the digital entity years after it was produced. They are the ones deciding whether to trust the digital entity and associated metadata presented to them or not. Figure 1 describes the situation where one reader or user is involved. This is of course applicable every time a user is involved.

In between you have the archivist trying to preserve the readability/usability of the digital entity, and at the same time trying to preserve the trustworthiness of the document. But this actor might produce errors and mistakes, decreasing the trustworthiness of the digital entity.

In addition you have the potential threats from outside, illustrated by a thief, but the consequence might be destruction or degradation of trustworthiness. The numbered points (stars) illustrate points in time where trustworthiness might be threatened. These are of different types that will be described in section 2.3.

Digital preservation is associated with several standards, of which some are briefly mentioned in Appendix A: Standards.

2.2 Trust and Trustworthiness

The critical question for the persons accessing digital content after a period of time is whether the content can be trusted or not. Trust is a subjective decision, e.g., I may trust something, or someone, while you do not trust the same thing or person.

According to [Øln01] trust can be defined as "perceived lack of vulnerability". A trust decision implies a (human) judgment about the vulnerability implied by a certain action. Trust decisions

are not necessarily rational. Trust is a subjective decision, based on perceived, not real, vulnerability. The decision may be made deterministic, e.g. by implementing the criteria in program code, but ultimately one still has a human decision in the definition of the criteria.

[Jøs96] separates trust decisions into trusting "rational entities", computers and the like that behave according to programmed instructions, and "passionate entities", which are human or humanly controlled and may behave according to will. Both aspects, trusting technology and trusting persons, organizations and processes, are relevant also in the long-term. For rational entities properties such as security, reliability and safety must be assessed. Likewise, these properties can be assessed for processes and organizations, and even for persons (e.g. security clearance).

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. Thus, one way of viewing this situation is that it may be possible to compute a degree of trustworthiness as a function of knowledge and assumptions, presumably also including assessment of the uncertainty related to the assumptions.

This computation is then also part of an assessment of a computation of the evidential value of the information. However, other elements such as formal or legal requirements may also come into play with respect to evidential value.

With respect to a trust decision, if trustworthiness could be computed, one would typically deduce that a trustworthiness value above a certain threshold would yield a "trusted" decision.

In the case of long-term digital records management, *trust is related to whether the user/reader believes in the digital record presented to her/him years after it was 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, and nothing else, years after time of creation is definitely not enough! In order to gain acceptance, both the digital records and the long-term preservation/records management systems must be trustworthy, i.e., worthy of reliance or trust.

Trustworthiness related to the digital record itself includes being able to keep/demonstrate important properties like integrity and authenticity (identity/origin/provenance), while trustworthiness in the long term preservation systems (records management systems) relates to reducing the system vulnerabilities in all possible ways. This includes reducing possible impacts by errors or intentional actions by humans, be they external or internal (potential) attackers.

2.3 Where Trustworthiness might be Threatened

One of the widely used standards is ISO 14721:2002, the Open Archival Information System Reference Model, called OAIS for short. OAIS is presented in Appendix B: OAIS Concepts.

Within the OAIS terminology, as illustrated in Figure 6, the phase between point 1 and two in Figure 1 is called the ***Ingest*** phase, while the phase between point n and n+1 in Figure 1 is called the ***Access*** phase. All in between defines the ***Digital Repository***, including management of content and metadata. This management includes changes in storage media, metadata, and bit streams. Below is listed some of the threats that might compromise trustworthiness.

2.3.1 Threats at Ingest

Trustworthiness might be compromised by:

- Insufficient inclusion of provenance/origin/context information
- Lack of integrity protection.
- Security breaches related to travel in space, e.g. on the Internet.

2.3.2 Threats within the Digital Repository

At the entry point trustworthiness might be threatened by:

- Insufficient capture of (provenance/origin) metadata
- Integrity is broken (se causes below)
- Signatures are not validated/verified

Trustworthiness might also be compromised by, e.g., the following actions:

- When content management processes are performed, e.g.:
    - new storage media or other type of technology is introduced;
    - conversion takes place, from one content format to another;
    - back-ends of an emulator is modified to fit new hardware/operating systems;
    - management of signatures, e.g. resigning of records;
    - new metadata is added, etc.
- When accidental or deliberately harmful modifications takes place.

2.3.3 Threats at Access

Trustworthiness might be compromised by:

- Security breaches related to travel in space, e.g. on the Internet.
- Incorrect presentation (display or otherwise) of information. (Note that presentation format may be different from preservation format.)
- Reduced accessibility to (e.g. impossible to verify) content, metadata, signature validation chains etc.

2.4 Evidential Value

Generally electronic records (digital artefacts) are saved for different reasons;

- they are used in the routine activities (of an organization), often called "administrative value";

- they indicate/prove what the person/organization has been doing, often termed "evidential value";

- they contain information of longstanding value, often called "informational value";

- they reflect aspects of a person's/society's/an organization's development, often termed "historical value";

Our main focus in this report is on the ***evidential value***.

We have not found any standardized definition of the term evidential value, but we have found a lot of descriptions/definitions from a lot of sources. In this jungle we have just picked one that was found most suitable at the time of our writing, to illustrate:

- "Evidential value is the quality of records that provides information about the origins, functions, and activities of their creator. Evidential value relates the process of creation rather than the content (informational value) of the records[2] ."

We said earlier, in section 2.2, that in the case of long-term digital records management, trust is related to whether the user/reader believes in the digital record and associated metadata presented to her/him years after it was created. One main question is, does he/she e.g. believe in the authenticity of the digital record?

Evidential value, in practice, is linked to the juridical system, which defines different rules for making a rational trust decision based on the evidential value of the digital record presented for the court years after its creation.

One might think of evidential value without an explicit link to the juridical system. But one main motivation for organisations to archive digital entities as records is the intention of being able to stand in court, if needed, even decades after a digital object was archived, having the archived digital object accepted as evidence in court.

It is also a difference between frozen digital content, i.e. something that is written and stored with the intention of not changing the content or meaning in any ways for as long as it is going to be stored, and so called semi-stable digital content and metadata, e.g. patient journals and associated attachments where new information is expected to be added, but old information remains unchanged. In the latter case, the security and privacy management must play a major role, in addition to keeping the authenticity.

2.5 Problem Areas to be discussed in this Report

Preserving trustworthiness through quality management, audits, and certification will be briefly presented in chapter 3. This approach [TRAC07] has gained support among larger archival institutions worldwide. But there are opposition, mainly related to the cost and effort needed such approaches and the fact that the focus is on institutional procedures and (management-) systems, and not on optimizing the design of durable digital objects.

[2] http://rpm.lib.az.us/alert/thesaurus/terms.asp?letter=e

In chapter 5 we present what we call a "best practice" example, referring to an approach implemented in the city of Antwerp [Bou05a, Bou05b]. This approach is based on encapsulation, using XML, of the original digital content (bit stream) associated metadata (ingested into the repository) and the content and associated metadata for all derivations of the original bit stream. The encapsulation also includes authentication mechanisms to be used in the archival context, and allows freedom on how to treat digital signatures.

Digital signatures are themselves unable to testify the identity and integrity of a digital document over time. The main value of a digital signature is lost after only one change in the bit stream of the digital object/document. This is discussed in chapter 6 in addition to strategies for testifying the existence, in the first place, of the identity and integrity of the signatures themselves.

In order for a digital record to be a "competent witness" of a juridical fact (commitment to obligations), the digital object/document must be accompanied by traces of all of the operations which it is susceptible to incur: creation, modifications, annotations, signature, conversion, transmission, etc. Finding suitable ways of describing authenticity that can last over decades and centuries is one of the main research challenges. Two different views on how to define and maintain authenticity of digital objects over time and past conversions are discussed in chapter 7.

In chapter 8 the need for security services that can last in a long-term perspective is emphasised. The different components are presented and aspects related to the long-term perspective are discussed

In chapter 9 different aspects of evidential value are discussed. Here, a case example from Belgium is presented, where three (fictive) persons use three different preservation strategies and their expected success rate in court is described.

3 Trustworthy Digital Repositories, a Quality Management Approach

3.1 Background

Trust can be accomplished by standardizing quality management around the digital repositories being responsible for long-term digital preservation.

One main approach is the approach towards *trusted digital repositories*. This effort to develop criteria for trustworthy digital repositories began in 2002 with the publication of the RLG-OCLC report entitled Trusted Digital Repositories: Attributes and Responsibilities [RLG02]. 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 be developed for the certification of digital repositories. In order to be worthy the label trusted digital repository, the idea is that a repository has to pass various audit and certification criteria. A new document, version 1.0 of the *Trustworthy Repositories Audit & Certification: Criteria & Checklist (*TRAC) was published in February 2007 [TRAC07] presenting criteria for audit and certification.

3.2 TRAC versus Digital Containers

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. Opponents to this approach claim that cost and effort needed to be certified excludes the vast amount of smaller digital repositories. If the certification itself is the costly part, then this is clearly a problem. However, if the real problem is that running a trustworthy repository with all necessary controls in place is inherently costly, then the smaller repositories may have a problem regardless of certification.

More importantly, the "medicine" might not be sufficient to provide trustworthiness of the digital records that resides within a certified trustworthy digital repository according to the opponents. [Gla08] writes "[...] Repositories are merely tools for housing and disseminating the best human artefacts. Straining to make cultural institutions do what they are ill-suited to accomplish makes little sense. Instead it wastes skills and resources that could be better employed. A lesson is evident. Prescribing how clerical procedures might achieve digital preservation by creating "Trusted Digital Repositories" is not the best available objective. Instead we should focus on structure and content that create usefully "Durable Digital Objects.""

However, even digital containers must reside in a trusted repository in the sense that they must be protected against deletion and modification. Checksums and other digital container protection measures only detect changes; they cannot by themselves correct changes (unless error correcting codes are applied but this is considered to be too costly and besides can be broken in deliberate attacks). Requirements on the repository itself may be more relaxed but in practice the two approaches must be applied together in some way, and cost-benefit analysis of different approaches may guide the repository design.

3.3 TRAC Coverage

The TRAC checklist is divided into three sections:

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

3.3.1 Organizational infrastructure

Organizational infrastructure includes but is not restricted to these elements: (a) Governance, (b) Organizational structure, (c) Mandate or purpose, (d) Scope, (e) Roles and responsibilities, (f) Policy framework, (g) Funding system, (h) Financial issues, including assets, (i) Contracts, licenses, and liabilities, and (j) Transparency.

Criteria addressing these elements are organized in these five groups:

- A1: Governance and organizational viability
- A2: Organizational structure and staffing
- A3: Procedural accountability and policy framework

- A4: Financial sustainability

- A5: Contracts, licenses, and liabilities

3.3.2 Digital Object Management

The digital object management responsibilities of a repository include both some "organizational" and technical aspects related to these responsibilities, such as repository functions, processes, and procedures needed to ingest, manage, and provide access to digital objects for the long term. Requirements for these functions are categorized into six groups based on archive functionality, allowing grouping under the well-known OAIS functional entities:

- B1: The initial phase of ingest that addresses acquisition of digital content.

- B2: The final phase of ingest that places the acquired digital content into the forms, often referred to as Archival Information Packages (AIPs), used by the repository for long-term preservation.

- B3: Current, sound, and documented preservation strategies along with mechanisms to keep them up to date in the face of changing technical environments.

- B4: Minimal conditions for performing long-term preservation of AIPs.

- B5: Minimal-level metadata to allow digital objects to be located and managed within the system.

- B6: The repository's ability to produce and disseminate accurate, authentic versions of the digital objects.

Requirements here assume familiarity with OAIS and/or with detailed repository practices.

3.3.3 Technologies, Technical Infrastructure, and Security

These requirements do not prescribe specific hardware and software to ensure AIPs can be preserved for the long term, but describe best practices for data management and security. In total, these criteria measure the adequacy of the repository's technical infrastructure and its ability to meet object management and security demands of the repository and its digital objects.

Criteria here are similar to the good computing practices required in international management standards like ISO 27002. Repositories or organizations that have undergone ISO 27001 certification are very likely to meet many of these criteria. Providing proof of certification to relevant IT management or security standards can serve as the required evidence for some of the criteria within section C.

