JATLIM International

http://www.jatlim.org

INTERNATIONAL JOURNAL OF APPLIED TECHNOLOGIES IN LIBRARY AND INFORMATION MANAGEMENT

Manuscript Number: JATLIM - 01.01/4/2015-010

International Journal of Applied Technologies in Library and Information Management 1 (1) 10-075 - 086

Digital Preservation of Electronic Records: Review of Concepts, Challenges and Strategies

Okechukwu Kalu Ogwo
ogwopg@gmail.com

Oliver C. Ugocha [PhD]
chukwudiugo@yahoo.com
Department of Library & Information Science,
Abia State University, Uturu

Abstract

This paper presents the tripartite (short, medium and long term) definitions of digital preservation by corporate bodies in the information profession field. The paper also clarifies the concept and discusses electronic records preservation in terms of content and context. It identifies challenging factors in preservation of digital sources and points out strategies for effective and efficient preservation of electronic information which includes steps to long term preservation project. The paper concludes that with recommendation that digital preservation of electronic records should be taken up as a project of paramount importance by librarians, archivists, and other information professionals especially in Africa so as to serve as a roadmap towards technologically based strategies to overcome the identified challenges.

Keywords: *Digital preservation, electronic records, preservation strategies, storage media, metadata*

1.0 Introduction

In explaining the scope, meaning and general concepts of the term 'preservation' It worth noting that it is an umbrella word or concept under which most librarians and archivists cluster all the policies and options for the managing, including conservation treatments of different formats of information materials. It has long been the responsibility of librarians and archivists – and the clerks and scribes who went before them – to assemble and organize documentation of human activity in places where it can be protected and used. Digital material preservation therefore is a way of preserving information materials, which refers to digital surrogates created as a result of converting analog materials to digital form (digitization), as well as those that are 'born digital' for which there has never been, a is never intended to be an analog equivalent, and digital records. The term 'digital preservation' refers or embraces two major categories, based on the process by which the digital materials came into existence or were created (Digital Preservation Coalition (DPC), 2002).

According to IFLA-ARL (2004) definitions, "preservation includes all the managerial and financial considerations including storage and accommodation provisions, staffing levels, policies, techniques and methods involved in preserving library and archive materials and the information contained in them". Conservation denotes those specific policies

and practices involved in protecting library and archive materials from deterioration, damage and decay, including the methods and techniques devised by technical staff. Restoration denotes those techniques and judgments used by technical staff engaged in the making good of library and archive materials damaged by time, use and other factors".

According to Ashok (2008) "Digital preservation" or "digital archiving" essentially aims at taking steps to ensure the longevity of electronic documents. It applies to documents that are either "born digital" or stored on-line (or on CD-ROM, diskettes or other physical carriers) or to the products of analog-to-digital conversion, if long-term access is intended. Digital preservation concerns two types of documents: namely 'born-digital documents' and 'digitally created' documents. Born digital documents: These refer to those materials that were initially created using some form of digital technology. They are often termed as 'electronic records'. Digitally created documents refer to those materials, which have been transformed from analog to digital form through some reproductive means such as re-keying the information or scanning the document or object etc.

An electronic record is a form of information produced in the course of an event or a routine activity by an individual or an organization. Electronic records are created and retained as specially designed output from an information system or personal computer. They may be selected for retention because of their comprehensiveness and portability, or they may not be retained since other forms of the same information (e.g., hard copy, microfilm) may be preferable for personal use, acceptable as evidence, or retained at lower cost. Most files output or saved from information systems or personal computers are not electronic records. Simple backups or downloads usually do not contain sufficient contextual information to authenticate and verify the accuracy of the content (Spindler, 2010). He further asserts that electronic records can include information originally created in an electronic format or information created in hard copy or analogue format and subsequently digitized through use of digital imaging, optimal character recognition technologies or direct analogue to digital conversion for audio and video information. Electronic records can include digital image files, audio files, and geographic information system files, as well as the more common documents produced by word processing or e-mail.

