# Digital Preservation: Issues and Challenges

*John Paul Anbu K*
*Marion L. N. Chibambo*

## ABSTRACT

*The paper takes stock of various preservation problems and challenges faced by different types of digital materials to ensure permanent accessibility of resources by present and future generations. It argues that digital preservation poses challenges fundamentally different from the traditional materials mainly due to varied and complex nature of digital resources. It concludes that there is an urgent need to draw up national digital preservation frameworks to protect the current digital information resources for future access and use.*

## KEYWORDS

*Preservation, Digital Resources, Digital Preservation.*

## INTRODUCTION

From times immemorial, man has created records and art works to express activities and convey feelings and eventually libraries and archives have collected, organized and managed these resources for use and posterity to serve as the institutional focal points for preservation of human culture and knowledge in different formats. Now libraries and archives consider rightly resource preservation as one of their core functions. These institutions over the years have developed an array of tools, methods and techniques for the preservation of traditional print and non-print resources with the purpose to make sure that works of distinction and valuable information are organized for utilization and posterity. The process

especially in libraries has been taken to prevent or retard deterioration of information resources in all media/format, avert their theft or loss, wherever possible to improve their condition and change, when necessary or useful, their format to save their intellectual content. A comprehensive approach entails choosing the most suitable method of preservation to be accomplished through storage of materials in proper conditions with careful handling and housing having security systems designed to eliminate mutilation and theft and above all renovation or replacement of damaged material and refreshment and migration of electronic files. The value of most materials lies primarily in the information they contain and when restoration and renovation of such materials becomes impossible or prohibitively expensive, their content need to be preserved through reformatting or conversion into other media. In recent decades, major libraries and archives e.g. The National Library of Australia, University of Buffalo Libraries, The Library of Congress, National Archives of Australia, The British National Archives etc. have established formal preservation plans for their resources by incorporating the appropriate needs and requirements into their overall planning and earmarked a regular budget for maintenance and disaster recovery. Some significant and successful methods employed by libraries and archives for their print materials include the provision for fire protection and adequate environmental controls; re-housing of acid-based print materials; microfilming of unique and endangered information resources; desktop publishing (DTP); use of acid-based paper instead of alkaline-based paper- more susceptible to discoloration and quick decay.

Besides many problems abound like the battle against acid-based papers and thermo-fax, nitrate film besides other fragile media continue to pose significant challenges to avert the situation in order to preserve the cultural, intellectual, and scholarly resources in the traditional print formats . In developing countries the problem is

further compounded by the absence and lack of the means for disaster prevention and recovery programs, inadequate environmental controls to monitor and control the environment in which the resources are housed to avoid catastrophic loss of their holdings; unfavorable environments of extreme heat and humidity which favour the growth of insects, pests, moulds, fungi, etc. But above all, inadequate funding that does not allow most developing countries to allocate budget for any meaningful and successful preservation plans contrary to developed countries where successful stories and regular use of established program plans for the print materials are well established for the preservation of archival materials.

Though no fool proof methods are established for preserving information resources, the traditional methods and techniques used for print-based information sources are very simple and require little planning. These techniques are straightforward that once the desired standards are met, less effort is required. On the other hand, the advent of digital/electronic materials has complicated these simple techniques due to nature of digital materials and media technologies associated with them. It is in this context that the authors take an overview of the problems and challenges confronting digital preservation.

## DIGITAL MATERIALS

Digital materials can be defined as, "*information resources that have been created in digital form, i.e. their life begins in digital form (i.e. Born digital resources), and those that have been converted from print or paper-based resources (i.e. analog) to digital formats (i.e. Made digital or Digitized resources)*" **(Digital Preservation Briefing Paper, 2006)**. It must be noted that the production process itself is not hermitically sealed along analog or digital because materials collected or generated for a television show, for example, may consist of both digital and analog components so tightly bound

together that at any point in their life-cycle, one may serve as a surrogate for the other or a sound recording product, for example ,includes its packaging i.e. the notes, artwork, and photograph of the artist, etc. Even on the Web, many sites offer digitized versions of print works as well as born digital documents; thus, archiving the Web itself can be seen as encompassing both *"born digital"* and *"made digital"* materials. Although there lies a thin line of distinction between *"born digital"* and *"made digital"* publications, this distinction is upheld in this paper to illustrate the essentially different problems inherent with each type of publication because each type of publication has its own complexities that require different approaches for preservation.

