DIGITAL RECORDS PRESERVATION

Where to Start Guide

October 2010

This document was produced by ISO TC 46/SC 11/WG 7 – the ISO working group responsible for Records Management Digital Records Preservation. Any recommendations, suggestions for future development of the document or new references are welcomed, and should be forwarded to the working group convenor at paulm@corconcepts.co.za

This publication is freely available from the ISO/TC 46/SC 11 home page http://isotc.iso.org/livelink/livelink?func=ll&objId=8800112&objAction=browse&sort=name

Table of Contents

1 Introduction

1.1 Audience

The intended audience is the person, groups or business units in an organization tasked with the development of plans for the preservation of digital records.

This includes managers of organizations, records, archives, information and technology management professionals, risk management and audit professionals, and other individuals with a duty to create and maintain records on behalf of the organization. This guide is not intended for digital preservation experts. The guide will also be helpful for those involved in the management of personal information and for digital record technology providers.

1.2 Purpose

Organizations are increasingly creating, receiving and managing digital records. Sound records management practices require that the same rules and principles should apply to all records regardless of their media or form. Records in digital format do however have some unique characteristics that require specific actions to be taken to ensure that they retain their value for as long as they are required. The purpose of this Guide is to provide guidance related to developing a preservation plan.

This "Where to Start Guide" relies on the concepts set out in ISO 15489 parts 1 and 2 and needs to be applied in conjunction with that standard. Applicable requirements included in ISO 15489 are not duplicated in this document.

1.3 Outcomes

Outcomes are:

- Understanding the issues specific to preservation of digital records;
- Guiding the development of the preservation plan;
- Contributing towards the safeguarding of digital records asset over time with confidence;
- References and links to what is happening elsewhere.

1.4 Concepts

Key concepts facing the preservation of digital records include the following:

- The unique nature of digital records;
- Functionality of digital records systems.

Definition of a digital record

A digital record is recorded information in a form that requires a computer to process it and that otherwise satisfies the definition of a record. Applicability of this guide is limited to those records which take the form of digital "documents" such as those created by office productivity tools like word processors and spreadsheets.

The unique nature of digital records:

The content of a digital record exists in an electronic format that cannot be interpreted without the use of software (a computer program) to render the content to the user. For example, the contents of correspondence typed into a word processing system cannot be read without the use of that system. The elements that constitute the record include:

- the data or content itself;
- formatting and control characters that are generally not visible to the user; and
- metadata about the record.

The type of information in the correspondence may require different technologies to translate and present it to the user. The content may have pictures embedded in addition to text, and each of these content-types may need different software to accurately recreate them.

The information may contain links to external information (such as hyperlinks to web sites). Without further technology and infrastructure, the information these links refer to may not be available.

Metadata is an essential part of a digital record, and yet all or parts of the metadata may be stored separately from the record.

Functionality of digital records systems:

Digital records may be created using a variety of different computer systems such as word processors, spreadsheets, line of business systems and e-mail. The structure and nature of these systems varies greatly, and whilst these systems may create and manage records, they are generally not designed with electronic preservation in mind.

Records may be transferred at some stage from these creating systems to a system designed to store and manage the records over the long term. These specialised systems are designed so that the records cannot be altered, and functionality such as classification schemes and retention and disposal rules can be applied, and all associated metadata stored.

1.5 Challenges

Most records created today are either "born" in digital format, or are converted to digital format through some technological transformation. Digital records have a number of fundamental differences from traditional records; hence need special treatment to preserve their integrity as records over time. These unique characteristics of digital records require unique preservation actions.

Organizations should be aware that digital records preservation actions ideally begin at records creation. In other words, the earlier in the process that the preservation activities start, the greater the assurance that the records will meet the requirements of reliability, completeness, authenticity and usability.

The unique nature of digital records creates a number of challenges that need to be addressed if the records are to be preserved over time. These challenges often arise long before the retention period of the records has expired. The primary challenges are:

- Physical format obsolescence and degradation (Media);
- File format obsolescence;
- Software obsolescence (operating system, etc);
- Hardware obsolescence.

Physical format obsolescence and degradation (Media)

Digital records are stored on media technologies that are subject to potential damage and degradation over time. Damage can result from inappropriate storage conditions, handling procedures and read equipment maintenance. In addition, the rapid change in storage media technology typically results in obsolescence in relatively short periods of time.