Archivists have struggled, over the years, to understand and explain the difference between an electronic record and electronic information or data. Archivists have made important distinctions between the content of information and its context, since elements of the context of information creation can certify the accuracy and authenticity of information content. Context is also essential for documenting how and when content changes have been made over time. Electronic records must have content and context in order to successfully document an event, an activity, or a business transaction.

Electronic publications often serve as a record of an activity or transaction since they generally are presented with attributions of authorship and sources. They are often granted the status of records because authors, footnotes, and bibliographies prove sufficient context for readers to have confidence in the accuracy and authenticity of the information system are intended to serve as records, but often that output does not include the contextual information necessary to authenticate the content of an electronic record. It is based on this background that this study is set out to review definitions, concepts, challenges and strategies for digital preservation and preservation of electronic records.

1.2 Objectives of the study

The study is carried out in order to:

i. Review the definitions of digital preservation
ii. Discuss the concepts of electronic records preservation in terms of content and context
iii. Identify challenges and factors affecting electronic records preservation
iv. Present strategies for effective and efficient preservation of digital or electronic resources

2.0 Literature Review

2.1 Definitions of Digital Preservation by ALA

During the American Library Association's (ALA) 2007 Midwinter Meeting, a working group within the Preservation and Reformatting Section (PARS) was charged to draft a definition for digital preservation to support the work of PARS, the Association for Library Collections and Technical Services (ALCTS) and the ALA, for use on the web, verbally, in written policy statements, and other documents. The work was reviewed at the 2007 Annual Meeting and approved for further distribution by the PARS and ALCTS Executive Committees. The definitional policy statements are as follow:

> Policies, strategies and actions: this phrase makes explicit the need for a declared intention to preserve, a plan for doing so, and engagement in measurable activities to realize that plan. These policies and strategies determine the precise actions required of a digital preservation effort.
>
> To ensure access to: the concept of access was initially used because of its fundamental place in the mission of libraries, but the working group has recognized that the phrase "to ensure usability of" may speak more effectively to the needs for computer processing and rights to access that are crucial for digital content, as well as the traditional interaction of our patrons with our resources.
>
> Accurate rendering of authenticated content: this phrase refers to well-established library and archival concepts for the authorship and provenance of a work as well as the integrity and functionality of a digital object. In digital preservation there may be a requirement to support the ongoing machine readability and future processing potential of digital content in addition to human interaction with the content.
>
> Born digital and reformatted content: The term "reformatted" carries a very particular meaning in the library community and among the PARS membership. This meaning is different than its colloquial use and its meaning in information technology. Consequently, the group is considering expansion of this phrase to "content that is born digital as well as converted to digital form."

Digital preservation is defined in tripartite ways: short, medium and long terms as tabulated in Table 1.

Table 1: Tripartite Definition of Digital Preservation

Short Definition	Digital preservation combines policies, strategies and actions that ensure access to digital content over time.
Medium Definition	Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.
Long Definition	Digital preservation combines policies, strategies and actions to ensure the accurate rendering of authenticated content over time, regardless of the challenges of media failure and technological change.

Digital preservation applies to both born digital and reformatted content. Gordon (2011) gives examples of digital records needing long-term storage:

Utilities – agreements for elements and locations of power, water, communication, and gas lines

Finance – bonds, loans, and proofs of stock ownership

Insurance – life and property policies

Environment – locations of hazardous waste

Medical – lifetime reports; longer for multi-general studies and genetic analysis

Pharmaceutical – laboratory notebooks, especially for multi-generation drugs and devices where new products draw on old research

Energy – geographic information system data, and (most long-lived) inventories for dry-cast, above ground storage of spent nuclear fuel rods where the half-life is 10, 000 years.

2.2 Challenges of Preserving Electronic Information

Although thirty years of work in acquiring and retaining electronic content has not resulted in complete solutions for electronic record preservation, but researchers have started to identify and catalogue the major barriers to successful electronic record keeping. A successful electronic record preservation strategy requires attention to seven major challenges. Table 1 illustrates these identified challenges.

