

# DIGITAL PRESERVATION IN LIBRARIES – WHY AND HOW?

**M. C. Lakshminarasimhappa**

Project Fellow, Dept. of Library and Information Science,  
Tumkur University, Tumkur, Karnataka – 572103(India),

Email: [lakshminarasimha431@gmail.com](mailto:lakshminarasimha431@gmail.com)

**and**

**M. R. Veena**

Support Information Analyst, Tata Consultancy Service,  
Bangalore, Karnataka-560069, (India),

Email: [veenamlisc@gmail.com](mailto:veenamlisc@gmail.com)

## ABSTRACT

Preservation of traditional materials became more successful and systematic after libraries and archives integrated preservation into overall planning and resource allocation. Digital preservation is largely experimental and replete with the risks associated with untested Digital preservation strategies are shaped by the needs and constraints of repositories with little consideration for the requirements of current and future users of digital scholarly resources. Archives contain millions of documents like manuscripts, rare books, paintings, photographs, and historical records. All these constitute human heritage. Paper manuscripts, with the passage of time, are getting fragile and brittle because of various reasons. Lamination does not seem to be a permanent solution of preservation this cultural heritage for posterity. Digitization technology brings with it untold benefits for heritage preservation access. Once a document has been properly digitized, it become immortal and can remain accessible long after the original has ceased to exist. The option of digital access further aids in preservation of originals through reduced need for physical handling. This paper states about need for digital preservation with its objectives in Library environment.

Key Words: Preservation, Digitization, Archiving, Digital Preservation,

## INTRODUCTION

Human past is splendid and full of wonders. The past is known by records, archives, books, and other library materials that constitute documentary sources. There are also other things like monuments, buildings, art objects, and other artifacts. All these constitute human heritage. They need conservation and preservation. Librarians, archivists, curators, chemists, and archaeologist the world over are concerned with the problems of conservation and preservation.

During the last few years, the library and information backdrop has changed considerably. Digital objects have become the dominant way that we create, delineate and exchange the information. Now, librarians and information scientists are anxious to provide access to information 24/7/365, as needed by the information society. Information and Communications Technology (ICT) is one of the major constituents of this divesting change, which is used in acquiring, processing, storing and disseminating information.

Preservation is a part of conservation. Preservation is concerned with problems like the repair, dusting, fumigation, de-acidification, air-conditioning, lamination, binding, and storage of manuscripts, books, films, disks, and optical materials.

## **PRESERVATION AND DIGITAL PRESERVATION**

### **Preservation**

**The American institute for Conservation of Historic and Artistic Work (AIC)** defines as, Preservation is the protection of cultural property through activities that minimize chemical and physical deterioration and damage and that prevent loss of informational content. The primary goal preservation is to prolong the existence of cultural property.

Preservation (in Library) is an “activities associated with maintaining library and archival materials for use either in their original physical form or in some other usable way”.

### **Digital Preservation**

**According to Cornell University Library**, “Digital Preservation encompasses a broad range of activities designed to extend the usable life of machine readable computer files and protects them from media failure, physical loss and obsolescence”.

**According to Russell (1998)**, Digital Preservation is a process by which digital data is perceived in digital form in offer to ensure the usability, durability and intellectual integrity of the information contained therein.

## **CONCEPT OF DIGITAL PRESERVATION**

The most precise concept of Digital Preservation is the storage, maintenance and accessibility of a digital material over a long term, usually as a consequence of applying one or more digital preservation strategies. These strategies may include technology preservation, technology emulation or data migration.

## **OBJECTIVES OF DIGITAL PRESERVATION**

The main ground of digital preservation is to achieve the following objectives:

- Providing continued access to digital material for both born digital and digitized material.
- Ensuring authenticity of preserved digital material.
- Preserve physical media to avoid damage/ deterioration by ensuring an environmental control.
- Changing the digital information in to newer and fresher format, if it is necessary.
- Achieving co-ordination of all efforts that are undertaken for preservation globally in order to achieve more synergy, to avoid redundancy and reduce cost. This provides an effective infrastructure for collaboration by connecting different networks, institutions and individuals that are working in this field.
- Focus the stakeholders on issues that desperately need attention in this area. The basic assumption being that action has to be taken at outset rather than at secondary stages.