File format obsolescence

Compounding the media challenge is the fact that data needs to be represented using software programs. These programs can only read and interpret data which is in a specific file format. Software vendors are constantly upgrading their systems to add new functionality, and this presents a risk that older formats may not be readable in the new software systems.

Software obsolescence

Allied to the above challenge is one whereby the software programs themselves become obsolete, hence rendering the file format useless. Software obsolescence could be in the form of application programs or operating systems upon which the applications run. Obsolescence of either of these platforms could result in records not being viewable.

Hardware obsolescence

Whilst the physical media may still be in sound condition, the technology may not exist in future to read the media. As with the software challenge, new advances in hardware are delivering faster, more robust technologies. As these hardware technologies advance so the way that information is stored and retrieved may change, and older media may not be supported.

2 Why be concerned about preserving digital records?

Whilst digital records clearly create unique challenges, at the heart of preservation is the requirement that records need to be kept, regardless of their format. The following extract from ISO 15489 highlights the rationale and importance of the preservation function:

"Records contain information that is a valuable resource and an important business asset. A systematic approach to the management of records is essential for organizations and society to protect and preserve records as evidence of actions. A records management system results in a source of information about business activities that can support subsequent activities and business decisions, as well as ensuring accountability to present and future stakeholders. Records enable organizations to:

- *conduct business in an orderly, efficient and accountable manner;*
- *deliver services in a consistent and equitable manner;*
- *support and document policy formation and managerial decision making;*
- *provide consistency, continuity and productivity in management and administration;*
- *facilitate the effective performance of activities throughout an organization;*
- *provide continuity in the event of a disaster;*
- *meet legislative and regulatory requirements including archival, audit and oversight activities;*
- *provide protection and support in litigation including the management of risks associated with the existence of, or lack of, evidence of organizational activity;*
- *protect the interests of the organization and the rights of employees, clients and present and future stakeholders;*
- *support and document current and future research and development activities, developments and achievements, as well as historical research;*
- *provide evidence of business, personal and cultural activity;*
- *establish business, personal and cultural identity; and*
- *maintain corporate, personal or collective memory."*

(ISO 15489-1 Information and documentation – Records management – Part 1 General, Chapter 4 "Benefits of records management")

In other words, records of decisions, actions, agreements and transactions are among the records that need to be preserved in a democratic society, to ensure the accountability of government and business to its citizens, stakeholders and clients. Since digital records are now used widely, it has now become a common category of evidence requested in legal proceedings and audits. Therefore, it is essential that all forms of digital records be appropriately classified, filed and preserved, so that they can be found when needed, and so that their authenticity, integrity and reliability can be proven.

3 What actions should be taken?

The following summary identifies important activities that should be initiated. All the activities detailed in this section should be formally allocated to individuals or groups in the organization.

A reliable routine should be developed to ensure that records are regularly incorporated into the preservation system. This routine needs to be invoked to ensure consistent handling from the moment records are taken from the creating system. The routine should be designed in such a way that it only applies to those records which have been identified for preservation.

Records should be stored in a safe environment, regardless of format or media.

Preservation options should be identified as a key component of the strategy. Different options may be valid for different types of records, media or formats. The options available generally include:

- converting into new formats/platforms (hardware and software);
- migration to new media;
- emulation.

Historical account logs need to be kept of all records management and administration actions relating to the digital records.

Routines to secure authenticity and reliability in a verifiable way need to be identified and implemented.

Environmental coverage of technical developments, conversion techniques and relevant standards should be conducted on a regular basis.

An accepted model of concepts needs to be developed and used as a foundation for the preservation programme.

An accepted metadata schema needs to be developed and implemented. Capture of all related metadata, transfer of the metadata to any new formats/systems, and secure storage of the metadata is critical.

4 How should I develop a preservation plan?

4.1 What is a preservation plan?

A preservation plan is a plan to preserve a specific collection or a part of a collection of digital objects, taking into account:

- preservation policies;
- legal obligations;
- organizational constraints;
- technical constraints;
- user requirements; and
- preservation goals.

It describes the preservation context, the evaluated preservation strategies (like migration, conversion and emulation) and the resulting decision for one strategy, including the reasoning for the decision. Furthermore a preservation plan defines a series of preservation actions to be taken by the responsible institution due to an identified risk for a given set of digital objects or records (called collection).

