

Digital Preservation in Libraries: An Overview

Imran Khan^{1,*}

¹Librarian, Zakir Husain Delhi College (University of Delhi), Jawaharlal Nehru Marg, New Delhi, India

Abstract

Preservation of conventional resources became more flourishing and organized when the libraries and archives started incorporating preservation into their overall planning and resource allocation. Digital preservation is fundamentally experimental as well as full of risks linked with untested Digital preservation strategies. Archives hold millions of documents like manuscripts, rare books, paintings, photographs, and historical records, which represent human heritage. With the passage of time, paper manuscripts start getting fragile and brittle because of various reasons including environmental and climate changes. Lamination does not seem to be a permanent solution of preservation of this cultural heritage for future generations. Digitization technology has a lot of benefits associated with it for the effective preservation and future access of this heritage. A document becomes perpetual and can be accessed for long after it proper digitization irrespective of the existence of the original document. Another benefit of providing access of a document in digital format is that it results in considerable reduction in physical handling of the original hard copy of the document. Digital Preservation is a set of various organized procedures and activities, which are required to be carried out for the purpose of ensuring everlasting availability of the material in digital format for its usage. It is a challenging task because of being a complex process. The present paper deals with the basic concept of Digital Preservation, its need, relevance, objectives, and strategies. It also discusses about the overview of technological approaches and strategies to digital preservation and its benefits and challenges in the libraries.

Keywords: Digitization, digital preservation, digital library, preservation planning, digital resources, preservation strategies

***Author for Correspondence** E-mail: ikhan@zh.du.ac.in

INTRODUCTION

In the present era of digital and virtual environment, it has become essential to preserve the treasure of information and knowledge for its sustainable use in future. There are several records, archives, books, and other library materials that constitute important documentary sources. In view of their great importance and relevance it is necessary to preserve them in such a format that has long life. This has led towards the concept of Digital Preservation. For digital preservation of the documents, the first step is digitization. Digitization is the process of converting a hard copy of a printed document into electronic format. By way of this conversion and storing the electronic copy on a server or a medium of central storage, the document can be accessed and viewed by

many users at the same time. Usually the digitized materials are made available on Internet for their wider accessibility. The documents and other materials that can be digitized include books, letters, manuscripts, photographs, maps, drawings, etc. The institution having collection of the digitized materials with a facility for the users for their retrieval and access is termed as Digital Library. Digital Preservation provides opportunities facilitating long term storage and access of the digital objects and documents without any limit of time or geographical location. Digital Preservation is a long-term error free storage of digital information. The process is complex because it requires technical skills besides having practical knowledge of ICT, Computer Science, and handling of various digitization equipment.

DEFINITIONS OF DIGITAL PRESERVATION

Digital Preservation is the process of dynamically securing the information stored in digital format using such procedures and practices that ensure its safety both from failure of various mediums of storage as well as software and hardware obsolescence. There are several other definitions too. According to Arora (2006) [1], the term digital preservation refers to preservation of materials that are created originally in digital form and never existed in print or analogue form, which are also prominently known as born-digital, and also those which are transformed from ancient documents and antiquities like photographs, printed documents and/or physical objects etc., into image formats using various imaging techniques like scanning, taking pictures using digital cameras etc. for access and preservation purposes. According to Cornell University Library (2005)[2, 3], digital preservation refers to a series of managed activities designed to ensure continuing access to all kinds of records in digital formats for a long duration as per their requirement and protecting them from any type of crashing of media, physical damage and staling.

Digital Preservation Coalition (2006) [4] describes digital preservation as a set of all the activities necessary and required to sustain access to digital objects and materials over and above the limits of media failure or change in technology. Those objects and materials may be digital records developed during the day-to-day affairs and operations of an organization, i.e. born-digital resources created for a particular purpose (like teaching resources), or the products of digitization projects. According to ALA (2007) [5] digital preservation combines policies, strategies and actions that ensure access to digital content over time. While according to Glossary of LIFE [6], digital preservation is a process of ensuring that a digital object is accessible over the long term.

NEED FOR DIGITAL PRESERVATION

There are innumerable documents in printed hardcopy form, manuscripts, archives, etc.

which contain a big amount of treasure of knowledge and hence need to be preserved for future reference and use because of their relevance and importance for the society. Digital resources play vital roles in fulfilling the needs and requirements of the users and knowledge seekers working in the field of higher education and research. In the present times education and research are heavily dependent on the digital contents and information in digitized form. People now require information 24x7x365 irrespective of their location. This is possible only if the same is available in digital format. Conventional libraries now are increasingly getting transformed into digital libraries. The availability of information in digital form via Internet is putting heavy pressure on the traditional libraries due to which they are allocating a large portion of their budgets for acquiring or accessing web-based online or full-text search services, CD-ROM products, online databases, multimedia products, etc. In addition to acquiring and buying access to digital information and collections, libraries are attempting to initiate digital library projects in their respective institutions to build their own digital collections. The libraries are heavily converting their existing print collections into digital formats. The digitization of all forms of cultural, educational and artistic heritage is becoming increasingly popular at all levels of society. The safeguarding of precious information is an important responsibility that people must accept for the sake of future generations.