## REQUIREMENTS OF DIGITAL PRESERVATION

Digital Preservation combines policies, strategies and actions to ensure that digital objects remains authentic and accessible to users and systems over a long period of time, regardless of the challenges of component and management failures, natural disasters or attacks.

Even though, it is impossible to define all the requirements applicable for all digital preservation needs, since digital preservation requirements depend, for instance, on the type, size and amount of data. It also depends on the goals of each organization, regarding the reuse of data. However, there are several generic and common requirements that can be surveyed, based on what someone in the future would require from information stored today.

1. Digital Preservation requires that a copy (or preservation) of any preserved digital object survives over the system's lifetime, which is usually unknown, but may be as long as decades or even centuries. This can be defined as a reliability requirement. Therefore, a digital preservation system must be designed to store data indefinitely without suffering any data losses.
2. A future consumer should be able to decide if the accessed information is sufficiently trustworthy. Usually, this requires the authenticity assurance of digital objects (which is already a common requirement for tangible objects). Also, the provenance of digital objects should be required, especially its creator or entity responsible for it. Moreover, it is crucial to assure the integrity of digital objects, guaranteeing that their informational content was not modified.
3. Digital Preservation requires that future consumers are able to obtain the preserved information as its creators intended, dealing with obsolescence threats. This requirement encloses several challenges, since a digital object, to be explored, requires a technological context defined by specific software and, in some cases, even by specific hardware.
4. Dynamic collection and environments for digital preservation require technical scalability to face technology evolution allowing, for instance, the addition of new components through incremental updates.
  - Existing static collection (with a fixed size) like, for instance, a digitized historical archive, where no new items will be added, will have a fixed data size.
  - Although it will not be necessary to add new components to increase the storage capacity, it may be necessary to replace components by others with more recent

technology (in order to achieve lower maintenance costs or simply because the initial technology was disrupted).

Fortunately, some typical requirements of normal storage systems are not crucial in digital preservation. For instance, data updates are uncommon because, usually, objects in digital preservation systems are supposed to remain unchanged. Almost all write access to the repository is to either ingest new objects or re-write the existing objects in new migrated formats.

## **DIGITIZATION EQUIPMENTS**

- Wide format scanner which can scan document width up to 42 inches
  - Digital Cameras
  - Scanners (8.5"x14") & Scanners (8.5x11.5")
  - Book scanners with V shaped cradle
  - Dark rooms with lighting equipment
  - Servers: IBM server with an installed storage of 20TB, Scalable up to 48TB
  - Backup equipment: LTOP Tap drives, hard drives, DVD writers.
  - Computers: Desktop and Laptops
- These are required based on the available documents which are going to be digitized.

## **METHODS OF DIGITAL PRESERVATION**

Digital Preservation is concerned with ensuring that records which are created electronically using today's computer systems and application remain available, usable and authentic in future use, so digital preservation consist of preservation more than just the records bit stream for interpret the survival of the records otherwise without interpretation the bit stream is nothing more than a meaningless series of 0's and 1's.

### **1. Preservation of Digital Materials**

Digital media, no doubt, have immense capacity to record information enabling the libraries to provide users seem-less access to information, but it is very fragile in comparison to print media. Archiving of digital information is more complicated than archiving printed information.

Digital preservation means taking steps to ensure the longevity of the electronic document in terms of the following.

- Data (this might be for text, image, video or audio stored in variety of format and standards)
- Index to the data
- Link to other data
- Metadata

- Software (relies upon hardware and Operating System)
- Storage medium

Hence, in the preservation of digital heritage material the following measures taken into account.

- Integrity of Digital material
- Physical presence
- Preservation format
- Digital material functionality
- Authenticity of the materials &
- Provenance.

## 2. Standards for Digital Preservation

Standardization is the secret behind quality, uniformity of measurement, norms, terminology which directly affect mass production that leads to economy of time, space, efforts, material, manpower and money and facilitates and acts as a tool of transfer of information which is required for digital information preservation.