These requirements are grouped into three layers:

- C1: General system infrastructure requirements.

- C2: Appropriate technologies, building on the system infrastructure requirements, with additional criteria specifying the use technologies and strategies appropriate to the repository's designated community(-ies).

- C3: Security–from IT systems, such as servers, firewalls, or routers to fire protection systems and flood detection to systems that involve actions by people.

4 Overview of different Preservation Strategies

There are several different preservation strategies. The baseline strategies are:

- Maintain technology: Keep all necessary hardware and software in order to process the archived formats.

- Emulation: Keep formats unchanged but develop and maintain software to process these formats on new platforms

- Conversion: Convert objects to new formats when regarded necessary in order to be able to discard old technologies.

Storage management of records includes three activities:

- Refreshing: Copy to another media instance of same type, without altering bits of representation or associated descriptive data – this may be necessary for all preservation strategies;

- Migration: Copy to a media instance of a different type (e.g. a new storage technology), without altering bits of representation or associated descriptive data – this may be relevant for all strategies but some formats may be tied to a particular medium rendering migration without conversion impossible;

- Conversion/transformation: Process which generates a new representational form while attempting to preserve information content.

Technology maintenance leads to a "technical museum" associated with an archive and is in general not feasible. However, one may not be able to emulate all kinds of objects; e.g. writing an emulator for a computer game designed for some old computer may be very time-consuming and new errors might be introduced. Similarly, there may be limits to migration and conversion technology, and information might be lost during transformations. Management of migrations might in itself be challenging depending on the frequency and amount of information. Conversion to (a limited set of) standard formats is preferable to maintaining a large number of formats.

Weaknesses in the above preservation technologies have resulted in the development of other more combined approaches, using (some of) the baseline approaches mentioned above as components. [THI02] gives an overview of different preservation approaches, as illustrated in Figure 2. We will here briefly describe a few of them.

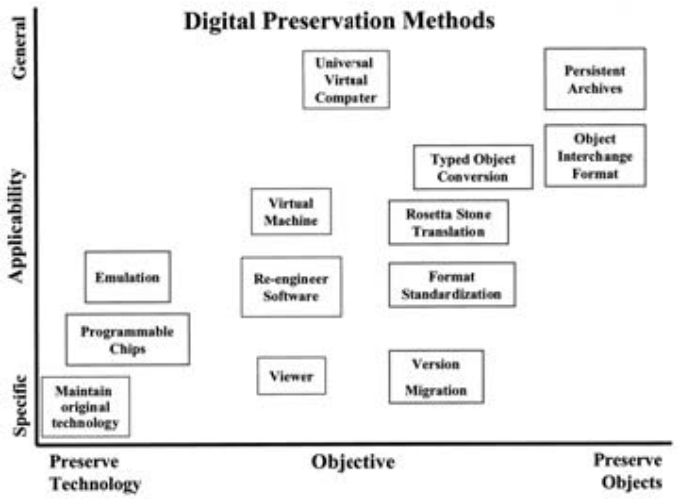

Figure 2: Digital Preservation Methods, reproduced from Kenneth Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years".

The Universal Virtual Computer, or UVC, is part of a UVC-based preservation method. This method is invented by R.A. Lorie (IBM Research Center in Almaden) and allows digital objects (like text documents, spreadsheets, images, sound waves, etc.) to be reconstructed in its original appearance in the future using a combination of emulation and conversion. The UVC is designed to be a general-purpose computer, implementing a universal Turing machine. The main criticism against emulation approaches is that writing emulators (probably) introduces new errors. The UVC approach reduces this risk, since the UVC machine has less complexity compared to an emulator. But the extra cost is that a transformation must (initially) take place, from the original format to the UVC compatible format.

The Typed Object Conversion, TOM, and Rosetta Stones Translation are two preservation approaches, based on conversion. TOM articulates the essential properties of each data type, to which digital data (objects) belongs. If e.g. the essential properties are "content" and "appearance", then a digital document can be stored either as PDF or Word and you have "respectful conversion" between the two. Rosetta Stones constructs representative samples of objects of a particular type, instead of articulating essential properties. It adds a parallel sample of the same object in another, fully specified type, and retains both.

Large amounts of data and metadata may have to be stored in order to maintain trust in the authenticity of a digital object and to be able to interpret or execute it correctly. Several preservation approaches involve encapsulation of all relevant metadata and (all derivations of the) content, in one way or another. We look into some aspect of encapsulation in the following. Other approaches include [Gla03d] who combines encapsulation with the UVC approach, and the Persistent Digital Archives initiative [Moo00] also involving the application of GRID technology.

5 Digital Containers, A "Best Practice" Example

The City of Antwerp in Belgium is implementing a solution that can be described as a best-practice example. The solution has been developed by Expertisecentrum David (eDAVID) and

is described in [Bou05a, Bou05b]. We could have selected a more state-of-art-in-research example, but have instead used an example that is used in practice.

5.1 Digital Containers, the eDAVID Approach

In the eDAVID preservation strategy [Bou05a, Bou05b] there is at least four options for reconstructing the records:

- emulation of the original format;
- conversion of the original format;
- conversion of the suitable archiving format;
- emulation of the suitable archiving format.

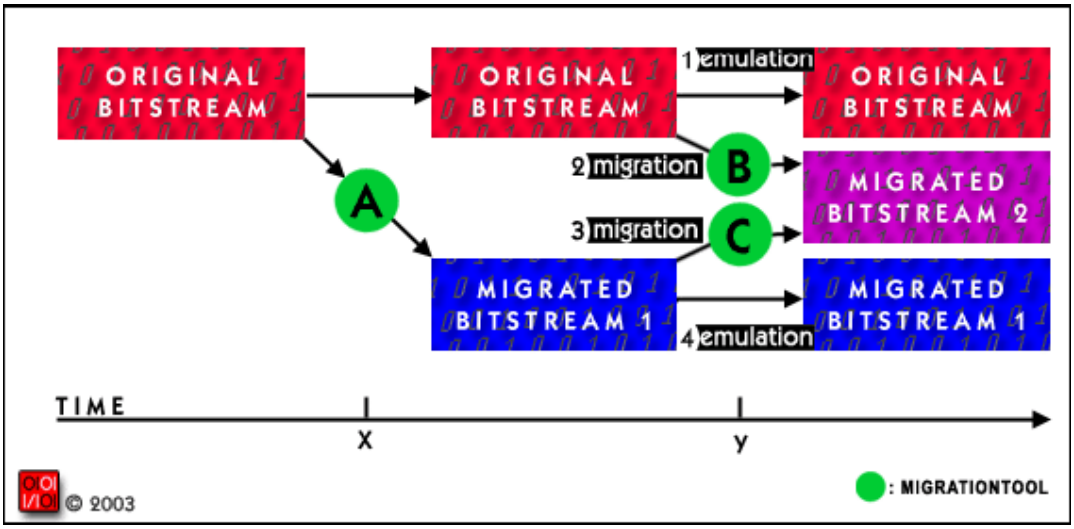

Figure 3: The eDAVID approach opens for both conversion (called migration in the figure)[3] and emulation, and keeps all bit streams, both the original and all derived ones.

By including both the original and the converted bit stream in the digital repository, one anticipates also the future technological evolution. 'Original' means the bit stream ingested into the repository. There is a presumption that this should if possible be the same as the bit stream that was submitted for preservation; however when the submitted information is not in a bit stream format, a conversion clearly has to take place at ingest:

- Paper is scanned into a bit stream format;
- Original content may be tied to a particular media type, e.g. sound or other media.

This preservation strategy, in addition to providing as many readability guarantees as possible, also offers that users can consult an electronic record in both the original bit stream and in a converted bit stream, depending on their preference or on the software applications they have.

[3] The figure is borrowed from [Bou05b] Filip Boudrez, Digital *Containers for Shipment into the Future,* where the term migration is used with the same meaning as conversion in the LongRec terminology .

Secondly, when the original bit stream is archived, authentication remains possible on the basis of technologies that relate to the original bit stream. An advanced digital signature is an example of this. A condition is that all elements of the 'validation chain' and the necessary metadata must be available. Thirdly, records in their original and converted bit stream can be compared or the conversion process can be reconstructed.

With most storage methods, the various components of an electronic record do not form a physical entity, but are stored at separate locations (in a database, a file system or a combination of both) and as different digital objects. Their mutual relationship is indicated by means of links, database relations, pointers and filenames. Archiving these relationships is not self-evident in the (medium to) long term. The fast evolution of information technology requires that the relationships between the digital objects are established in a clear and permanent manner. This is not an insurmountable problem, but it is an important point and can involve a challenge as time passes. In addition, the danger always exists that relationships might be lost.

Preserving the components of an electronic record separately always involves a risk. As soon as mutual relationships are broken and cannot be reconstructed the record must be considered as lost. Metadata are indeed essential for the long-term preservation of and access to the electronic record, including the existence of persistent, unique identifiers in both data and metadata. The archivist can avoid this risk by including metadata in the files that contain the documents. By combining both components in one physical object, the relation between the record and its metadata is prevented from becoming lost.

Keeping metadata and data together is not a prerequisite for permanent electronic record-keeping, but it is well worth considering since it provides important advantages:

- The metadata are inextricably connected with the record. One does not have to worry about links or pointers between digital objects and their metadata. Encapsulation also facilitates management in the (medium to) long term.

- All components of an electronic record can easily be transferred and migrated together.

- The electronic records are self-descriptive and autonomous: they identify and document themselves.

- The embedded metadata can be extracted at any time and stored centrally.

- The objects in the digital repository have record status without needing external information. Electronic records rather than digital objects form the basic units of the repository

- The consequences of disasters might be less serious (risk assessment):

    - the digital repository still contains records;

    - metadata can be extracted from the records.

5.2 Encapsulation of AIPs

The main structure of an AIP in eDAVID consists of three parts:

- the identifier for the AIP;
- all representations and the essential metadata of the record;
- the checksum.

The identifier and the checksum serve mainly for the management of the AIPs. The identifier contains the unique ID of the computer file with the AIP as content and is the reference to the AIP. Preferably, this should be a permanent ID so it can serve as an identifier for the AIP on a long-term basis. The checksum functions 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 hashing algorithm.

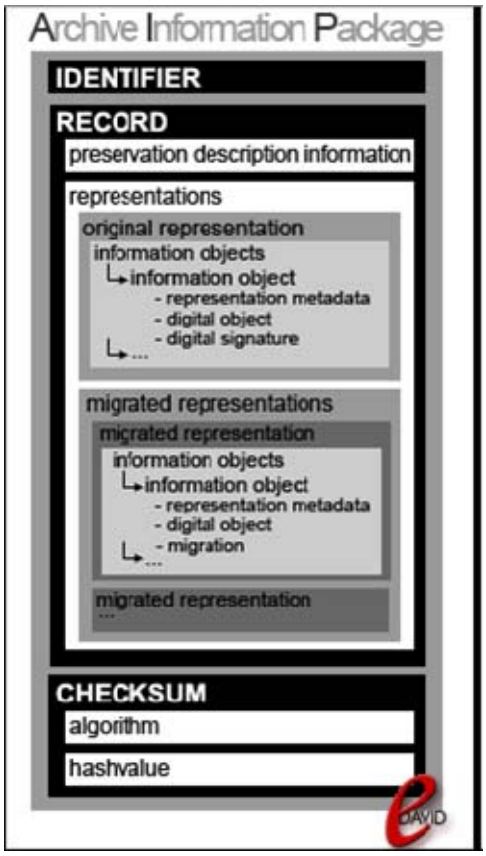

Figure 4: Encapsulation of AIPs in the eDAVID approach.

The second part in this AIP structure contains all components of the electronic record and is split further into several sub elements. The archival descriptive metadata and the records management metadata are included in the sub element 'preservation description information'. These metadata relate to every representation of the electronic record and therefore only have to be stored once. The second sub element ('representations') contains all representations and the

technical metadata of the electronic record. The structure provides space for one or more archiving file formats besides the original representation of the record. A record can have more than one suitable archiving format or, in future, new conversions can be needed. Each representation may consist of one or more computer files ('information objects'), as there might be a one-to-one or a one-to-many relationship between a record and computer files.