Table 2: Seven challenges of electronic record preservation

S/N	Challenge	Example
1.	Physical degradation of storage media	CD does not read anymore since the recording surface has been degraded by air pollutants.
2.	Physical obsolescence of storage media	5.25 inch floppy discs does not fit in CD-R drive.
3	Incompatibility/noninteroperability of storage media	One manufacturer's DVD disc will not play in DVD players produced by a different manufacturer.
4	Software, operating system, or encoding incompatibility/ Noninteroperability	New software release will not run file from old release. New software release opens old file, but contents are corrupted. Software is not compatible with operating system. Product designed for display with one Internet browser does not look the same in a different browser. Proprietary backup software and storage format not compatible with new release or other backup software. Propriety codes form HTML editing packages do not convert to XML.
5.	Human error/vandalism	Accidental or malicious deletion. Website is vandalized
6.	Backups and snapshots	Wrong files backed up. Incomplete snapshot: Video server backed up, Web server not backed up. Snapshots of different servers not contemporaneous.
7.	Metadata	Insufficient context to verify source and authenticate content.

Source: Spindler (2010) ELIS. P 1685

3.1 Digital Preservation

Digital preservation refers to the management of digital information over time. Unlike the preservation of paper or microfilm, the preservation of digital information demands ongoing attention. According to Das and Das (2008), "this constant input of effort, time, and money to handle rapid technological and organizational advance is considered the main stumbling block for preserving digital information beyond a couple of years. Indeed, while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost." Digital preservation can therefore be seen as the set of processes and activities that ensure the continued access to information and all kinds of records, scientific and cultural heritage existing in digital formats. The Research Libraries Group (2004) defined digital preservation as follows:

"*The managed activities necessary: 1) For the long term maintenance of a byte stream (including metadata) sufficient to reproduce a suitable facsimile of the original document and 2) For the continued accessibility of the document contents through time and changing technology*."

3.2 Difficulties in Digital Preservation

In his own view, Ashok (2008:304) itemized difficulties in digital preservation:

Increasing complexity of digital objects (incorporating text, images, audio, video in various formats) and their increasing software dependence (e.g. storage in databases);

Rapidly increasing number of digital objects and proliferation of document standards and formats;

Lack of planning to incorporate preservation needs in systems and lack of availability of off-the-shelf products supporting preservation needs.

Lack of consideration of long-term access requirements when creating digital products; lack of long term thinking about management of the digital information.

Absence of widely-accepted standards which will assure access over time.

Copyright/intellectual property rights that may interfere with the ability to preserve digital objects through systematic

Table 3: Three levels of Digital Preservation Functions

Content creation	Content Integrity	Content Maintenance
Clear and complete technical specifications Production of reliable master files Sufficient descriptive, administrative and structural metadata to ensure future access Detailed quality control of processes	Documentation of all policies, strategies and procedures Use of persistent identifiers Recorded provenance and change history for all objects Verification mechanisms Attention to security requirements Routine audits	A robust computing and networking infrastructure Storage and synchronization of files at multiple sites Continuous monitoring and management of files Programs for refreshing, migration and emulation Creation and testing of disaster prevention and recovery plans Periodic review and updating of policies and procedures

Source Gordon (2012) Tabulation – Authors'

4.1 Preservation Issues

While there are many strategies for preserving the availability of digital records, there is no single solution, no best practice, and no established policies or procedures that meet widespread needs. The issues fall into four categories: storage media, hardware, software, and governance (Gordon 2012). Drawing from his practical experiences, Gordon (2012) lists and explains other preservation issues including strategies and steps to long term preservation project in relation:

i. Storage Media Longevity

Many variables affect the lifespan of storage media, and no comprehensive, scientific evaluations advise the consumer. However, it is clear that each medium has an Achilles' heel. For example, prudent care reduces, but does not prevent, deterioration of digital linear tape, such as:

Increasingly brittle tape

Failure of the adhesive that attaches the magnetic particles to the tape

Exposure to magnetic fields

Optical disks use organic dyes that biodegrade over time, and few blank disks come with a date of manufacture. The conditions under which disks are shipped and stored are generally uncontrolled. It is hard to predict the longevity of a spindle of DVDs produced of unspecified materials in western China, shipped over land to Shanghai, transported across the Pacific Ocean on a container ship, and sent by rail to a distributor in the United States. Even solid state memory degrades over time due to cracking seals, unstable material, and environmental factors, including cosmic rays.

ii. Hardware Lifespan

The lifespan of computers and their peripherals is shorter than that of persistent records. Only specialists retain disc drives for the 5.25-inch floppies of the 1990s or even the subsequent 3.5-inch disks. An Apple representative recently opined that its future

computers would not support removable media; all storage and access to stored information would go through the cloud.

i. Software Obsolescence

Similarly, most software is obsolete before the end of persistent records' lifespan. Disk operating system (DOS), which was the most common operating system (OS) on personal computers in 1990, is unseen now. Microsoft stopped supporting its Windows 97 OS in 2007, and it makes no commitments to backward compatibility in future versions of Windows. Generally, application software changes even faster. Updated versions of line-of-business programs debut regularly. The rapid churning of software developers and manufacturers means there are no guarantees of continuing support and compatibility. For mission-critical applications, the source code can be held in escrow, but putting that to work is expensive and sometimes impractical.

iv. Governance Continuity

This may be the area of greatest vulnerability for persistent records. It reeks with uncertainty because it requires action today based on predictions and assumptions about the future. Two basic issues cast doubt on the maintenance of persistent records because of governance.

First, each of the electronic options for long-term preservation, described below, has serious limitations. For example, a basic tenet of records management is provenance, the practice of organizing records in useful sequence, such as date of creation or entry into a records management program. Without special care, a digital record's date (often entered as metadata) may be corrupted by something as beneficent as a virus scan. Similarly, when a record is transferred from an old medium to a newer one, the date of origin might be altered.

Second, there is no question that maintaining electronic record viability over the long haul requires effort and resources. Today, it is easy and irresponsible for a records manager to fail to make plans and investments for the future. In fact, current demands of most records' programs fully consume the resources of their directors. Few have the assets, means, and foresight to make substantive provisions for the next 25 years and longer.

It seems presumptive to assume that the next-generation records management staff will have the skills, time, technology, money, and motivation to maintain accessibility to persistent records from today, when current records managers have made no plans to help them.

4.2 Preservation Strategies

Given the challenges to preserving records' availability over extended lifecycles, it is fortunate that several prudent strategies exist. Clearly, there is no single solution that is best for all situations. In noting the various technologies and techniques available, the best strategists will consider all and create a blueprint that best meets their needs.

i. Print to Paper

Tried and true, paper lasts hundreds (if not thousands) of years. Longevity depends upon stable materials and controlled storage. The techniques and risks are well known. Printing digital records is usually most appropriate for small quantities of records, as accessibility may suffer and costs will increase as volume rises. Printing in fonts easily recognized by scanner software facilitates a return to digital media upon records' retrieval. A challenge in printing digital records is to retain associated metadata with the printout.

ii. Establish a Computer Museum

Some organizations maintain their aging records by stopping the calendar, technologically speaking. IT directors

sometime assert that records within computer systems are viable, as long as the systems are active. They only see an issue when systems are retired. This is valid as long as the peripheral devices and supporting infrastructure still work with the aging system. When records' retention period does not exceed the longevity of the system, this can be a usable strategy. It has limits, however. Recently, a septuagenarian who programs COBOL, which is one of the earliest computer programming languages, reported he continues to work for ever-higher pay because of the scarcity of his skill.