## 3. Standards for Architecture

In a distributed digital environment the management of digital collection cannot be responsibility of just one central organization. In such environment it is important to agree on concepts, definitions and procedures. For this there are several standards on record management architecture.

- ISO/DIS 15489:** is a draft international standard on record management and this standard enable organization to standardize for mainly to design and implementation of record system.
- AS 4390:** In December 1995, Australia became the first country in the world to develop a standard on record management, i.e. AS 4390-1996. Following the approval and release of the standard, the international record management community began work on the development of an International Standard.
- DOD 5015.2.STD:** is being developed by department of defense of USA. The DOD 5015.2.STD implementing and procedural guidance on the management of document management.
- OAIS (Open Archival Information System):** It is developed by Consultative committee for space data system (CC SCS) of the NASA. The OAIS reference model described both the information flow and archival requisite and it is being reviewed as an ISO/DIS. This architecture is implemented by various types of Digital library and Archives.

## 4. Standards for Preservation content

Standards for preservation content should be depending upon ageing process of the semantic and physical recoverability of the document that is being preserved. A standard can only provide such longevity when the standard itself does not change and backward compatibility is provides. More often XML and PDF are put forward as two rivals to preserve a document for long term preservation.

- PDF (Portable Document Format):** PDF is the de facto document standard and is the proprietary of Adobe. It uses the image model of the post script language in order to depict text and image as exact copies of the original. The PDF have two types of format.

- Text based PDF outline font technology of postscript PDL(Page Description Language) for describe format of a page and
- Roster scanned image PDI without text outline font OCR (Optical Character Recognition)

**b. XML (eXtensible Mark-up Language):** XML is subset of the standard SGML (Standard General Markup Language) and is related to the Web language HTML Hypertext Markup Language (XML, 2001). With the help of the XML the structure of a document can be saved in a specific type document so called Document Type Description (DTD). For the specification of the form of document style sheet can be used. Cascading style sheet (CSS), extensible style sheet language (XSL) or XSL transformation (XSLT) can be used. Finally the content of the document can be stored is ASCII format with XML “tags”.

## 5. Standards for Preservation Access

Preservation of the bit stream of the document in some standard format is not enough to preserve a digital document over an indefinite period of the time, so it is necessary for the description of the digital, object or various parts into one object means file/ formats stored in different physical places in the information system and linked by cross reference. Such types of description are called metadata which included the contextual information to manage, retrieve and interpret the electronic information overtime.

- a. Dublin Core:** The Dublin core has especially been developed for cross disciplinary network discovery. It exists upon 15 elements (i.e. Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights) for resources description and discovery.
- b. MARC:** MARC defines the representation and communication of bibliographic and relation definition in machine readable forms. Machine records contain a guide to its data or little “sign post” before each piece of bibliographic.
- c. Z39.50:** Z39.50 is an ANSI/NISO standard for information storage and retrieval under supervision of the Z39.50. Maintenance Agency (Library of Congress) of Z39.50 and its possible successor ISO Z39.50 is protocols which specifies data structure and enable searching or interchanging bibliographic information on different platform in a distributed environment.

## 6. Standards for Interoperability

Interoperability is ability of multiple systems with different hardware or software platform, data structure and interface to exchange data with minimum loss of content and functionality. Therefore interoperability is a critical problem in the networked environment with increase in diverse computer systems software application, file formats, information and users. But it is important for digital preservation in digital library and archives.

- a. ODMA (Open Document Management API):** The Open Document Management API enables integration of proprietary document application into ODMA compliant document

management system. Now many document management software packages are already ODMA compliant.

- b. **DMA (Document Management Alliance):** the DMA is a comprehensive standard for interoperability among electronic document management systems (EDMSs). The Document Management Alliance tries to solve the problem of island of information that created different proprietary DMSs (Document Management Software), e.g. MS-Word.

## ADVANTAGES OF DIGITIZATION

1. **Remote Access:** Connecting people globally by providing continued free online access.
2. **Multiple Accesses:** One of the most important advantage digitization is multiple Access documents. If a manuscript is in physical format, then it can be handled by only one person in the specific period, but through the digitization process, several users can access the specified document at a time.
3. **Preservation and Conservation:** Addressing heritage loss due to environment, ignorance, and destruction.
4. **Dissemination and Promotion:** Saving invaluable treasure to enrich the present and enlighten the future. Dissemination of knowledge and culture via the internet is a 21st century phenomenon.