The plan defines a well-documented procedure of actions to ensure the long term access and usage of the collection. The preservation plan includes:

- preservation context;
- selected preservation strategy;
- the evaluation results of different preservation alternatives and the decision;
- roles and responsibilities for the preservation plan and for its monitoring;
- triggers that initiate the execution of the plan;
- cost (maybe estimated) of realising the preservation plan:
    - cost for preparing the preservation plan;
    - cost for executing the preservation plan.

The preservation plan should ensure that all records, whatever their format and storage media, are retained with a view to being accessed at some later date. Digital records differ from traditional records in that the digital bit stream itself is not enough for the digital record to be usable.

In order to safeguard the authenticity of digital records throughout their useful life, preservation plans require that preservation systems that are a combination of technology and organizational procedures are used.

Preservation plans should ensure that digital records are at least:

- searchable and available for access in a timely manner;
- interpretable (usability, presentation, representation, view, encryption);
- retrievable, including the appropriate metadata;
- protected against loss of rights such as intellectual property, confidentiality and copyright;

- available for access for as long as required, by those authorised to access the record;
- monitored for quality of the access provision (availability, timeliness, delivery, use history).

The preservation plan should cover all activities, from initial planning, preservation activities, monitoring and on-going reviews.

4.2 What is the scope and structure of the plan?

The preservation plan should include all the processes and procedures used for the preservation of digital records. This should enable the organization to implement reliable digital preservation processes, in line with the organization's general records management policy.

The processes and procedures for digital preservation should include:

- the determination of which records should be retained in a digital form;
- the implementation of disposition schedules;
- the identification of formats and storage media (e.g. optical, magnetic) for digital records;
- the management of additional metadata requirements for digital records;
- the implementation of preservation actions needed to ensure reliability and authenticity of digital records;
- compliance with legal and regulatory requirements, specific to the digital record within each jurisdiction;
- the identification and management of the risks associated with technological obsolescence;
- the ability to demonstrate the authenticity, reliability and usability of a digital record;
- regular reviews and audit requirements.

These processes and procedures should ensure that the essential characteristics of a digital record are preserved, along with any additional characteristics related to the preservation processes.

4.3 What records are included in the plan?

Typically, records that are at risk from technology obsolescence should be included in the preservation plan. Records identified as being at risk should be reviewed based on the required retention period and the expected life of the storage software / hardware platform.

4.4 How does the plan relate to policies/strategies?

Organizations with good records management strategies and policies will have approved retention schedules for all records. These schedules will identify legal, regulatory and business needs for the retention of records over time. A preservation plan will enable access to records throughout their retention period, even when storage software and hardware technologies become obsolete or unsustainable.

4.5 Who is responsible for the preservation plan?

Responsibilities for the development and testing of the preservation plan should rest with the corporate records manager, with input from Information Technology (IT) staff. Input from IT should include plans for any changes to software used for access to records and any changes to storage technology.

Where a preservation plan is to be implemented, the following should be represented on the conversion / migration project:

- organizational management (to approve budget and resources);
- information technology (IT) staff (to implement the processes);
- records management (to ensure that records remain authentic etc.);
- users (to ensure that converted / migrated records are accessible);
- auditors (to ensure that the processes are audited once completed).

It is important to note that a digital preservation plan cannot be created and implemented in isolation. A key aspect in developing the plan should be collaboration with various internal and external stakeholders.

4.6 What are the specific technical challenges?

The general technical issues to be managed are detailed in Section 1.5 - Challenges. Included in this section are discussions on:

- media obsolescence;
- file format obsolescence;
- software obsolescence;
- hardware obsolescence.

Media obsolescence includes the inability to expand or maintain storage media due to the lack of additional media or spare parts. This issue also includes the situation where storage technology is to be replaced with new, more efficient / cost effective / faster storage media.

File format obsolescence is usually the result of upgrades or replacements to software for displaying records.

Software obsolescence can be the result in changing software suites, for example when changing email software from MS Outlook Express to MS Outlook.

Hardware obsolescence is likely to result where major pieces of IT equipment, especially storage systems, are replaced. This may be, for example, the replacement of magnetic tape or optical storage technology with on-line Storage Area Network solutions.

4.7 The metadata challenge

Retaining important metadata during a conversion / migration process can be a challenge. For example, when transferring records from one storage system to another, the ‘date stored’ field is often replaced with the date the records are stored on

the new media. This results in the original storage date being potentially lost, unless special care is taken.

Thus, as part of the conversion / migration project, metadata that needs to be retained should be identified and measures taken to facilitate this retention.

It may also be necessary to create new metadata, to enable the identification of conversion / migration processes that have taken place with stored records.