Digital resources are more complex, right from their inception. The *"born digital"* and *"made digital"* resources need different preservation methods or approaches. The *"born digital"* resources do not have any print or paper-based backups and some of these publications are difficult to preserve because these will only work with a specific type of hardware, software and other allied equipment. On the other hand, the *"made digital/digitized"* resources are usually converted from old texts and brittle documents which were created long before the digital era. The biggest challenge with such types of material is to capture the authenticity of the original content while retaining the original source for future innovations. The third type of digital materials ( *born digital originally*) when printed into paper formats for wide circulation/readership or for other reasons and later reconverted into digital formats( e.g. newspapers, conference papers, etc) are relatively easy to maintain as these do not pose as many challenges as the typical *"born digital"* and *"made digital"* publications.

Other than this, digital publishing has created new and different categories of document types from simple analog documents to multimedia document types, with varying types of media embedded

in-between. The advent of new and exciting digital formats such as graphs, audio, video has transformed the document types into very colorful, yet complex multimedia documents at times. The embedded audio, video, graphics, sound and other aids have enhanced the quality of digital documents. Digital publishing especially that of *"born digital"* materials encourages these types of new innovations. The current digital preservation community is therefore forced to strategize ways of preserving these unique document types. For example, decision is to be made regarding aspect(s) of the metadata unfolding the digital materials which need preservation in a particular format.

## DIGITAL PRESEVATION

The word preservation is derived from the Latin verb ***"praeservare"*** *a* composite term of two words: ***"prae"***,means *before*; and ***"servare"***- to *keep safe*. Thus, **"praeservare"** means *"to keep the past things safe"* **(New International Webster's Comprehensive Dictionary of the English Language, 1999, p. 997)**.

However, Digital preservation is broadly defined as "a series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value" **(Digital Preservation Briefing paper, 2006)**. On the other hand, it can also be defined as "the actions needed to assure enduring access to the full content of digital resources over time" **(Smith, 2002, p.137)**. Content has wider implications than simply assuring that a given file can be accessed. For example, hierarchical and structural relationships among the files and metadata that make the files usable must be preserved along with the files themselves. Digital files are like the pages of a book, while metadata is the descriptive and administrative aspect of the data contained in the files here such as author, originator, performer, title, place, name, publisher, source, etc. Thus, digital preservation is not only a technical issue, but it is

also a strategic and long-term issue that requires constant review and selection of what has to be preserved and in what format depending on the type of digital object under consideration.

The complexity of digital preservation is again aggravated by the need to preserve the digital objects within their context, i.e. the context in which they were created - the hardware, software, document type, document format and physical media. **Smith (2002)** observes that users want digital information to remain intact and appear unchanged the objects originally created and simultaneously dynamically accessible and interactive with digital content using the most advanced software and hardware technology available to them. This is a paradoxical challenge that poses the choice between preserving the originality of the document on one end and the use of the latest technology available on the other end. The challenge is therefore more complex because technology is dynamic, rendering digital preservation a continuous process and not a one-step solution as is the case with the print-based materials. As a continuous process, the concept of digital preservation includes the constant planning, resource allocation and application of methods and techniques necessary to ensure that digital information resources of continuing value remain accessible and usable now and for future generations.

## SIGNIFICANCE AND GLOBAL SCENARIO

The importance of digital preservation stems from various factors associated with the nature of digital materials. When the possibilities of digital publishing was established in the early 1990's many publishers and academic institutions ventured into digital publishing because the process was very simple and affordable, especially with the emergence of microcomputers and DTP software. However, not much consideration to technological innovations was considered. Statistics indicate that almost forty percent of websites disappear every year due to various factors, such as revision of sites, migration

and merging of databases, changes in URL addresses, etc **(Biswas, 2008, 29 July)**. Further, diverse digital objects, such as: e-books; e-journals; online databases on websites, networking discussion list serves; e-mails, institutional digital documents, etc. get aggravated by the constant changing technologies associated with their storage and retrieval. Besides the traditional reasons for preservation, the diversity of digital objects and the uncertainty of technological innovations associated with their storage and retrieval makes it imperative that they are preserved at all costs, or else all the information could just vanish into thin air. It is thus imperative for libraries and other information repositories to establish budget lines for digital preservation on a continuous basis in order to cater for the migration or conversion of the digital materials to new technologies, while at the same time preserving the originality of the document's content. Many people consider backups of digital objects as a form of preservation. Unfortunately, backups are temporary preservation methods against present calamities, because backups are neither insured against future calamities nor against any technological obsolescence. It accrue from the above that digital preservation raises challenges of a fundamentally different nature from those of the traditional print and other formats.