## BARRIERS IN DIGITIZING

1. Initially no one understood what digitization was, and so everyone was hesitating in submitting his or her manuscript for digitization.
2. There is no doubt that in present times 27 million documents from the government, universities, and personal collection are to be digitized.
3. Initially they used to send requests to the public to send their manuscripts for digitization but now the scenario has changed; the request is being sent by the public.
4. It will take approximately ten years to fulfill the requests that are pending.
5. There is a great need to create awareness among the masses about the preservation of the archives so that this heritage can be handed over to the future generation.
6. Digitization requires a huge amount of money. Therefore, they have introduced a new scheme of “adopt one book” for digitization.

## CONCLUSION

Digital preservation in all its aspects is going to require some form of organizational transformation. Long term preservation of documents is now hot issue in the present time. Accessioning to the digital documents in twenty or hundred years from now will be out of the question, if people not process the bit stream underlying digital documents. Digital documents are not fragile. The speed of technological obsolescence makes digital preservation an important issue for everyone.

Today, we have acknowledged the canonical needs of digital preservation, and the major libraries and archives of the world are concerned with the abiding digital preservation for traditional information materials and born digital information for the future generation. The

ever growing amount of material being available digitally, not only drives the need for feasible access and delivery, but also for preserving digital objects in the medium and long run. But the problem of digital preservation is not static and will continue to evolve with the technological developments.

## REFERENCES

1. American Institute for Conservation of Historic and Artistic Works. (2001). AIC Definition of Conservation Terminology. Washington D.C.: American Institute for Conservation of Historic and Artistic Works (AIC). 21-22.
2. Weber & Dorr, M. (1997). Digitization as a method of preservation? Amsterdam: European Commission on Preservation and Access.
3. Conway, Paul. (1994). Digitizing Preservation. *Library Journal*. 12(1), 42-45.
4. Lynch, Clifford. (1994). The integrity of Digital Information: Mechanics and definitional Issues. *Journal of the American Society for Information science*. 45. 737-744.
5. Russell, Kelly. (1998). Skills requirements of LIS professionals in the new e-world. *Library science with a slant to documentation*.36, 141-149. Retrieved from <http://www.ariadne.ac.uk/issue18/cedars/>
6. Hagg, D. (2002). XML and Digital Preservation. Retrieved from [www.digitaleduc.zamhind.xl/does/xml.pdf](http://www.digitaleduc.zamhind.xl/does/xml.pdf)
7. Arora, J. (2004). Building Digital Libraries: an overview. Conference Proceeding of INFOGOLD. BIT, Mesra.
8. Dempsey, L & Heery, R. (1998). Metadata: a current view of practice and issues. *Journal of Documentation*. 2 (54), 145-172.
9. Pandher, Balbir kaur. (2012). Conservation and Preservation of Archives: A case study of Punjab Digital Library. *World Digital Libraries*. 5 (1), 75-84.
10. Ghani, Syed Raiyan. (2012). Digital Preservation: A paradox or sophism. *SRELS Journal of Information Management*. 49(5), 529-534.
11. Tanwar, Sundar Singh. (2012). Digital preservation. *PEARL- A Journal of Library and Information Science*. 6(2), 83-88.
12. Halder, Sambhu Nath. (2013). *Library and Information Science in Changing Paradigm*. Ess Ess Publication: New Delhi, 207-215.
13. Digital Preservation, available at [http://en.wikipedia.org/wiki/digital\\_preservation](http://en.wikipedia.org/wiki/digital_preservation). accessed on 28 October, 2014.
14. Hedstrom, M. (1997). Digital Preservation: A Time bomb for Digital Libraries. *Computers and the Humanities*. 31(3), 189-202.
15. Preservation (library and archival science) available at: [http://en.wikipedia.org/wiki/Preservation\\_\(library\\_and\\_archival\\_science\)](http://en.wikipedia.org/wiki/Preservation_(library_and_archival_science))