4.8 Retention period and access requirements

When planning a conversion / migration project, the length of time that records need to be retained should be taken into consideration.

Where records are past their retention period, it is appropriate to delete them as part of the conversion / migration project. It is important to retain records of all such deletions.

Where records are close to the end of their retention period, a view needs to be taken as to whether access will be required (or is anticipated) before their deletion. It is likely that a risk assessment will indicate that including these records in the conversion / migration project will not be cost effective.

4.9 What are the non-technical issues?

As well as the technical issues discussed above, there are often non-technical issues. Some of these issues are included in this section.

Staffing

Conversion / migration projects typically require additional resource than that normally available within an organization. As well as IT resource (including both staff and available test systems) to create and operate software necessary to progress and monitor the project, user involvement will be necessary to assess how new systems will integrate with working practices. There may also be additional IT and user resources required during the testing phase of the project, and as part of an overall project monitoring team.

Management

Discussions will need to be held prior to a conversion / migration project being undertaken, to ensure that the right levels of funding and resource are available for a successful project conclusion.

Funding can be an issue with conversion / migration projects. Costs such as the availability of testing systems and the cost of procedural changes may not have been considered when assessing budget requirements.

Training of staff in preservation activities and systems can be a significant issue. As preservation is seldom seen as a day-to-day operational issue, specific training interventions will be required on a regular basis.

Monitoring

The monitoring of the records management system both during and after the conversion / migration project will important for a successful outcome. This monitoring should include assessing how the new systems are being used, and of user confidence in the new systems.

Authenticity

Are stakeholders (e.g. management, IT staff and users) confident that the conversion / migration process has not compromised the authenticity, reliability and accessibility of the records? Such confidence will only be achieved by the implementation of an appropriate preservation plan, accepted by all stakeholders prior to its use.

4.10 Critical organizations events

Critical organizational events that could trigger preservation actions include:

- migration of IT system to new software / hardware platforms;
- the introduction of new storage media types;
- the introduction of new storage file formats.

5 Will my plan work over time?

5.1 Introduction

It is important that an organization defines a preservation plan for the digital records that are registered in its records management system. The preservation plan is designed to help ensure that the records retain their characteristics of authenticity, reliability, completeness and availability over time.

5.2 Should I update my plan periodically?

The preservation plan allows the organization to ensure its records are properly preserved, in spite of technological advances in hardware and software. Organizations should review and update their preservation plans on an annual basis, in order to take advantage of new developments and changes in technology.

During the annual review, the organization should identify changes in the technology that could result in changes to the records themselves and determine whether to update the preservation plan. It is advisable to anticipate the impact of technology change on the records being preserved. In this way, the organization can ensure that the preservation plan is up to date, thus avoiding loss of records due to changes in technology.

5.3 How can I feel confident that my plan is working?

Digital records are complex by nature and the technology used to create and manage them is constantly evolving. Therefore, it is not possible to guarantee that any specific plan will be effective over time.

However, there is sufficient experience to suggest that the adoption of preservation procedures that are based on system standards and generally accepted practices will result in properly preserved records[1]. The preservation plan should define procedures and controls that can be evaluated as to their performance and effectiveness in preserving digital records. These procedures and controls should include:

- formats and the media in use;
- backup copies and their revision;
- permanent analysis of risks;
- quality control of records and systems;
- security, access and legibility.

Each organization should appoint people to take charge of these controls, set the pre-established processes to be used, and ensure the processes are routinely followed. The outcome of these controls should be documented so that the plan can be modified should the current controls be no longer effective.

5.4 Have I taken the correct preservation actions?

Even though the organization can never be truly sure of having taken all the correct preservation actions, they can be reasonably sure of success if sound principles,

[1] See Section 7 for a short list of recognized standards and practises.

methods and procedures supported by records management and information systems professionals are followed.

In any case, the organization should have experts that can evaluate technological advances and the risks they pose in relation to the preservation of digital records. These experts should have knowledge of the organization's system standards and thus be able to determine the need to update the preservation plan and actions.

With controls in place and an on-going awareness of technology changes and their impact on preserving records, an organization can be reasonably assured that it is taking the appropriate preservation actions.

6 What help is available?

Section 7 of this publication is a list of references and details of existing work in the area of digital preservation. The list of references is provided for additional detail and guidance. This is not intended to be a comprehensive list, and has been updated to the best of the committees' knowledge as at October 2010. As there is on-going research being conducted in this field, users are requested to advise us of the latest information. For contact details, please see the first page of this publication.