By using XML as the file format for the container files, each organisation can work out a custom-made container model for the AIPs depending on its own needs and approach.

For the implementation of the above-described storage method using XML container files, eDAVID developed various XML Schemas for the City of Antwerp. These XML Schemas define the formal model for the XML documents. There are XML Schemas for:

- the XML container file or the AIP;
- a general record-keeping metadata set for the management of electronic records (work in progress);
- the archival descriptive metadata in conformity with ISAD(G);
- the document types for which XML is used as the archiving format: e-mails, calendars and databases.

This strategic choice of XML results in a combined application of XML. First, XML is used as a language in which all parts of an AIP are packed as electronic records. Here XML is used as an encapsulation format. Second, XML is also used as a suitable archiving format for several document types. Third, XML is also used as the metadata format for the essential metadata. These metadata are stored directly in XML.

5.3 Trustworthiness in the Digital Container Strategy

The encapsulation of the metadata at ingest, using XML, improve the trustworthiness, reducing the risk of loosing meta-information of value, e.g. about the origin of the records. Checksums are used to *detect* modifications.

The digital container approach also takes into account that the digital signature problem is not solved and opens for different strategies concerning how to handle digital signatures. By keeping the original bit stream within the encapsulated object actions like verification of signatures, can be performed even if the content is no longer readable. Digital signatures will be further discussed in the next chapter.

The vulnerability in this solution, like all long-term digital preservation solutions, lies in the conversions (transformations) where the transformation method might produce lossless transformations or information might be lost. The challenge will always be to produce an authentic derivation. The digital container approach stores information about each transformation and encapsulates every derivation. In this way you may have several readable derivations available at any given time. This also gives some assurance.

6 Digital Signatures

6.1 The Role of the Signature

The politics in most parts of the world are geared towards widespread acceptance of electronic communication. In paper-based communication, signatures fulfil important functions, which are reflected in laws and regulations. Thus, legal compliance and the (at least medium-term) need to relate electronic communication to accepted procedures for paper-based communication create legal requirements for electronic signatures.

A vital question to ask is: why do we sign? There is not a single answer to this. Answers will differ dependent on culture, practice, and the legal system in various countries. One suggestion for the purposes of a signature is:

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

It is fairly clear that a digital signature can fulfil all these purposes. This is confirmed by the American Bar Association, which states that a signature efficiently serves the functions of evidence, ceremony and approval.[4]

A further question may be if, and under which conditions, electronic communication without digital signatures can fulfil such purposes. As explained below, many different types of electronic signatures exist. Specific regulation must define (minimum) requirements for electronic signatures for each use case. In some European countries, the direction is to require or at least recommend use of an advanced or qualified (i.e. digital, see below) signature whenever a legal requirement for signatures exist. In other countries, such a parallel between paper and digital signatures is not drawn and other forms of electronic signatures are generally accepted.

6.2 European Legal Framework and Classes of Signatures

In the EU a harmonized legal framework for electronic signatures was put in place by the E-signature directive [EUDIR99] issued in 1999. The E-signature Directive defines various types of electronic signatures (see Figure 5) and attaches particular legal consequences to one of them: The qualified electronic signature.

4 American Bar Association: http://www.abanet.org/scitech/ec/isc/dsg-tutorial.html

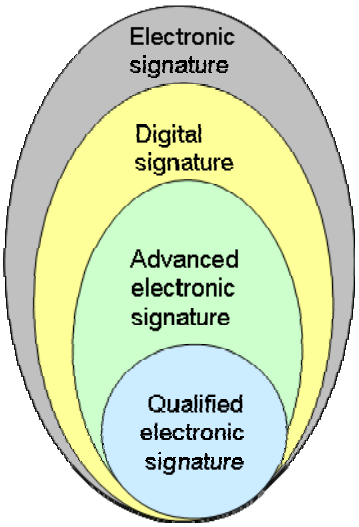

Figure 5: Different types of electronic signatures.

Electronic signature is a technology neutral term that is used to denote any data in electronic form that is attached to or logically associated with other electronic data and serves as a method of authentication (Article 2, 1° E-signature Directive). For example, putting your name under an ordinary e-mail can be regarded as a very basic form of an electronic signature.

More weight is attached to a specific kind of electronic signatures, namely the advanced electronic signature (AES). Such a signature is uniquely linked to the signatory, is capable of identifying the signatory and is created using means under the sole control of the signatory. Also, it is linked to the signed record in such a manner that any tampering is detectable (Article 2, 2° E-signature Directive). Although AES is also in principle a technology neutral term, in the current state of technology, only digital signatures can fulfil all these requirements. A digital signature is not necessarily an AES, e.g. this is not the case if the signatory is not the only one actor that can control signing.

A subset of advanced electronic signatures enjoys a particularly privileged status, namely the qualified electronic signature (QES). Not only must a QES be admissible as evidence in legal proceedings, it must be accorded the same legal consequences as a handwritten signature would receive in similar circumstances (Article 5, §1 E-signature Directive). The main benefit of using a QES is the uniformity of its treatment in the entire EU. This property is very attractive to anyone seeking to maximize legal certainty. A QES is an AES with additional requirements imposed. The QES shall be accompanied by a qualified certificate (QC), which is an eID certificate (PKI certificate) issued by a certification authority (CA) adhering to specific requirements. The QES shall also be created by a certified[5] secure signature creation device (SSCD); this requirement is usually fulfilled by storing the signer's private key in a certified smart card or similar device.

Although the E-signature Directive is written in terms that are supposedly neutral towards the signature technologies available on the market, the conditions of an AES are tailored to digital signature technology. As of yet, the predominance of digital signatures remains unchallenged.

[5] According to CEN Workshop Agreement CWA 14169, which specifies a Common Criteria (ISO/IEC 15408) profile for secure signature creation device. Evaluation assurance level EAL4+ is required..

Note that the E-signature Directive and the QES term are relevant in Europe only. Most industrialized countries outside Europe also have legislation in place for electronic communication and e-signatures but requirements vary.

6.3 Current use of Digital Signatures

Many agree that the presence of electronic signatures is much less than expected. However, the reasons given differ. Some blame the fragmentary legislations; others say that the economical model is wrong, whereas others again say that the technical solutions aren't developed enough.

In Europe, the Survey on the standardisation aspects of e-signatures [Study07] shows that there are still remaining issues in standardisation. To this one may add that the study and the E-signature Directive focus only on Europe and European standards, to a large extent neglecting the global scope of use of e-signatures. The IDABC study on signature interoperability across borders [IDABC07] concludes that cross-border use of electronic signatures is infeasible today. The IDABC study recommends introduction of trusted validation services as a means to achieve interoperability. This is also suggested by [Ølnes07].

Wang [Wang06] concludes that "the divergent and fragmentary legislations [for e-signatures] around the world do not constitute an environment under which e-commerce would flourish, and to some extent create new barriers to international e-commerce." This may be the cause for the much slower than anticipated uptake of digital signatures. The Report on the EU Directive from 2006 [EUREP06] says that "the use of qualified electronic signatures has been much less than expected and the market is not very well developed today". There are indications of increased use since 2006 but the volume is still small.

The EU-report [EUREP06] also states that "Another practical reason for the reluctance to implement e-signature applications is that the archiving of electronically signed documents is considered too complex and uncertain. Legal obligations to keep documents for as long as over 30 years require costly and cumbersome technology and procedures to ensure readability and verification of such period of time."

While qualified certificates are available in almost all European countries, SSCD products are at present (start of 2008) available in less than half the countries. This means that AES can be used across Europe while QES is only available in some countries. There is varying emphasis on QES in various countries. While QES is required (or at least highly recommended) in some countries, other countries only require AES or simpler electronic signatures.

This reflects another statement and intention of the E-signature directive: That an electronic signature (of any kind) shall not unduly be denied legal value only on the grounds of it being electronic; although only a QES will have a *guaranteed* legal value.

A typical use of a simple electronic signature is a reporting/submission application where the user logs on to the service using an eID of sufficient quality (requirements may range from username and static password, via one-time passwords to use of PKI-based eIDs). This logon together with an explicit "submit" action (such as pressing a submit button) is regarded as an electronic signature, provided that the log functionality of the system can be used to show the link between authentication, submission, and the content submitted.

6.4 Standards for Long-Term Electronic Signatures

There are a number of standards and recommendations that deal with electronic signatures in general, so we here we only present those that concern long-term preservation. For an extensive overview of the EU e-signature standardisation work see page 119 in the e-signature-survey [Study07].

The European Telecommunications Standards Institute (ETSI) has made two standards that define Electronic Signature Formats, ETSI TS 101 733[6] and ETSI TS 101 903[7]. Both: "defines a number of Electronic Signature Formats, including electronic signature that can remain valid over long periods. This includes evidence as to its validity even if the signer or verifying party later attempts to deny (repudiates) the validity of the electronic signature. (...) specifies use of trusted service providers (e.g. Time-Stamping Authorities), and the data that needs to be archived (e.g. cross certificates and revocation lists) to meet the requirements of long term electronic signatures." The main difference between the two standards is the format they use for defining the syntax of the signature. TS 101 733 uses ASN.1, whereas TS 101 933 uses XML. For long-term archival the ETSI-standards define a format called 'Archival Electronic Signature' (ES-A). Figure 10 (page 22) in TS 101 733 v.1.7.3 illustrates this format.

RFC 5126[8], CMS Advanced Electronic Signatures (CaDES), is technically equivalent to ETSI TS 101 733 v.1.7.4 and is an attempt at getting the ETSI specifications adopted by the IETF. RFC5126 states that "the technical contents of this specification is maintained by ETSI".

"The Internet Engineering Task Force (IETF) working group LTANS (Long-Term Archive and Notary Services) deals with the same topic and has already defined requirements, data structures and protocols for secure usage of archive services."[KOV06] LTANS[9] has issued RFC 4998[10] on Evidence Record Syntax (ERS) and RFC 4810 on Long-Term Archive Service Requirements.

In Norway, the main guideline on electronic signatures for long-term archival is the third deliverable from the SEID working group. It is about a data object for long-term archival and exchange of electronic signatures, named SEID-SDO[11] (Norwegian: Dataobjekt for langtidslagring og utveksling av elektroniske signaturer). The Norwegian BankID initiative (common eID and signature solution for Norwegian banks) has defined a "BankID SDO" that also aims at fulfilling requirements for long-term preservation of signed objects.

6.5 Shortcomings of Digital Signatures in Long-Term Perspective

As discussed in [Bla06], the initial enthusiasm generated by cryptographic signatures, which led many to praise it as intrinsically superior to handwritten signatures,[12] is usefully compared

6 ETSI TS 101 733: http://portal.etsi.org/docbox/EC_Files/EC_Files/ts_101733v010501p.pdf

7 ETSI TS 101 903: http://uri.etsi.org/01903/v1.2.2/ts_101903v010202p.pdf

8 RFC 5126: http://www3.tools.ietf.org/html/rfc5126

9 LTANS status pages: http://www3.tools.ietf.org/wg/ltans/

10 RFC 4998: http://www3.tools.ietf.org/html/rfc4998

11 SEID deliverable 3: http://www.npt.no/iKnowBase/Content/44963/SEID_Leveranse_3_v1.0.pdf

[12] The best example of this line of thinking is offered in [Ford, (W.), Baum, (M), Secure Electronic Commerce: Building the Infrastructure for Digital Signatures and Encryption, Upper Saddle River, NJ, Prentice Hall, 2000]: "Throughout history, lawmakers of both civil and common law jurisdictions have sought rules that achieve the type and level of non-repudiation made possible by digital technology. Signatures, seals, notaries, recording offices, and certified mail are all examples of traditional mechanisms employed in efforts to supply and bolster non-repudiation. … Explicit consciousness of this powerful issue has surfaced only very recently, as society has faced the challenge of first matching and then exceeding traditional legal protections in the emerging digital communications environment." (564)

alongside that generated by DNA profiling in criminal law. While this technology was initially granted a status of irrefutable proof of identification, it met with a surprising defeat during the course of the O.J. Simpson trial in 1995. As three sociologists of science explain, "[…] by following the samples from the crime scene to the laboratory, and then from the laboratory to the tribunal, one realizes that the genetic fingerprint may only serve its role of competent witness if and only if the succession of transactions during sampling, transport, preservation, digitization, and analysis of the sample is itself testified to by witnesses, certified and duly registered by responsible authorities. To be considered as such, the truth contained in the automatic signature (the genetic bar code) must be accompanied, surrounded by a whole series of bureaucratic traces: handwritten signatures on standard forms, actual bar-codes affixed on bags containing the samples, etc." [Lyn97]. It is those traces that were successfully contested during the Simpson trial, because, as archivists have long known, no evidence is ever self-intelligible.