A related strategy creates a museum of computer hardware and software to accompany record archives. The idea is that when records are needed, all components can be re-activated to deliver old records. This potentially expensive strategy fails to address the instability of storage media, degradation of hardware, and the potential lack of human skills necessary to make old systems productive.

iii. Migrate Periodically

The history of digital storage is littered with obsolete devices and related software designed to preserve digital information. An array of drives, cartridges, juke boxes, silos, tape players, and the like are now useful mainly as boat anchors and avant-garde sculpture materials. To counter this parade of obsolescence, storage programs committed to removable media practice regular, wholesale migration of records from old media to new. Anecdotal information indicates other large organizations also employ staff groups dedicated to systematic renewal of storage media. For organizations that can afford large-scale programs, migration may be a useful procedure. Issues of provenance and metadata maintenance require serious attention.

A bigger issue is readability. This can be ascertained only through the concerted efforts of a succession of records managers. Also, records staff need to work with IT colleagues to ensure old records that were encrypted can still be decrypted. Similarly, IT, the records staff, or both need to maintain passwords that allow access to protected records. Only well-defined policies and procedures will ensure password maintenance for 25 years or more.

iv. Use an Archival Format

Many archivists' favorite strategy for long-term preservation is converting digital images into non-proprietary, widely accepted formats. While this does nothing to overcome hardware obsolescence, it raises the likelihood that today's records will be readable in the future if the storage media retain integrity. For example, the Minnesota Historical Society stores images in TIFF format, which maps every bit without loss, even when compressed. TIFF files are relatively large, but they are complete. The format enjoys wide acceptance and stability, although it is neither an international standard approved by the International Organization for Standardization (ISO) nor an American National Standard approved by the American National Standards Institute. In fact, the format has not undergone a major update since 1992.

One concern about the TIFF format is that its copyright is privately held by Adobe Systems Incorporated, so its future is in private hands. That concern does not apply to another favored format from Adobe, PDF/A, a public version of the popular portable document format that has been approved as an ISO standard, ISO 19005-2:2005 *Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)*. PDF/A was well-designed for archiving, and like the other strategies described here, it is a useful tool in the records manager's tool belt.

PDF/A limitations, however, make it less

than a panacea for long-term preservation. For example, its large file size suggests its best use is with smaller quantities of records. PDF/A does not accept encryption, potentially making it inappropriate for personally identifiable information. Perhaps most worrisome, Adobe has made no promises of backward compatibility for future versions, leaving questions about long-term accessibility.

v. Use Microforms

Microfilm, invented in 1839, is a relatively low-tech strategy appropriate for many persistent records. In climate-controlled environments, manufacturers tout a 500-year life expectancy for some types of film. In general storage conditions, 20 years of viability is reasonable.

Microforms, including microfiche, are not appropriate for audio and video, and color images raise the cost. They are affected by the same environmental factors as paper, but are less resistant to high heat. Retrieval of individual records can be slow, although proper indexing and computer-aided finders improve retrieval times.

Relatively few records and IT leaders understand that many microfilmed records, such as text documents (using optical character recognition) and spreadsheets (when printed as 2-D barcodes), can be returned to process-able, digital files, complete with metadata and embedded formulae. When re-digitized, the records can enter a content management system for workflow, rapid retrieval, and more. In records management, however, immutability is a key principle, and standard microfilm's stability is an advantage.

vi. Consider Emerging Technologies

The door will never close on improvements in records preservation. For example, the current ultimate in size reduction and longevity is Rosetta HD from Norsam Technologies, which writes analog or digital records with an ion beam onto stable media, such as nickel. The engraving is 10 microns wide, which reduces character size more than 20 times from standard microfilm. The nickel is stable, unaffected by temperature, humidity, magnetic fields, and more.

More immediately applicable is a hybrid system widely used in the United Kingdom (UK), C-Cube Software, which offers a battery of technologies matched to the specific needs, values, and risk tolerance of each user. During a user conference, representatives from the UK's National Health Service reported that dozens of facilities of Britain's national health trust use C-Cube for long-term storage of medical records.