Digital preservation has been considered as an extremely serious undertaking, especially in the developed countries. Many countries have recognized the need for digital preservation and formulated specific programs. For example, in 2001, the United Kingdom formed a "Digital Preservation Coalition" (DPC) (http://www.dpconlone.org) to spearhead the digitization and digital preservation programme for United Kingdom. Then, in 2005, the United States of America initiated a "National Digital Information Infrastructure and Preservation Programme" (NDIIPP) (http://www.digitalpreservation.gov) and allocated close to U$100 million for digital preservation. The Library of Congress, as the

National Library for the USA was involved with the coordination of this Programme and initiated a dialogue among a number of federal agencies and other players in the information industry to establish "Guidelines for Digitization and Standards for Digital Preservation" (http://www.ditizationguidelines.gov). Further, the European Union has formulated the "Digital Preservation Europe (DPE) Strategy" (http://www.digitalpreservationeurope.eu) to oversee Digital Preservation Initiatives in the European Union. There are similar national initiatives springing up in Australia, Netherlands, Japan and other countries. These initiatives may go a long way in formulating sound digital preservation strategies and could be used as benchmarks for the developing countries instead of costly course of re-inventing the wheel.

## MAJOR CHALLANGES

### Obsolescence

Information created, stored, and accessed digitally has a risk to lose in two ways: obsolescence and physical damage. Obsolescence can affect hardware, software, and even organization of the data in a stored medium and can occur at an alarmingly fast pace. Probably more insidious and challenging than the media deterioration is the problem of obsolescence in retrieval and playback technologies (usually associated with the equipment and software for accessing and retrieving information) for digital materials. Innovations in computer hardware, storage and software is growing at a rapid pace resulting in increased storage and processing capacities at lower costs, at least for the developed countries but out of reach for most developing countries whose budgets can neither accommodate nor cope up with the rapid changes.

The biggest challenge of digital preservation is digital obsolescence. Driven by market forces, devices, processes, and software for recording and storing information is regularly replaced with new

products and methods on a 3 to 5 year cycle. Probably more insidious and challenging than the media deterioration is the problem of obsolescence in retrieval and playback technologies for accessing and retrieving the information from digital materials. File format may be superseded by newer versions which may no longer be supported by the current vendor or relevant standards; storage media may be superseded by newer and denser versions of that medium, or by new types of media (smaller, denser, faster, and easier to read); the devices needed to read a stored medium may no longer be manufactured; software used to create, manage or access digital content may be superseded by newer versions or newer generations with more capabilities using the most current technologies; computers of every size and scale being continually superseded by faster and more powerful machines that can store and process more and more content; vendors of all technologies compete, merge, emerge and fade off, making it even more difficult to maintain digital content over time; computer components and media may physically fail due to human error, natural events, and even due to the passage of time. The foremost challenge in digital preservation is therefore to fight digital obsolescence because the storage and software industries are yet to solve the problem of technical obsolescence.

Digital information is also vulnerable to physical threats. Like obsolescence, physical damage can occur to multiple components required to access digital information, namely, hardware and software. New ideas necessitate new and innovative components while new components need new and updated software and operating systems. New software and operating system necessitates new formats, thus the digital world keeps on changing continually. Most of the equipments used in the beginning of this decade are now out of date and so the software or operating system. It is estimated that thousands of software programs written in the early 90s are no longer available.

Relative to print formats, born digital resources are short-lived even under ideal conditions. Equally vulnerable, are the retrospectively converted (*digitized*) information resources from either paper or microfilm formats to digital formats. Many scenarios exist for the delivery of the information. For instance, e-books: the e-book device is connected to another computer that is linked to the Internet. The user goes to a specific web site and selects the desired title. The website could be that of the e-book producer, a portal that represents several publishers, a single publisher, or an academic or corporate site. The e-book has a build-in modem and is directly connected to the Internet by a phone line for downloading. The e-book device could be connected through a kiosk at a bookstore or a library etc. The e-book is connected by a wireless modem to the selected website or other remote location. Whatever the case, the e-book title is stored on a remote storage system and is routed to the e-book directory of the remote computer. No single data location of all e-book files will exist and mergers and personnel change at the hosting site may affect the long-term storage and preservation of the information. A publisher could decide to drop certain titles or go out of business.