The same principle applies to electronic records: in order to be a "competent witness" of a juridical fact (commitment to obligations), an electronic document must be accompanied by traces of all of the operations which it is susceptible to incur: creation, modifications, annotations, signature, conversion, transmission, etc. Likewise, digital signatures are unable to testify in and of themselves of the identity and integrity of a document, and to be effective, must also be accompanied by the various traces that testify to their own identity and integrity as evidence — public key certificates, revocation lists, certificate chains, audit trails, hash fingerprints, etc.

In the long-term perspective, there are several reasons why the evidential value of a Digital Signature will decrease. Notably, the following must be considered [ØlSe02]:

- Lifetime (expiry, revocation) of the keys and certificates used. The challenge is to verify that these were valid at the time of signing even if later expired or revoked. This requires a trusted time for a signature.

- Lifetime of the signing method, i.e. hash and cryptographic algorithms and size and quality of cryptographic keys. Given advances in technology, cryptography that is secure today is probably not secure over decades. In addition, flaws and weaknesses may be detected such as the 2006 attacks that effectively broke the MD5 hash algorithm

- Lifetime of formats of content, signature, signed data object, certificate, and other supporting information like time-stamps. Software to process the formats must be available, and format conversions necessarily invalidate the original signatures.

- Lifetime and continued service offer of (trusted and other) actors upon which the verification process relies. If the CA goes out of business, and its CRLs become unavailable (not accessible or impossible to verify), the verification process may fail.

In order to revalidate a digital signature, the state at the time of signing must either be captured in a reliable way or it must be possible to reconstruct the state. The ETSI standards for long-term SDOs aim at capturing state inside the SDO, while in other approaches one may opt for a solution where for example the revocation information valid at that particular point in time can be obtained when needed from a trusted source. Note in particular that the record cannot be converted to a different format, because then the bit stream of the record changes, and hence makes the signature validation impossible.

In his paper "The digital signature dilemma" Blanchette [Bla06] "argues that discrepancies between technical, legal and archival responses to the problem of long-term preservation of digitally signed documents are founded on diverging understandings — physical vs. contextual — of electronic authenticity." He concludes that "while legislation can provide a rich framework to support this engagement, efforts to dictate its precise rules are still premature at best. "

6.6 Preservation Strategies for Digitally Signed Documents

The goal in long-term conservation of digital signatures is to be able to demonstrate former validity of a signature. According to the guidelines from the National Archives and Records Administration [NARA guidelines] there are two main approaches to this:

- 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 [NARA guidelines], and to preserve the link or association between the various components of a signed record over time [ESRA guidelines].

The Norwegian standard for system requirements of digital archival systems in government administration [NOARK4] suggests re-signing as a solution. You can here either sign using a secret key linked to the post reception of the company, or chose to let the one making the conversion put his/her own signature. The latter is usually desired. It should be noted that NOARK does not put any requirements on the archival signature, which implies a possibility that it may be weaker than the original signature.

The ETSI standard TS 101 733 on electronic signature formats states: "It would be quite unacceptable, to consider a signature as invalid even if the keys or certificates were later compromised. 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." The standard 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[13]. 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 [KOV06].

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 Blanchette [Bla06]:

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

- 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" [Gil05].

Berbecaru et.al. describe a modular framework for concrete application of electronic signatures [BLMMR]. 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 [MiTa07] proposes a solution they call Signature Validity Extension. Their solution makes use of the long-term signature format in the standard RFC3126[14]. They claim the following main advantages over traditional solutions:

- 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.

- Since time stamp services are always provided using the cryptographic technology whose safety has been confirmed at the relevant point of time, obsolescence of the technology is not a concern.

14 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.

7 Authenticity in Long-term Digital Preservation

Evidential value is closely linked to the concept of authenticity which is a broader concept than authentication. Due to the of digital signatures in long-term perspective, it is even more important that sufficient metadata is supplied in order to keep authenticity past conversions (transformations) taking place during long-term preservation.

This chapter presents two different views on authenticity in a digital preservation context.

7.1 Rothenberg's Perspective on Authenticity

Jeff Rothenberg gives an in-dept discussion of the concept of authenticity in an article, [Rot00], from 2000 and the main points will be presented in the following.

Different disciplines may have their own explicit definitions of authenticity; however, in interdisciplinary discussions of authenticity, the dependence of a given definition on its discipline is often manifested only implicitly. The term authenticity is, according to Rothenberg, hard to define:

Its meaning is not restricted to *authentication*, as in verifying authorship, but is intended to include issues of integrity, completeness, correctness, validity, faithfulness to an original, meaningfulness, and suitability for an intended purpose.

The goal of preservation is to allow future users to retrieve, access, decipher, view, interpret, understand, and experience documents, data, and records in meaningful and valid (that is, authentic) ways.

If a paper document is "preserved" in such a way that the ink on its pages fades into illegibility, it probably has not been meaningfully preserved. Yet even in the traditional realm, it is at least implicitly recognized that informational entities have a number of distinct attributes that may be preserved differentially and some attributes may be lost. As an example of the latter, many statues, frescos, tapestries, illuminated manuscripts, and similar works are preserved except for the fact that their pigments have faded, often beyond recognition. Another example is that the original US Declaration of Independence has been preserved, but most of its signatures have faded into illegibility.

There is no accepted definition of digital preservation that ensures saving all aspects of such entities. By choosing a particular digital preservation method, we determine which aspects of such entities will be preserved and which ones will be sacrificed. For traditional informational entities, like old printed/written books, we can save the physical artefact; however, there is no equivalent option for a digital entity.

According to Rothenberg meaningful preservation:

- Implies usability in the future;
- Implies authenticity, in one way or another;
- Does not (necessarily) mean preservation of all attributes of a (digital) information entity.

7.1.1 Strategies for Defining Authenticity

Rothenberg describes some traditional strategies for defining authenticity, and introduces a new one. The traditional approaches are:

Originality strategy, focusing on whether an informational entity is unaltered from original state, this can follow several tactics:

- *Intrinsic properties tactic* – providing criteria for whether each property is present in its proper original form; a test against such criteria.

- *Process-centric tactic* – focus on the process by which an entity is saved, relying on its provenance or history of custodianship to warrant that the entity has not been modified, replaced, or corrupted and must therefore be original.

Intrinsic properties strategy, based on the intrinsic properties tactic above.

This involves identifying certain properties of an informational entity that define authenticity, *regardless* of whether they imply the originality of the entity.

Rothenberg says that the original-centric strategies work well in traditional preservation, but is problematic for digital informational entities.

The paper further discusses the process-centric tactic within the original-centric strategy as follows: "Relying on this tactic to ensure the authenticity of records involves two conditions: First that an unbroken chain of custodianship has been maintained; and second, that no inappropriate modifications have been made to the records during that custodianship. The first of these conditions is only a way of supplying indirect evidence for the second, which is the one that really matters. An unbroken chain of custodianship does not in itself prove that records have not been corrupted, whereas if we could prove that records had not been corrupted, there would be no logical need to establish that custodianship had been maintained. However, since it is difficult to obtain direct proof that records have not been corrupted, evidence of an unbroken chain of custodianship serves, at least for traditional records, as a surrogate for such proof."

The new strategy for defining authenticity, presented by Rothenberg, is as follows:

Suitability strategy, define authenticity in terms of whether an informational entity is suitable for some purpose or use.

This strategy would use various tactics to specify and test whether an informational entity fulfils a given range of purposes or uses. This may be logically independent of whether the entity is original.

Although the suitability of an entity for some purpose is presumably related to whether certain of its properties meet prescribed criteria, under this strategy both the specific properties involved and the criteria for their presence are derived entirely from the purpose that the entity is to serve.

Since a given purpose may be satisfied by means of a number of different properties of an entity, the functional orientation of this strategy makes it both less demanding and more meaningful than the alternatives[15].

The range of uses that an entity must satisfy to be considered authentic under this strategy may be anticipated in advance or allowed to evolve over time.

7.1.2 Authenticity Principles and Criteria

Due to the difficulty of defining authenticity abstractly, Rothenberg, instead suggests defining a set of *authenticity principles*: To encapsulate the overall intent of authentic digital preservation from different perspectives (legal, ethical, historical, artistic ...); further, the authenticity principle should be a functional statement from a specific perspective describing authentic preservation.

Only in cases when the authenticity principles are described as functional statements, can the authenticity principle be used in verifying whether a given preservation approach satisfies a given principle.

Further it would be ideal to device only a relatively small number of alternative authentication principles that captures perspectives of most disciplines/domains.

From each authenticity principle Rothenberg suggest to derive a set of authenticity criteria, to serve both as:

- Generators for specific preservation requirements, and

- Test of success of specific preservation techniques.

7.1.3 Authenticity Principles based on Expected Use Ranges

It is important to describe the range of expected uses of informational entities that is relevant to that discipline or organization, and derive authenticity criteria based on use. These descriptions should:

- Consist of a set of premises, constraints, and expectations for how particular kinds of informational entities are likely to be used;

- Include the ways in which entities may be initially generated or captured (in digital form, for digital informational entities);

- Include the ways in which they may be annotated, amended, revised, organized, and structured into collections or series; published or disseminated; managed; and administered;

- Describe how the informational entities will be accessed and used, either by the organization that generates them or by organizations or individuals who wish to use them in the future for informational, historical, legal, cultural, aesthetic, or other purposes.

[15] According to Jeff Rothenberg.

Different ranges of expected use may result in different authentication principles:

- One is that a given range of expected uses might imply the need for a digital informational entity to retain as much as possible of the function, form, appearance, look, and feel that the entity presented to its *author*.

- Another range of expected uses might imply the need for a digital informational entity to retain the function, form, appearance, look, and feel that it presented to its *original intended audience or readership*.

- A third range of expected uses may delineate precise and constrained capabilities that future users are to be given in accessing a given set of digital informational entities, *regardless of the capabilities that the original authors or readers of those entities may have had*.

Whereas retaining all the capabilities that authors would have had in creating a digital informational entity requires preserving the ability to modify and reformat that entity using whatever tools were available at the time, retaining the capabilities of readers merely requires preserving the ability to display, or render, the entity as it would have been seen originally. In the latter case capabilities might range from simple extraction of content to more elaborate viewing, rendering, or analysis, without considering the capabilities of original authors or readers.

It is possible to identify alternative authenticity principles that levy different demands against preservation. For example, the following sequence of decreasingly stringent principles is stated in terms of the relationship between a preserved digital informational entity and its original instantiation:

- same for all intents and purposes;

- same functionality and relationships to other informational entities;

- same "look and feel";

- same content (for any definition of the term);

- same description/metadata.

Rothenberg also states: "An authenticity principle must also specify requirements for the preservation of certain meta-attributes, such as authentication and privacy or security. For example, although a signature (whether digital or otherwise) in a record may normally be of no further interest once the record has been accepted into a recordkeeping system—whose custodianship thereafter substitutes its own authentication for that of the original—the original signature in a digital informational entity may on occasion be of historical, cultural, or technical interest, making it worth preserving as part of the "content" of the entity, as opposed to an active aspect of its authentication. Similarly, although the privacy and security capabilities of whatever system is used to preserve an informational entity may be sufficient to ensure the privacy and security of the entity, there may be cases in which the original privacy or security scheme of a digital informational entity may be of interest in its own right."

7.1.4 Definition of the Digital-Original Information Entity

Rothenberg finally presents a definition of the "original", the digital-original information entity, since the concept of an "original" is so pervasive in our culture. A *digital original* is defined to be:

- Any representation of a digital informational entity that has the maximum possible likelihood of retaining all meaningful and relevant aspects of the entity.