7 List of references and current digital preservation work

General

Open Archival Information System (OAIS)
http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html

Dutch Digital Preservation knowledge bank website
http://en.archief.nl/knowledge-base/digital-preservation

Digital Preservation Coalition
http://www.dpconline.org/graphics/index.html

Digital Preservation Coalition Preservation Handbook
http://www.dpconline.org/advice/preservationhandbook

Digital preservation tutorial from the Inter-University Coalition for Political and Social Research
http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html

Digital Curation Centre (UK)
http://www.dcc.ac.uk/

National Library of Australia - Preserving Access to Digital Information
http://www.nla.gov.au/padi/

National Digital Information Infrastructure and Preservation Program, Library of Congress, USA
http://www.digitalpreservation.gov/

NARA's Electronic Records Archive project
http://www.archives.gov/electronic_records_archives/index.html

Digital Preservation, Digital Library Federation
http://www.diglib.org/preserve.htm

Online Computer Library Center
http://www.oclc.org/digitalarchive/support/default.htm

File Formats

GDFR (Global Digital Format Registry) - provides sustainable distributed services to store, discover, and deliver representation information about digital formats
http://www.gdfr.info/

PRONOM- an on-line information system about data file formats and their supporting software products. Originally developed to support the accession and long-term preservation of electronic records held by the UK National Archives, PRONOM is now being made available as a resource for anyone requiring access to this type of information.
http://www.nationalarchives.gov.uk/pronom/

TIFF 6.0 Specifications
http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf

PDF/A specification
http://www.pdfa.org/doku.php

JPEG2000 specification
http://www.jpeg.org

Digital Repository

Nestor criteria - Catalogue of Criteria for Trusted Digital Repositories - Version 2
http://files.d-nb.de/nestor/materialien/nestor_mat_08_eng.pdf

Center for Research Libraries: Trustworthy Repositories Audit & Certification: Criteria and Checklist
http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Library and Archives Canada: Trusted Digital Repository – In all domains the amount of digital information is increasing at a rapid rate, this raises crucial questions of preservation. Our intellectual capital, as laid down in educational, scientific, public, cultural and other intellectual resources, is increasingly at risk by the volatile character of digital objects and the rapid developments in information technology
http://www.collectionscanada.gc.ca/digital-initiatives/012018-4000.01-e.html

Space data and information transfer systems -- Open archival information system -- Reference model – ISO 14721:2003
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683

Preservation Metadata

Library of Congress, USA

PREMIS (Preservation Metadata) - A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials –
www.loc.gov/standards/premis

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

Simple Rights Schema – A METS extension schema for rights declaration
http://www.loc.gov/standards/mets/news080503.html

Data Dictionary for Administrative Metadata for Audio, Image, Text, and Video Content to Support the Revision of Extension Schemas for METS
http://lcweb.loc.gov/rr/mopic/avprot/extension2.html

Online Computer Library Center USA

A Metadata Framework to Support the Preservation of Digital Objects
http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

Digital Archive Metadata
http://www.oclc.org/support/documentation/pdf/da_metadata_elements.pdf

Global Digital Format registry
http://hul.harvard.edu/gdfr/

Preservation Metadata for Digital Objects: A Review of the State of the Art
http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf

PREMIS (Preservation Metadata Implementation Strategies)
http://www.oclc.org/research/projects/pmwg/

Digital Archive Metadata
http://www.oclc.org/support/documentation/pdf/da_metadata_elements.pdf

Others

Flexible and Extensible Digital Object Repository Architecture (FEDORA)
http://fedoraproject.org/get-fedora

National Library of New Zealand, Metadata Standard Framework, Preservation Metadata
http://www.natlib.govt.nz/catalogues/library-documents/preservation-metadata-revised

Assessment of Metadata Needs for Imaging Projects

Adobe's Extensible Metadata Platform (XMP) – also local implementations
http://www.adobe.com/products/xmp/main.html

Technical Overview

Glossaries of Technical Terms

Technical Advisory Service for Images
http://www.tasi.ac.uk/glossary/glossary_technical.html

Storage and Digital Preservation

Conversion Specifications, American Memory, Library of Congress
http://memory.loc.gov/ammem/techdocs/conversion.html

NDLP Project Planning Checklist, Library of Congress
http://lcweb2.loc.gov/ammem/prjplan.html