Then, the software and hardware used for creating most of the digital materials a decade ago are currently obsolete, completely replaced by the more advanced and latest versions. For example, the recording and storage medium (magnetic tape, disk, etc.); the Operating systems (DOS, Windows, etc.); the data encoding Formats (Binary, ASCII, sound, video, etc.); Data coding standards/systems or Markup languages (e.g. HTML, XML, PDF, etc.); Metadata (e.g. bibliographic or stylistic encoding) are all constantly changing. This scenario poses a great challenge to libraries, archives and other information repositories because no institution has infinite financial resources to cope with the changing technologies especially developing countries running endemically short of adequate financial resources to increase even their traditional paper-based

information resources.

**Storage Media**

Though storage media, a part of digital obsolescence, need separate treatment because the core of digital preservation lies in the storage media. While the genesis of digital resources itself is complex, the format in which they are made available can also pose a challenge. Different digital materials are available or stored in different physical formats such as computer files, compact discs, microfilms, cassettes, and other upcoming storage media.

Firstly, storage media transform and evolve into different and new types of media almost every year; for example, over the years, cassettes have advanced from simple audio cassettes, to video cassettes, to compact disks (CD-ROM) and now to DVD's. Though some optical storage media promise a life span of 30 years and some even 100 years, the continuous innovations in the technological front poses a major threat to the digital preservation initiatives. The current High Density Read-Only Memory (HD-ROM) technology which uses the ion beam to inscribe information on pins of stainless steel, Iridium or other material seems to be the future technology for mass storage media **(Hedstrom, 1998, p.198)**. However, with new inventions rocking the digital world almost daily, it would be presumptuous to conclude that it is the future of storage media.

Secondly, most of the storage media is re-usable. Re-usable recording media for digital materials, such as magnetic and optical media are vulnerable to deterioration and catastrophic loss. They can deteriorate rapidly thus, the time-frame for decisions and actions to prevent loss of content and quality of the recorded information is a matter of years and not decades as is the case with print formats. Most preservationists feel that the limited life of magnetic and optical media poses a significant problem in digital preservation.

Thirdly, the play-back hardware and software technology to access

information in these different media and formats pose yet another problem. The retrieval technology is also rapidly changing due to technological advancements, with increased capacity at each stage of innovation. For example, while the 5 ¼ inches floppy discs were the in-thing a decade ago, they were replaced by 3 ½ inch floppy discs, which have now been replaced by flash discs. The 24 years history of hard disk drives has seen the capacity of these disks increase from a mere 5MB to a monstrous 400GB, and the physical size shrinking from 52 feet diameter to a mere 3 ½ inches diameter.

**Migration**

Migration is another biggest challenge in digital preservation. Digital migration is broadly described as moving into the new or advanced digital standard without compromising the realities of the current standard. It can be defined as "a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another or from one generation of computer technology to a subsequent generation" **(Goodman, 2001, p. 8)**. Migration is the most commonly used preservation strategy, especially for non-interactive digital objects such as images, databases or textual documents.

Before any digital object is migrated, two important decisions must be made, format and application. One must decide on the target format that should be used in order to accommodate the properties or authenticity of the original object; and the application software to carryout the migration process. This activity constitutes the first stage of any migration process; and the optimal combination of target format and migration/conversation software should be the ultimate aim of any digital preservation plan.

**Conversion**

Similar to migration, conversion is another widely used option for

digital preservation. Conversion consists of the reorganization of the information elements that comprise the digital object into the logical structure as defined by a different format. Conversion is transforming the present data into a current and more widely accepted and accessible format **(Hedstrom, 1998, p. 199)**. From the preservation point of view, it consists of setting up a conversion application and executing it against a collection of digital objects. Some scripts may have to be developed in order to automate the whole process. The difference between migration and conversion is that the conversions are applied only when the rollback facilities are not available. The problem of digital conversion is that at times the authenticity of the document could be lost.

**Authenticity and Context**

Probably, the biggest challenge in digital preservation is the preservation of the authenticity of the document's content within the context in which it was created. Technological advancements in hardware and software technologies, the migration and conversion strategies, and technological obsolescence are a threat to digital preservation. Ideally, digital material should be preserved within context, i.e. within the context in which it was created, along with the same software, hardware, document type/format, etc. yet, this is practically impossible because technology is forever changing.

The cost to retain the older technology and maintaining the new and the old could be one of the biggest challenges due to cost implications and other reasons. On the other hand a school of thought also feels that the emulation techniques should be used to capture the same authenticity. Again this will require a lot of resources. The best compromise would be to migrate or convert or re-encode the objects to the latest technology.