"This definition does not imply a single, unique digital-original for a given digital informational entity. All equivalent digital representations that share the defining property of having the maximum likelihood of retaining all meaningful and relevant aspects of the entity can equally be considered digital-originals of that entity. This lack of uniqueness implies that a digital-original of a given entity (not just a copy) may occur in multiple collections and contexts. This appears to be an inescapable aspect of digital informational entities and is analogous to the traditional case of a book that is an instance of a given edition: it is an original but not the original, since no single, unique original exists."

7.2 Gladney's and Bennett's Perspective on Authenticity

H.M.Gladney and J.L.Bennett also present an in-depth discussion of the concept of authenticity in an article, [Gla03a], from 2003 and the main points will be presented in the following.

The authors state that what makes the literature about authenticity confusing is that it often fails to declare which among the questions below it is addressing at each point, and that it makes unannounced shifts from one question to another. According to the authors it is prudent to separate questions that are distinct, such as: What do we mean by "authentic"? What do we mean by "evidence for authenticity"? What kinds of authenticity evidence might be available for something at hand? How can information producers create such evidence that will be useful long in the future? How can such evidence be preserved until it is wanted? To what extent should a consumer trust a document received and supporting evidence? Is the authenticity evidence sufficient for the application in question? How might evidence be different for different information genres (e.g., performances, reviews, written music) and different information representations (e.g., on musty paper, photocopy, digital representation, printed copy generated from a digital representation ...)? For a particular genre, what are the criteria of authenticity, and how can we test them?

7.2.1 Authenticity Criteria

Gladney and Bennett 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, they choose the following criteria:

- Distinguish as clearly as possible between objective facts and subjective judgments.

- Within the work represented by [Gla03b], 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.

7.2.2 Definition of Authentic, for Lossless Derivations/Transformations

The authors give the following formal definition of *authentic* in the case of faithful transformation, i.e. the existence of an inverse of every transformation function:

- 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, "$C(y) = T_n(\ldots (T_2(T_1(y)))$, with every Tk having an inverse"

- Then, if V is related to Y according to R

    - We say that V has *integrity* compared to Y,

- if R and S are true

    - We say that "by X as part of event Z" is a *true provenance* of V,

- if V has such *integrity* compared to Y and *true provenance*

    - We say that V is an *authentic copy* of Y

7.2.3 Definition of Authentic, for Lossy Derivations/Transformations

They also give the following formal definition of *authentic* in the case of:

- 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, " $C(y) = D_n(\ldots (D_1(y)))$, in which some D_k lose information"

- Then if V is related to Y according to R, and if C conforms to social conventions for the genre and for the circumstances at hand

    - We say that V has *sufficient integrity* relative to Y

- if R and S are true

- We say that "by X as part of event Z" is a *true provenance* of V

- if V has such (sufficient) *integrity* compared to Y and *true provenance*

  - We say that V is an *authentic copy* of Y

7.2.4 A Complete Definition of Authentic

In the general case Gladney and Bennett give the following formal definition of authentic:

- 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:, "$C(y) = T_n(... (T_2(T_1(y))))$"

- 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. Gladney describes trustworthy packaging for objects and metadata [Gla03c]. An object's accumulated metadata are the digital equivalent of a traditional audit trail for a physical archival holding.

7.2.5 Evidence in the Provenance and the Copy Functions

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 Dk(y).

For a signal that cannot be interpreted without representation information — particularly for a digital object — this extra information might further include the producer creating information to enable consumers' correct interpretation [Gla03d].

8 Security Services over Time

One main threat to the persistence of electronic records may well be accidental errors and events that may cause deletion, modification or other undesired effects. However, in order to maintain evidential value protection against such events (which is covered by the READ research area of LongRec) is not sufficient. The security, i.e. protection against deliberate modifications and deletions of records or their supporting information (such as metadata), must also be assured.

During the lifetime of a record, security technologies must be assumed to evolve and be replaced by new technologies. Ownership and authorizations will change. The entire environment of records, and thus the protection offered by the environment, must be expected to change. Then, how shall consistent security measures be ensured?

8.1 Authentication

Authentication is the process of verifying the digital identity of a process/computer and/or the physical identity of a person, i.e. user authentication. Authentication is thus the gatekeeper for other security tasks such as confidentiality, integrity, non-repudiation and availability.

Authentication is done at the time of access, from a user or between systems. The long-term aspects are thus not linked to the authentication process itself but rather to its traces as part of evidential value, e.g. at which times did a certain user log on to the system and what mechanism (with what strength) was used for authentication? This requires a certain trust in the logs of the system, including trust in the time stamp in the log.

The identity of all users and other entities that have authorizations to the system should be preserved even when these authorizations are revoked. Note that this may in cases be in conflict with privacy requirements. User names should not be reused. In the long-term perspective, this information may even have to be kept across technology changes, e.g. when a new archive system replaces the old one (see discussions for authorization and access control below).

Authentication mechanisms based on public key cryptography offer better evidence than other mechanisms because the user (or other entity) *signs* a challenge using a private key in the sole possession of the user. However, the time of the authentication still needs to be correct, and this form of digital signature must in principle be subject to the same considerations as a signed document.

8.2 Availability

Availability means ensuring timely and reliable access to and use of a record by authorized actors with a legitimate need. In a long-term perspective, accidental or deliberate deletion is the most severe threat, while denial of service attacks etc. can be considered as less important.

Recommendations for preservation unanimously point at the need for storing two or more copies of each record, whether this is in the form of backup copies or mirrored repositories. To protect against deliberate deletion, one could consider placing different copies under different governance regimes, e.g. by not having the same persons responsible for all copies or storing copies with a trusted third party (notary service).

Access control to the stored information is the sole most important measure to protect against unwanted deletion (and modification, see integrity below). Note that several copies also gives an attacker more opportunities to gain access to the (content of the) object.

Finally, availability is linked to success of the preservation process. If there no longer is software available for making you able to read/use both content and metadata, talking about availability becomes meaningless.

8.3 Authorization, Access Control, Ownership

Authorization means setting the rules for access to records (and to the systems holding them). Access control means enforcing access according to the authorizations. Two main access control models exist:

- Mandatory Access Control is used for information that is classified at different sensitivity levels such as classes of military or other national security relevant information. Information cannot be accessible at lower levels, and this is not under the control of the users (mandatory enforced). This is not considered further in this report although the model may be relevant in some cases.

- Discretionary Access Control is used for systems where the information owner, or any other user that is authorized to do so, 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:

- Identity based access control: Set authorizations for individual users and enforce accordingly.

- Role based access control: Authorizations are granted to defined roles, and users are assigned their relevant roles. A user having the appropriate role is granted access.

- Task based access control: Authorizations are assigned to the users, possibly also roles, involved in a particular instance of a process, and all users involved are granted access accordingly.

Role based access control is the most used model today in the world of record management. In this model, authorizations may be assigned in a three-step process, each of which may be performed by different actors for increased security:

- For each record, directly or by the record's membership in some kind of group, access rights are assigned to defined roles. This includes access to the IT-systems themselves.

- User identities must be managed – creation and other management of user accounts.

- Users are assigned to roles and thus assigned the authorizations implied by the roles.

It is recommended to perform such user and role management in a separate identity management system offering standard interfaces. This way, user and role management is common across all IT-systems, and an archive system (for example) may be replaced with the new system reusing the same interface and the same user and role information. For evidential value, the identity management system should also ensure:

- Preservation of historical information on roles and which users that were assigned to the roles "at all times". This may be a difficult task given today's identity management systems. Keeping the historical information with the identity management system may be OK but the information should be protected against later modification even by authorized administrators.

- Easy and reliable transfer of the information (upon replacement) to the next generation identity management system. Export and import are usually supported in such systems but transferring historical information (previous bullet point) may not be easy. Often, only a snapshot of the currently valid information will be transferred.

Given a well-managed role system, changes in the user population, roles and authorizations are more easily managed. Changing the authorizations allocated to a certain role may be more or less cumbersome, as will deletion or addition of new roles.

Note that role based access control should be used with identity based logging to ensure that the individual user and not only the role can be held responsible for actions. User identities should be persistent and not be reused.

Ownership to a record is a very important property and must be well defined. The easiest model is where ownership is transferred from the originator to (a role associated with) the repository. If this cannot be done, the owner must have some way of authorizing accesses to the object, and to change such authorizations (like declaring an object public). Restrictions on the owner's access, such as for deletion or modification, must be determined and enforced, and the same applies to authorizations and restrictions that pertain to roles associated with the repository. There is a need to define rules that apply when the owner is no longer in a position to act, e.g. he or she is dead.

The evolving and rapidly changing digital environment in which digital objects reside suggests that references to these objects have a high probability of becoming inoperable in a few short years. Therefore each record in a repository should be assigned a persistent identifier (PID); a name that can be used in perpetuity to refer to and retrieve the record. This simplifies authorization and access control.

Another aspect of authorization and access control is intellectual property rights and digital rights management, see below.

8.4 Confidentiality/Privacy

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 the permission of the owner.

The primary method for maintaining confidentiality/privacy of stored information is access control. Encryption can also be used, but this must be done with care to ensure that one does not lose the ability to decrypt the information, which will then be lost. Encryption must be used when confidential information is transmitted over networks.

Since compliance with privacy rules can often depend on factual circumstances only manifest after a given query has been made it is simply impossible to rely on control over query (data collection rules) alone to protect privacy; provision of transparency and accountability to rules and policies can be used to achieve a second level for protection of privacy.

8.5 Integrity

Integrity means ensuring that a data record is accurate, complete, and not modified in an unauthorized way. The single most important measure to ensure integrity of stored information is access control.

Additional protection is provided by checksums that may be applied to individual records, files or disk structures. There are many flavours of checksums from simple ones to cryptographic methods such as message authentication codes (MAC), digital signatures, cryptographic checksums or keyed hashes. Digital signatures may be said to be the most powerful checksum since it also authenticates the signer. Watermarking based methods can also be mentioned and different mechanisms may be used in combination.

Checksums are used to *detect* modifications but one cannot regenerate the correct content. Also, a simple checksum stored with the object does not provide protection against an adversary that can change both the object and the checksum. Algorithms for error correcting codes exist but at the expense of a large storage overhead, and even in this case an attacker may be able to change both the object and the checksum/code. Cryptographic checksums protect against unauthorized changes in checksums.

Checksums may also be stored separately from the objects and under special protection. This is commonly used for system integrity (e.g. Tripwire) and may be used for object/record integrity as well. One may then run regular integrity checks or compute checksums at time of access of an object.

The best protection when reconstruction is needed is to store several copies of each record in separate systems under separate administration and possibly also in separate locations. If checksums are applied in a way that ensures detection of modifications, two copies may be sufficient. Else, at least three copies should exist in order to enable a majority vote to determine the correct version.

8.6 Non-Repudiation

ISO's non-repudiation framework[16] defines non-repudiation as: "The goal of the NR service is to collect, maintain, make available and validate irrefutable evidence concerning a claimed event or action in order to resolve disputes about the occurrence of the event or action."

"Irrefutable evidence" probably does not exist neither in the digital nor the real world, and thus the term non-repudiation may go out of use; e.g. in the key usage extension of X.509 certificates there is no longer a "non-repudiation bit" but rather a "content commitment" bit. As a consequence, this report does not use the non-repudiation term but rather uses the terms "evidential value" and "authenticity".

In essence, non-repudiation means authentication/authenticity and integrity preserved over time. ISO's non-repudiation framework points in particular at use of digital signatures and trusted notary services as services that can provide non-repudiation. As discussed in chapter 6 a digital signature in itself may not necessarily provide sufficient evidential value unless other preconditions are met.

If a notary service is used as a deliberate mechanism in some interaction (like an electronic commerce protocol), the notary must ensure that is runs a secure, trusted repository/archive as discussed in this report. The notary's conditions for delivery of material, in particular the metadata, formats and authenticity requirements, must be made clear.

8.7 Authenticity

Authenticity is defined by the InterPARES[17] as: "The trustworthiness of a record as a record; i.e, the quality of a record that is what it purports to be and is free from tampering or corruption." This quality is attributed to an original or a true and faithful copy. It includes issues of integrity, completeness, correctness, validity, faithfulness to an original, meaningfulness, and suitability for an intended purpose. Validating authenticity entails verifying that a record is indeed what it claims to be, or what it is claimed to be by external metadata.