4.3 Steps for Long-term Preservation Project

There are many obstacles to long-term preservation of digital information, and there are many strategies for meeting the needs. Today's practitioners must amalgamate a compromised hybrid of policies, procedures, hardware, and software to best meet their organization's need. All solutions will leave some imperfections and risk, and none will be permanent. The needs and the tools with which to address those needs will always evolve. To begin a long-term preservation project, follow the steps below:

i. Identify Acceptable Risk Level

This process will help define the needs of a long-term preservation solution. For example, there could be an identified “core” class of information that is not quite “vital,” but still receives high priority for long-term preservation because losing access to it would be highly risky. Other peripheral information may not justify extensive retention, and its loss would be within stated risk tolerance.

ii. Assess Current Effectiveness

This evaluation answers the question: How well is the current solution meeting the organization's needs? A recently created maturity model for long-term preservation of digital records that measures the effectiveness of a program and compares it to an organization's acceptable level of risk is available at (www.SaingTheDigitaWorld.org/MaturityModel.html.) If maturity metrics reveal a chasm between what is concrete and what is acceptable, there is a clear call to action.

iii. Survey Existing Tools

During this process, the practitioner surveys existing tools, strategies, and tactics to cobble together a solution, with accompanying policies and procedures. This is art, as well as science, and it always involves compromises, especially in terms of budget, cost, and effectiveness. After implementation, another maturity assessment measures the effectiveness of the new preservation program.

5.0 Framework for Digital Preservation in Africa

A conducive framework for sustainable digital preservation in Africa according to Aina, Mutula & Tiamiyu (2008) should be anchored on the following:

National policies;
Legislation;
National coordination;
Human resource capacity building;
Standardization;
Research and development; and
Disaster planning and recovery

Conclusion

There have been several universal initiatives to deal with digital material preservation. Among others, International Federation of Libraries Associations and Institutions (IFLA) and UNESCO are major international organizations that are concerned with digital preservation. Digital materials can be 'born digital' or 'made digital'. The former refers to materials not intended to have an analogue equivalent, either as the originating sources as a result of conversion to analogue form. While the later refers to creating digital files by scanning or otherwise converting analogue materials, and the resulting digital copy is classed as a digital material. The challenges of preserving digital information have been there over the years; however strategies for overcoming the challenges have been suggested as reviewed. The issue of digital preservation of electronic records should be taken as project of paramount importance by librarians, archivists, and other information professionals especially in Africa so as to meet up with overseas counterparts who have gone far in developing technologically based strategies to overcome the identified challenges.

References

Aina, L.O., Mutula, S.M & Tiamiyu, M.A. (2008) *Information and Knowledge Management in the Digital Age: Concepts, Technologies and African Perspectives*. Ibadan; Third World Information Services Limited.

ALA (2007) Definitions of Digital Preservation. *Prepared by the ALCTS Preservation and Reformatting Section, Working Group on Defining Digital Preservation*. ALA Annual Conference, Washington, D.C., June 24, 2007

Ashok, R. (2008) Digital Preservation: An Overview. In: *Ramaiah, L.S. & Sujatha, G. (2008) (Eds.) Preservation of Library, Archival and Digital Documents:*

Problems and Perspectives. New Delhi Ess Ess Publications

Association of Research Libraries (2004). *Recognizing Digitization as A Preservation Reformatting Method.* Available at: http://www.arl.org/preserv/digitfinal.html.

Das, S. K. & Das, B. (2008) Fundamentals of Digital Preservation and Repository: an overview. In: *Ramaiah, L. S. and Sujatha, G. (2008) eds. Preservation of Library, archival and Digital Documents: Problems and Perspectives*. 189-197

Digital Preservation Coalition (2002) DPC History. Accessed on 14/5/2011 from:

http://wwww.dpconline.org/graphics/about/history.html.

Gordon, E.J.H. (2012) Future Watch: Strategies for Long-term Preservation of Electronic Records. *Information Management*. May-June 2012. Accessed 24/5/2014 from:
http://cententarma.org/IMM/Authorist/IMMauthorgodonejhoke.aspx

Spindler, R. P. (2010) Electronic records preservation. In: *Bates, M.J. & Maack, M.N. (eds.). Encyclopedia of Library and Information Sciences*, 3rd ed. Vol. III. New York; CRC Press; Taylor & Francis Group, 1682