**Cost**

The cost implication has an important bearing on digital preservation because no institution has access to unlimited funds. Cost implications depend mainly on the socio-economic dimension of the information along with the storage, hardware and software cost; and cost over the human resources. The ability to employ and develop appropriate skills in digital preservation entails training and re-training of information professionals as the technology changes.

As noted earlier, digital preservation is a continuous process that requires a constant inflow of financial and other resources for any successful preservation project because a break in the middle of a preservation project could result in loss of information and other negative effects. It is also important to note that only co-operative and consolidated efforts in creating strategies and solutions alone will provide cost-effective preservation solution. It is with this backdrop that consortia come into play for the acquisition, maintenance, and preservation of digital materials. It is only through co-operative endeavors that digital preservation can be successful, not only because of cost implications, but also due to the enormity of the exercises. No single institution has the human, technological or financial capacity to embark on a successful digital preservation programme.

**Awareness**

A major strategy for a sound digital preservation is needed to create awareness amongst the digital producers. The value of long-term preservation should be stressed right from the onset, i.e. from the data creation stage as it will influence the choice of the appropriate digital medium, software and hardware for preservation. Similarly, it is important that awareness is created for the existing or current standards in digital preservation including the importance of documentation. In this way, the producers of digital documents can

plan for long-term funding and sustaining any future technological innovations.

## CONCLUSION

The paper has explored the importance of digital preservation and its attendant problems/challenges; such as: the nature of digital material (i.e. whether *"born digital"* or *"made digital"*); obsolescence of the technology used in the creation, storage and access of digital resources; problems of retaining the authenticity and context when migrating or converting digital resources; the rapid technological advancements in ICT; and the cost implications for long-term preservation of digital materials., besides an overview of the current preservation strategies.

Considering the role of libraries in information management, it is imperative that libraries, archives and other information centers be equipped with the necessary skills to handle digital resource management and preservation. The institutions like university libraries, national libraries and national archives especially in the developing countries could start the ball rolling by creating institutional repositories (IR) using the available technologies and equally spearhead national digital preservation initiatives in order to avoid a possible digital blackout of their national digital information resources. It is gratifying to note that the developed countries like U.S.A, the U.K, Japan, Netherlands, etc. have taken the lead in designing national frameworks for digital preservation. These initiatives could serve as benchmarks for institutions in the developing countries to design their own national digital preservation strategies instead of the costly route of re-inventing the wheel by starting from scratch but with some local modifications and innovations suitable to local digital environment and socio economic set up.

## REFERENCES

Abrams, S. L. (2004). The Role of Format in Digital Preservation. *Journal of Information and Knowledge Management Systems*, 34(2), 49-55.

Biswas, S. (2008,). Authorities finally wake up to digital content preservation. *Economic Times*.

Building a National Strategy for Digital Preservation: Issues in Digital in Digital Media Archiving Retrieved December 15, 2008 from http://www.clir.org/pubs/reports/html

Digital Preservation Briefing Paper. (2006). Retrieved December 12, 2008 from http://www.jisc.ac.uk/publications /publications/pub_digipreservationbp.aspx

Digital Preservation Coalition (DPC) Retrieved December 12, 2008 from http://www.dpconlone.org

Digital Preservation Europe (DPE) Strategy Retrieved December 12, 2008 from http://www.digitalpreservationeurope.eu

Guidelines for Digitization and Standards for Digital Preservation Retrieved December 15, 2008 from http://www.ditizationguidelines.gov

Goodman, L. & Palmon, D. (2001,). Technological innovations and the work of the accounting historian. *Accounting Historians Journal*, pp. 1-13.

Hedstrom, M. (1998). Digital Preservation: A Time Bomb for Digital Libraries. *Computers and the Humanities*, 31(3),189-202.

National Digital Information Infrastructure and Preservation Programme" (NDIIPP) Retrieved August 12, 2008 from http://www.digitalpreservation.gov

*New International Webster's Comprehensive Dictionary of the English Language*. (1999). Delux Encyclopedic edition.

Florida: Trident Press.

Searle, M. (2004). Selection for Digital Preservation. *Library Hi Tech.*, 22(2), 119-121.

Smith, B. (2002). Preserving Tomorrow's Memory: Preserving Digital Content for Future Generations. *Information Services and Use*, 22, pp. 133-139.