Authenticity must involve the entire process from submission of information to a repository, creation of the record containing the information and the necessary metadata, and security and reliability of the stored information record. Validation of the information at time of submission is crucial. This includes secure transmission and authentication but may also extend into requirements on the processes producing the information, such as ensuring who is the author or owner of the information. Authenticity is strongly linked to accountability as discussed below.

Cryptographic mechanisms such as digital signatures may be used to enhance authenticity properties of a record. Version control is also a useful tool for preserving the authenticity of digital records.

8.8 Accountability/Auditing

Accountability is the ability to provide a report, explanation, or justification of decisions, events, actions, conditions, or understandings. Records auditing allows organizations to maintain accountability with regard to the use of protected documents, because they can know precisely:

[16] ISO/IEC 10181-4 *Information technology - Open systems interconnection - Security frameworks in open systems - Part 4: Non-repudiation framework*". ISO/IEC,1996

[17] InterPARES II glossary: http://www.interpares.org/ip2/ip2_term_pdf.cfm?pdf=glossary

- How a recipient has used a record;
- How often each type of usage occurred;
- When that usage occurred.

The activities are audit trail/logging and tracking each use of a record (session and transaction logs), maintenance of logs over time, accessibility (including search and retrieval) of log information over time, and accountability of logs (their security and trustworthiness). Logs may be kept separate from the repository system and subject to a separate administrative regime if desired in order to increase protection of logs. Cryptographic mechanisms such as digital signatures may be used to increase log security (logs signed by system or system administrator). In addition to logs, some events may be recorded in (the metadata of) the records themselves, e.g. actions implying modifications.

For evidential value one may in principle want to keep logs for eternity. To increase lifetime of logs, they must be in an open, system independent format that can survive even if the logging system is replaced. Logs may in this respect be treated as information records subject to the same measures and with the necessary metadata attached. Note that privacy regulations and concerns may limit allowed lifetime of log information. Note also that for some systems the sheer volume of the log information makes it difficult to use, and that log analysis tools are needed. This is a general problem and not in particular related to long-term aspects.

For log information to be useful over time, names of users, records and roles must be eternal and not reused.

The goal of audit is 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 from an economical and organizational perspective. Thus an audit trail provides supporting aggregated information about the records being stored and ensures that one can demonstrate that records are authentic. The supporting information should include:

- The name of the author and/or owner of the information record;
- The time the record was stored;
- The names of actors who accessed or made changes to the document and the roles and/or processes in force for the accesses;
- Details of the changes made to the record; version information;
- Details of movement of the record from medium to medium, and from format to format;
- The authentication measures used when the file is moved;
- Evidence of the controlled operation of the system in which the record is stored.

8.9 Intellectual Property Rights/DRM

Intellectual property rights (IPR) issues, in a broad sense, can if not properly addressed impede long-term storage and access activities. Ownership and IPR necessarily change over time, must comply with relevant laws and regulations, and dictate management of authorizations given to other actors. IPR can be considered as parts of the context information of a record. IPR has a substantial impact on a trusted, long-term storage of digital documents. Simply copying digital documents onto another medium, encapsulating content, or converting content to new formats or platforms, all involve activities which can infringe IPR. Some of the additional complexity in IPR issues relates to the fact that digital documents are also easily copied and re-distributed. Rights holders are therefore particularly concerned with controlling access and potential infringements of IPR [Abi03a-b, Abi04a-c. Abi05, Abi07]. Therefore engendering trust among customers is essential in long time preservation of rich media.

9 Evidential Value

Different aspects of evidential value are briefly discussed in the following, ranging from different threats to evidential value, including loss of authenticity, to evidential value within the juridical system. Here different technical approaches might result in different outcomes.

9.1 Threats to the Evidential Value over time

There are varieties of events related to issues such as file formats, protocols, software implementations, or encryption algorithms that may affect evidential value over time and ways in which to guard against such events are the primary requirements to be met. Like any other records storage the threats to the preservation of evidential value include [Mas06]:

- Malicious Modification or Destruction - For any of a number of reasons, long-term storage may be subject to attack. It is insufficient to rely on multiple copies without mechanisms for ensuring that a concerted attack on storage sites will not cause information to be destroyed or changed.

- Loss of Interpretability - Because the goal is to keep retrievable records of human activities, and not just sequences of bits, it is important to ensure that the data stored can be interpreted in the future. Thus, a long-term record keeping needs to be concerned with the data formats used and not just bit sequences.

- Loss of Context - In a large archive of data, each record must be specifically identified and its context supplied. Often, in day-to-day use of records, this kind of contextual information is implicit—not explicitly represented and therefore at risk of being later forgotten.

- Loss of Guarantee of Authenticity - Records of transactions may be subject to manipulation. Simple maintenance processes or media refresh may cause loss of clues about record origins or dates and interfere with processes to manage chain of custody of records which might otherwise be used to determine authenticity.

- Authorization Failure - There are situations where multiple data storage locations might still be subject to centralized regulation, legal intervention, or other events affecting archived data even though the effects (such as data change or destruction) might be contrary to the original wishes of the principal creating the archive.

- Loss of Confidentiality - Almost every organization has some records that are confidential. Secrecy of the content or even the existence of such records may be essential to evidential value, and may be weakened if breaches in confidentiality occur.

9.2 Judicial Challenges- an Example from Belgium

An example, from Belgium, presented in [Dek05] illustrates how the outcome in court might be very different depending on the strategy of preservation chosen by three (fictive) persons called Alice, Bob, and Carl.

"One might expect that the law remains indifferent to the way in which evidence is preserved, as long as its authenticity can be demonstrated to the courts. As a rule, this approach is followed in criminal cases and generally regarding the proof of all matters of fact. In these cases, any and all evidence is admissible regardless of its form The records preserved by Alice, Bob and Carl can all be presented in court, leaving it to the judge to make up his own mind whether these records are convincing.

In many jurisdictions, limits apply to the admissibility of evidence where legal transactions are concerned. The reasoning being that the parties generally plan these transactions beforehand and are in a position to document the process in a reliable fashion. The parties are not allowed to burden the courts with shaky evidence unless they have a good excuse.

In Belgium, a signed document – in original form – must be presented for all private agreements exceeding the value 375 €.15 'Original form' means the document that features the parties' original signatures. In the case of paper documents, it means the piece(s) of paper that was (were) in the hands of the parties and signed by them. In the case of digital documents, the meaning of the term 'original' is not as clear cut. Certain is that the digital file with digital signatures as appended by the parties qualifies as an 'original'. The advantage of presenting an original as evidence of an agreement is that it constitutes sufficient proof of the terms of the agreement on its own. Alice, Bob and Carl all start out with such original documents.

Carl does not preserve the digital originals, but replaces them with copies on paper. In principle these copies do not carry any weight in court. However, if the adversary does not demand that an original is produced, the judge may not request this of his own accord. The copy will be treated as if it were an original to prove the agreement. Even if the adversary protests the lack of proper evidence, all is not lost. Subject to certain conditions, any written piece of evidence may be presented. Of course, a document that you created yourself is not evidence, only documents originating from the adversary count. Moreover, to compensate for the lack of a signature, supporting evidence must be provided. So Carl must demonstrate that he has a paper copy of an electronic original emanating from his adversary and confirm its contents with additional evidence.

Alice discards all digital signatures and replaces them with metadata. As such, metadata containing authentication data can be considered a simple electronic signature. The judge may not disregard this signature just because it is in electronic form or is not a QES. Belgian law may very well reject this generic electronic signature because it is not the original signature as it was created by the signatory. Alice's documents will be treated as copies, just like Carl's. The fact that Alice re-signs the converted documents has no bearing on their status as copies, neither does the involvement of a third party in the archiving process.

Bob makes every effort to preserve his signed documents in their original form. If successful, he can present proper documentary evidence to a judge and has fulfilled his burden of proof completely. Problems may arise if the digital signatures can no longer be validated due to bit degradation or incompleteness of the validation chain. Though the document loses its status of an original, it may still be regarded as a copy if all the conditions are met. The situation is worse if the record is no longer readable and no emulator is available. Little does it matter that the digital signature is valid if the underlying document is inaccessible. From a legal point of view, having a converted version of the record as well, doesn't change the fact that the original is illegible. Bob can of course present the converted record as evidence in its own right, though it too will be treated as a copy.

The insistence on preservation of 'originals' is not unique to Belgian Law. Other European countries have similar rules in place. Nor is the predilection for originals limited to contractual law. An illustration of both these points is found in the Directive on Electronic Invoicing.

In order to have a valid invoice, the authenticity of its origin and the integrity of its content must be guaranteed. These goals can be achieved through a multitude of archiving strategies. Member States may however demand that all invoices be preserved in their original form. Belgium is one of the countries that have used this option, thus barring Carl's hard copy strategy for invoices. Moreover, the Directive favours two techniques for the creation of valid invoices, namely the advanced electronic signature or an EDI format agreed upon by the parties. Only these two techniques must be accepted by Member States, others are optional. Bob's strategy of preserving original records complete with signatures fits neatly into this framework. Alice, however, may run into trouble. In Belgium, it seems that converting invoices runs afoul of the obligation to preserve invoices in the form in which they were received. Discarding the signature of an invoice authenticated by AES is not allowed either."

This example show that the law's appreciation of the different preservation strategies does not necessarily correspond with archival criteria. This should be kept in mind when selecting a preservation strategy.

9.3 Authenticity Challenges

Jean-François Blanchette discusses, in his paper [Bla06] issues and problems related to digital signatures, authenticity, and evidential value of digital/electronic documents. He says that the gap between the responses offered by the legal, technical and archival community over the long-term preservation of digitally signed documents is best understood as a clash between two differing conceptions of electronic authenticity.

The first, espoused by the technical community and adopted by some segments of the legal community, is based on the measurement of a physical property of the document — bitwise integrity — whereby "data has not been altered in an unauthorized manner [i.e., by insertion, re-ordering, inversion, substitution, or deletion of bits] since the time it was created, transmitted, or stored by an authorized source" [Men96].

The appeal of such a measure lies in the hope that authenticity may become susceptible to precise quantification, to be given a simple "thumbs up" or "thumbs down".

From the point of view of the archival mission, such a physical measure of authenticity is highly useful at specific points in the document lifecycle — for example, when transmitting documents across space.

However, as the primary method for establishing authenticity, it effectively compounds the preservation problem[18] [Dur02]. Archivists prefer to rely on a second conception of electronic authenticity, one best described as contextual, which documents the totality of the controls and procedures, whether human or computer-based, that insure the identity and integrity of an electronic record throughout the totality of its lifecycle[19].

The lesson here is that criteria for electronic authenticity will not be established by a technological silver bullet [And94]. Just as signatures themselves were once technological novelties around which social practices gradually coalesced [Fra92], the evidential value of electronic documents will emerge out of the slow and gradual engagement of relevant social groups with the various technical means supporting claims of authenticity. While legislation can provide a rich framework to support this engagement, efforts to dictate its precise rules are still premature at best.

10 Concluding Remarks

This report has presented a broad go-through of the different elements involved when discussing trustworthiness in long-term preservation, related to the concept of evidential value of digital records.

In order to provide a high evidental value of an electronic record, enough traces of authenticity has to be kept and maintained for decades and centuries in order to be capable of a successful authenticity validation in the future. Authenticity includes issues of integrity, completeness, correctness, validity, faithfulness to an original, meaningfulness, and suitability for an intended purpose [Rot00]. Validating authenticity entails verifying that a record is indeed what it claims to be, or what it is claimed to be by external metadata.

Digital signatures are increasingly been used within the juridical system for their high evidential value, proving identity, authorisation, and integrity. With politics geared towards widespread acceptance of electronic communication, some types of digital (electronic) signatures have been reflected in laws and regulations, as equals to written signatures [ØlSe02]. But in the long-term perspective digital signatures have proven unable to testify the identity and integrity of a digital document (object) over time. Digital signatures are related to the bit stream and when the bit stream is changed, as it is during conversion, the main role of the signature is lost [Bla06].

A main challenge in the long-term perspective is therefore to put enough evidence of authenticity in the metadata accompanying the digital content intended for long-term storage. Successfulness is very much dependent on providing enough authenticity related metadata at ingest, when the digital object/document enters a repository. What is sufficient metadata depends on what definition of authenticity is used. It is not fruitful to add more and more metadata without an accompanying definition of authenticity.

[18] This is what the InterPARES research project expressed when declaring that "it is impossible to preserve an electronic record as stored physical object; it is only possible to preserve the means to make this document manifest".

[19] Criteria for this type of context-based authenticity have been offered by the InterPARES research project as benchmark et baseline requirements.

The quality of authenticity has to survive every conversion (migration) from one content bit stream to another, and every conversion imposes a threat to authenticity. Fewer conversions (transformations) imply less risk of loosing authenticity. By using a simple definition of authenticity [Gla03a] suggests a minimal definition of authenticity, separating subjective and objective matters where the latter can be (partly) automated. This may ease the way to handle authenticity when deriving new bit streams.

Keeping the same content in several derived bit streams and being able to compare the content gives some assurance of the authenticity. This is applied in several approaches being implemented today [Thi02] [Bou05b], but a comprehensive visual and functional comparison will be difficult if the digital content is of a complex nature. Still it is better to have more derivations to compare against than to have only one. The same applies of course to have several backups with redundancy and diversity in storage media.

Encapsulating both the original digital content (bit stream) and associated metadata, together with (all) derivations of content and associated metadata, seems to be fruitful [Bou05b] [Moo00] [Gla08]. XML is a preferred language for encapsulation. Encapsulation reduces the risk of loosing valuable metadata, but there has to be associated policies on what to be stored within an encapsulated object, e.g. software for accessing the content should be stored separately. There is a trade-off between encapsulation and searchability which may result in duplication of information into several subsystems, but from an authenticity perspective encapsulation with associated checksum and unique identity seems to be a good idea.

All through the lifetime of a stored digital record there are security threats, regardless of the perspective of the storage period. Both content and metadata faces unauthorised modification or deletion as major threats. In the case of long term preservation, security services should be picked based on what is suitable in the long term. Logs of actions on the digital records are necessary to achieve a high evidential value. There is also a distinction between stable documents, and semi-stable documents. An example of the latter is patient journals to be preserved during the lifetime of the person. Here privacy and confidentiality plays important roles.

11 Bibliography

[Abi03a] H. Abie, B. Blobel, J. Delgado, S. Karnouskos, R. Marti, P. Pharow, O. Pitkänen, and D. Tzovaras, *DigiRight: Relevance to and Potential Impact on Europe's Need to Strengthen the Science and Technology Excellence on DRM*, First International Mobile IPR Workshop: Rights Management of Information Products on the Mobile Internet, IEEE, pp. 127-135, Helsinki, Finland, 2003.

[Abi03b] H. Abie, J. Bing, B. Blobel, J. Delgado, B. Foyn, S. Karnouskos, P. Pharow, O. Pitkänen, and D. Tzovaras, *DigiRight: Network of Excellence for a Framework for Policy, Privacy, Trust and Risk Management for Digital Rights Management*, First International Mobile IPR Workshop: Rights Management of Information Products on the Mobile Internet, IEEE, pp. 117-127, 27-28, Helsinki, Finland, 2003.

[Abi04a] H. Abie, P. Spilling, and B. Foyn, *Rights-Carrying and Self-Enforcing Information Objects for Information Distribution Systems*, in the Proc. of ICICS'04, Malaga, Spain, LNCS 3269, pp 546-561, ISBN: 3-540-23563-9, Springer Berlin / Heidelberg, 2004.

[Abi04b] H. Abie, B. Foyn, J. Bing, B. Blobel, P. Pharow, J. Delgado, S. Karnouskos, O. Pitkänen, and D. Tzovaras, *The Need for a Digital Rights Management Framework for the Next Generation of E-Government Services*, International Journal of Electronic Government, Vol. 1 No.1, pp 8-28, Inderscience Publishers, (ISSN: Print 1740-7494, Online 1740-7508), 2004.

[Abi04c] H. Abie, P. Spilling and B. Foyn, *A Distributed Digital Rights Management Model for Secure Information Distribution Systems*, International Journal of Information Security (IJIS), Springer-Verlag, 2004.

[Abi05] H. Abie, A. Skomedal, *A Conceptual Formal Framework for Developing and Maintaining Security-Critical Systems*, IJCSNS International Journal of Computer Science and Network Security, Vol.5, No.12, 2005

[Abi07] H. Abie, *Frontiers of DRM Knowledge and Technology*, IJCSNS: International Journal of Computer Science and Network Security, Vol.7, No.1, pp 216-231, 2007.

[And94] Anderson, (R.), *Why Cryptosystems Fail*. CACM, 37(11), pp. 32-40, 1994.

[Bla06] Jean-François Blanchette, *The digital signature dilemma*, 2006.

[Bou05a] Filip Boudrez, *Digital Signatures and Electronic Records*, retrieved from http://www.expertisecentrumdavid.be/docs/digitalsignatures.pdf, 2005.

[Bou05b] Filip Boudrez, Digital *Containers for Shipment into the Future*, retrieved from http://www.expertisecentrumdavid.be/docs/digital_containers.pdf, 2005.

[BLMMR] Berbecaru, D., Lioy, A., Maino, F., Mazzocchi, D., Ramunno, G., *Towards concrete application of electronic signature,* Italy.

[Dek05] Hannelore Dekeyser, *Preservation of Signed Electronic Records*, DLM Conference, Budapest 5-7 October 2005.

[Dur02] Duranti, L., et al., *Strategy Task Force Report, in The Long-term Preservation of Authentic Electronic Records*, Vancouver, InterPARES, 2002.

[EUDIR99] *Directive 1999/93/EC of the European parliament and of the council of 13 December 1999 on a Community framework for electronic signature,* 1999.

[EUREP06] *Report on the operation of Directive 1999/93/EC on a Community framework for electronic signatures*, 2006.

[Fra92] Fraenkel, (B.), *La Signature: Genèse d'un signe*, Paris, Gallimard, 1992

[Gil05] Gilliland-Swetland (A.), *Electronic Records Management*. ARIST 39, pp. 219-25, 2005.

[Gla03a] H.M.Gladney and J.L.Bennett, *What Do We Mean by Authentic? -What's the Real McCoy?*, D-Lib Magazine, Volume 9 Number 7/8, ISSN 1082-9873, 2003, retrieved from http://www.dlib.org/dlib/july03/gladney/07gladney.html.

[Gla03b] H.M. Gladney, *Trustworthy 100-Year Digital Documents: Executive Summary of a Digital Preservation Proposal*, 2003.

[Gla03c] H.M. Gladney, *Trustworthy 100-Year Digital Documents: Evidence Even After Every Witness is Dead*, 2003.

[Gla03d] H.M. Gladney, *Trustworthy 100-Year Digital Documents: Durable Encoding for When It's Too Late to Ask*, 2003.

[Gla08] H.M. Gladney, *Durable Digital Objects -Rather Than Digital Preservation*, DRAFT, 2008, retrieved from http://home.pacbell.net/hgladney/hmgpubs.htm.

[IDABC07] *Preliminary study on mutual recognition of eSignatures for eGovernment applications,* IDABC study, 2007.

[Jøs96] A.Jøsang, *The Right Type of Trust for Distributed Systems*, Proceedings of the 1996 New Security Paradigms Workshop, 1996.

[KOV06] Kunz, T., Okunick, S., Viebeg, U., *Long-term security for signed documents: services, protocols, and data structures*, 2006.

[Lyn97] Lynch, (M.), McNally (R.), Daly, (P.), Le tribunal : *Fragile espace de la preuve.* La Recherche, 300, pp. 112-115, 1997.

[Mas06] Larry Masinter, Michael Welch, *A System for Long-Term Document Preservation*, Adobe Systems Incorporated; San Jose, CA, IS&T Archiving 2006 conference, 2006.

[Men96] Menezes, (A.J.), van Oorschot, (P.C.), & Vanstone (S.A.), *Handbook of Applied Cryptography*, Boca Raton, FL, CRC Press, 1996.

[Moo00] Moore, Reagan et al., *Collection-Based Persistent Digital Archives—Part 1*, D-Lib Magazine 6(3), 2000, retrieved from http://www.dlib.org/dlib/march00/moore/03moore-pt1.html.

[MiTa07] Miyazaki, K. and Tanaka, M. *EVERSIGN: Preserving the Long-Term Authenticity of Electronic records*, 2007.

[RLG02] RLG/OCLC Working Group on Digital Archive Attributes, *Trusted Digital Repositories: Attributes and Responsibilities*, RLG-OCLC Report, 2002, retrieved from www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf.

[Ros06] Heiko Roßnagel, *On Diffusion and Confusion - Why Electronic Signatures Have Failed*, 2006.

[Rot00] Jeff Rothenberg, *Preserving Authentic Digital Information*, in Authenticity in a Digital Environment, ISBN 1-887334-77-7, Council on Library and Information Resources, Washington, D.C., 2000.

[Study07] *Study on the standardisation aspects of eSignature* (22/11/2007), 2007.

[Thi02] Kenneth Thibodeau, *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*, in -The State of Digital Preservation: CONFERENCE PROC-EEDINGS, ISBN 1-887334-92-0, CLIR, Washington, 2002.

[TRAC07] *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.

[Wang06] Wang, Minyan, *A review of electronic signatures regulations: do they facilitate or impede international electronic commerce? I ICEC*, 548-552, 2006, retrieved from http://doi.acm.org/10.1145/1151454.1151458.

[Øln01] Ølnes, J., *A Taxonomy for Trusted Services*, First IFIP Conference on e-Commerce, e-Business, e-Government (I3E), Zurich, 2001 .

[ØlSe02] Ølnes, J., Seip, A.B., On *Long-term Storage of Digitally Signed Documents,* Second IFIP Conference on e-Commerce, e-Business, e-Government (I3E), Lisboa, 2002.

[Ølnes07] Ølnes, J, Andresen, A, Buene, L., Cerrato, O., Grindheim, H. *Making digital signatures work across national borders*, ISSE/SECURE Conference, Warszawa, 2007.

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

[ESRA guidelines] *Electronic Signatures and Records Act (ESRA) guidelines*, 2007.

12 Appendix A: Standards

Below is listed some of the standards that are applicable within the domain of long term digital preservation, addressing trustworthiness at some level:

The ISO 9000 family of standards addresses quality assurance components within an organization and system management that, while valuable, were not specifically developed to gauge the trustworthiness of organizations operating digital repositories.

Similarly, the ISO 27000 family of standards is developed specifically to address data security and information management systems. Like ISO 9000, it has some very valuable components to it but it was not designed to address the trustworthiness of digital repositories. Its requirements for information security seek data security compliance to a very granular level, but do not address organizational, procedural, and preservation planning components necessary for the long-term management of digital resources.

ISO 15489-1:2001 and ISO 15489-2:2001 defines a systematic and process-driven approach that governs the practice of records managers and any person who creates or uses records during their business activities, treats information contained in records as a valuable resource and business asset, and protects/preserves records as evidence of actions. Conformance to ISO 15489 requires an organization to establish, document, maintain, and promulgate policies, procedures, and practices for records management, but, by design, addresses records management specifically rather than applying to all types of repositories and archives.

ISO 15801:2004, "Electronic imaging- Information stored electronically- Recommendations for trustworthiness and reliability", describes the implementation and operation of information management systems which store information electronically and where the issues of trustworthiness, reliability, authenticity and integrity are important. The whole life cycle of a stored electronic document is covered, from initial capture to eventual destruction. It does not cover processes used to evaluate the authenticity of information prior to it being stored or imported into the system. However, it can be used to demonstrate that output from the system is a true reproduction of the original document

ISO 14721:2002, the Open Archival Information System Reference Model, OAIS, provides a high-level reference model or framework identifying the participants in digital preservation, their roles and responsibilities, and the kinds of information to be exchanged during the course of deposit and ingest into and dissemination from a digital repository. OAIS will be described in depth in another appendix of this report.

13 Appendix B: OAIS Concepts

13.1 OAIS Architecture

Standardization at different levels is a valuable contribution to establishment of trustworthiness in general. Several standards could be mentioned, but the most important one is the ISO reference model for Open Archival Information System, called OAIS, ISO 14721:2003. In this section, the key concepts of the OAIS model[20] are presented. The main motivation for going into details is to increase the readability of the discussions in this report.

The Open Archival Information System (OAIS) presents views of the archive (and archival process) at different levels. At its highest level, it may be viewed as a black box receiving content from producers and sending content to consumers. Inside the black box there are a number of processes, which transform the material received into an archival form. We are not presenting these functions and processes any further, but are only focusing on the presentation of the different information object and packages involved.

The architecture of the OAIS reference model[21] is shown in Figure 6 below.

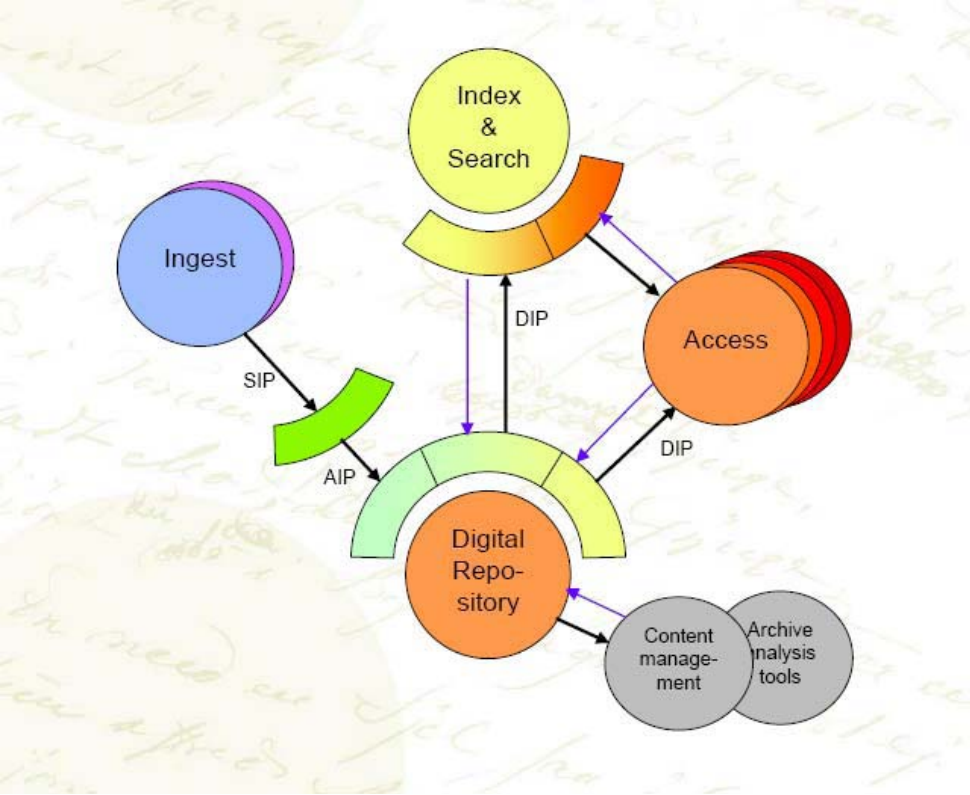

Figure 6: OAIS reference model.

[20] Most of the sections are reproduced from James Currall, Claire E. Johnson, Pete Johnston, Michael S. Moss, Lesley M. Richmond "No Going Back?" The final report of the Effective Records Management Project, 2001, www.gla.ac.uk/infostrat/ERM/Docs/ERM-Final.pdf with information added.

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

As shown the model references three different packages. These are further described below:

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

The OAIS reference model defines six areas of concern:

- Ingest (the process that accepts submissions from producers and transforms this into a representation (AIP) suitable for the repository)
- Data Management
- Archival Storage
- Administration
- Preservation Planning
- Access

The relations between the areas are shown in Figure 7 below.

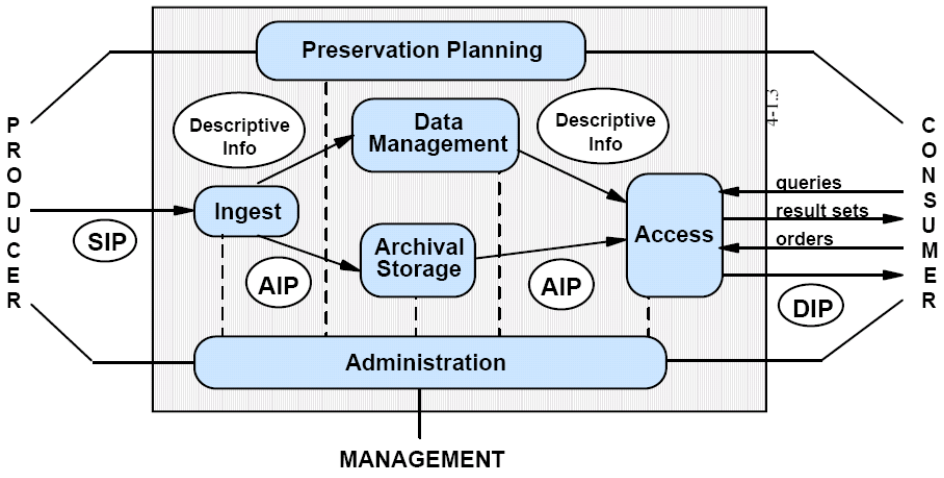

Figure 7: Areas of concern in the OAIS model, and their relationships.

13.2 Information Object

Figure 8 below illustrates the OAIS Information Object. An Information Object is composed of two components:

- Data Object
- Representation Information

The Data Object is the digital file that is to be archived. An example might be a Microsoft Word file containing the minutes of a particular meeting. According to the OAIS reference model, the Data Object may also be a physical object but this is not considered relevant here.

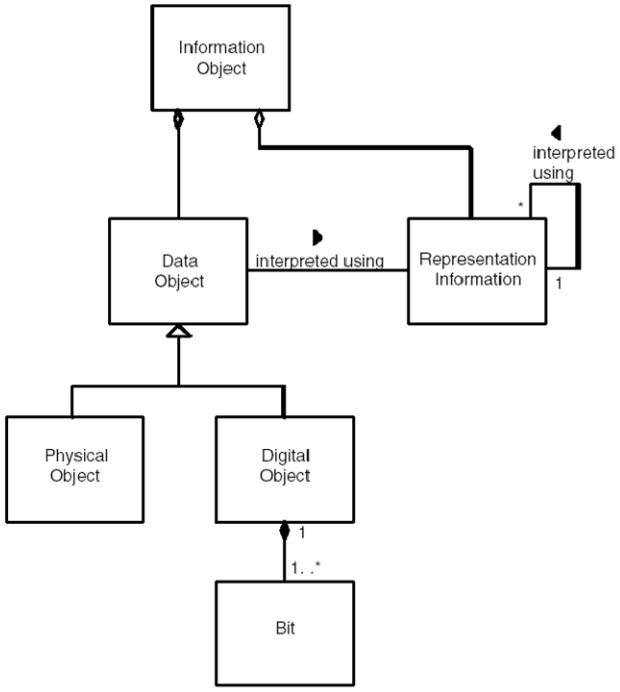

Figure 8: An OAIS Information Object, reproduced from figure 4-10 in the OAIS Reference Model

The Representation Information is the information which is required to be able to access the Data Object. In the example above, this includes the fact that it is a Microsoft Word 97 format file. The information must be sufficient to enable the data object to be interpreted and its content rendered in an intelligible form. In our case there must be a reference to either a full description of the file format (at the bit level) or a means of rendering it, which is available to users of the archive (for as long as the data object must be accessible).

Representation Information may involve 'indirection' in that the Representation Information actually stored with our Microsoft Word file needs to identify it as that type of file but may point elsewhere in the archive to the detailed description of the format which needs to be recorded only once for each file type. Initially the detailed description may indicate where a copy of the software may be found, but this may be changed later if that software is no longer available and other approaches need to be taken. There is of course a recursive element to Representation Information. The Representation Information itself will be stored (as a Data Object) in some file format (perhaps as ASCII text) and there must be Representation Information for that also.

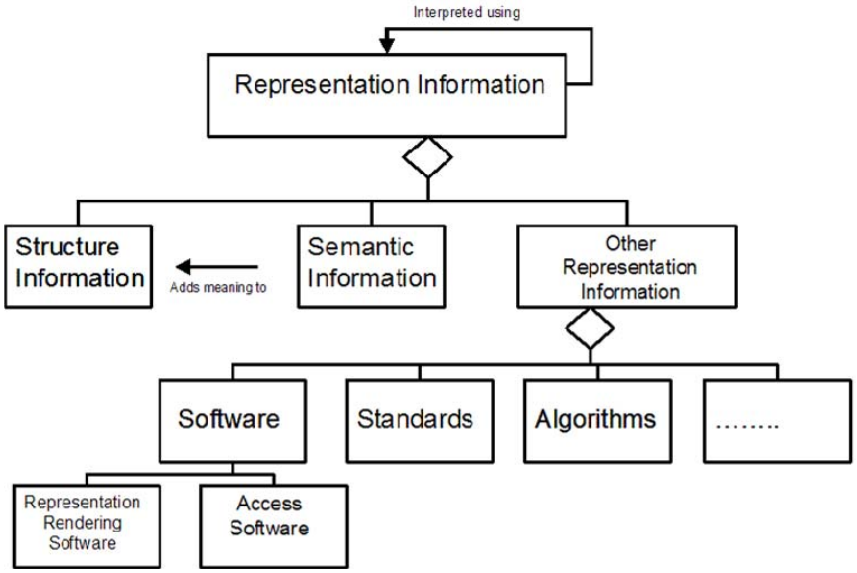

Figure 9: Types of Representation Information in the OAIS Reference Model

The relationship between the *Data Object* and *Representation Information* may be summarised as:

Data Object -- *interpreted by* --> Representation Information -- *yields* --> Information Object

13.3 Types of Information Objects

Two types of Information Objects are presented below.

13.3.1 Content Information

An Information Object containing content information is perhaps the primary information of interest. The Content Information is an Information Object which therefore contains a Data Object (in our example, the minutes themselves) and their Representation Information.

13.3.2 Preservation Description Information

In order for the Content Information to make sense, it is necessary to have additional information about the content (minutes), which will enable readers in the future to understand their context and the degree of confidence they may have in the content. These 'metadata' are termed Preservation Description Information in the OAIS model. They provide information in four areas:

- Reference - identifies what the content is - basic description and metadata;
- Provenance - describes the creation environment of the content (who, why, when, where), and the management history from creation to archiving, etc.;
- Context - describes the relationships that the content has with other content and organisational structures etc., so that users of the content can gain an understanding of where it fits in;
- Fixity - describes the ways in which content may be verified and its authenticity established – through for example checksums or digital signatures.

13.4 Information Packages

An Information Package is a container for

- Content Information
- Preservation Description Information

Packaging Information relates the Content Information and Preservation Description Information and provides the information necessary to identify where the actual files concerned are. The Information Package is the 'unit' which archival finding aids identify and which are then of interest to users.

There are three different types of Information Packages involved in the OAIS model.

13.4.1 Submission Information Package

The Information Package, which is deposited with the archive, will be in the format in which the producer or creator of the information holds it. The archive is likely to make certain stipulations regarding minimum standards of Representation Information (what form is the Data Object in?) and Preservation Description Information (details of description, creation environment, context and fixity) that are required as a condition of deposit.

13.4.2 Archival Information Package

The information stored in the archive is likely to be stored in a different arrangement to that of the submissions. The submissions may be single items submitted serially over time, whereas the Archival Information Packages may be aggregations of submissions. A single submission may be added to a number of Archival Information Packages. These decisions will be made according to the policies of the particular archive.

13.4.3 Dissemination Information Package

When a request for information is made to an archive, the materials required to meet the request must be assembled and prepared for the consumer/user as a Dissemination Information Package or Packages. These will be assembled from Archival Information Packages and may be constructed to exclude information to which the consumer in question has no right of access.