OBJECTIVES OF DIGITAL PRESERVATION

While preserving a book or any document in print format, one needs to preserve the paper on which the book or the document is printed, in order to preserve the scholarly information contained in it. On the other hand, preserving a book or a document in electronic format, its medium of storage can be changed or refreshed any number of times by way of replacing the original CD-ROM or any other medium on which the same has been recorded/stored. The basic objective behind digital preservation is the preservation of

access to the intellectual contents rather than the physical object or medium. The other objectives of Digital Preservation are as under:

1. Preservation and providing persistent access to digital information present in the form of both the born-digital and digitized material.
2. Ensuring authenticity of the preserved digital materials.
3. Preserving the integrity of the digital objects.
4. Converting/changing, if necessary, the format of digital materials in order to preserve their scholarly content.
5. Preventing from harm and corrosion of the physical media by ensuring an environmental control.
6. If possible, undoing the damage caused to the materials.

STRATEGIES OF DIGITAL PRESERVATION [7, 8]

Digital preservation comprises of selecting and executing an expanding range of strategies to accomplish the long-term accessibility of the different types of materials and information. Several strategies have been proposed but it may not be possible to find a single solution suitable for all data types, situations, or institutions. Some of such strategies are as under:

Long-term Preservation

Continued access to digital materials, or at least to the information contained in them, indefinitely.

Medium-term Preservation

Continued access to digital materials beyond changes in technology for a defined period but not indefinitely.

Short-term Preservation

Access to digital objects and materials for a definite time period to which their requirement is anticipated but not for a very long time and/or till the time it gets inaccessible because of changes in technology.

In libraries, we are more concerned with the long-term preservation of the important and rare documents and materials and therefore,

for the purpose of long-time access, following preservation strategies can be suitably adopted for the preservation of digital objects:

Selection of the Material

Digital preservation should be based on the condition, value, utility, and rarity of the material. On the basis of these parameters the objects, documents, or material should be selected for carrying out their preservation.

Digitization

It is a highly cost consuming preservation process which can either be carried out in-house or outside the organization by way of outsourcing. It consists of evaluation, estimation, and selection of the source material, digitization assessment, full digitization quality assessment, editing and reviewing, application of metadata, etc.

Refreshing

It is the process of storing the digital information from one long term storage medium to another storage medium of same type so that there is no change or alteration of the data. It is always very important due to the deterioration of physical media. It is a process to overcome the decay and obsolescence issues related to the storage media. This strategy will possibly require to be used jointly with Migration in case if the hardware or software necessary to read or interpret the data is not available or is not being able to recognize the data format.

Migration

It is a periodic process which holds the core concept of a digital data during conversion from one hardware or software to another newer form and preserving its opacity and veracity. It may be considered as an enhanced form of refreshing but differs from it in the sense that exact copy of any database or any digital object is not possible due to hardware or software change.

Canonicalization

It is the process of converting data having more than one possible representation or version into a standard canonical

representation. This may be carried out in order to:

- i. compare different representations or versions for similarity.
- ii. count the number of distinct data structures.
- iii. improve the efficiency of various algorithms by eliminating repeated calculations; or
- iv. to make it possible to improve a meaningful sorting order.

Replication

It is the process of creating duplicate copies of data from one format to another. It is necessary to save the data from natural, social and technological disasters for life-long usage and perpetual access of digital objects. The purpose of replication is to increase the durability of digital documents while maintaining their legitimacy and veracity through copying and the use of multiple storage locations.

Emulation

It is the process of replacement of functionality of an obsolete system by a newer one. It is a trendy strategy for retaining the functionality of the older system. It involves formation of emulators that translate code and instructions from one computing environment to the newer one for its proper execution in the later one.

This strategy is technically complex, cost-intensive, labor-intensive and requires a high degree of expertise. Moreover, it is still in the research and testing stage. Insufficient documentation of software or by non-standard use of file formats may disrupt effective emulation. Another problem is that emulation of all features of a system or application may not be possible.

Encapsulation

It is a method of grouping digital objects and metadata essential to provide access of digital object. Digital objects include reference, illustration, attribution, fixity, context information with appropriate types of metadata for the encapsulation, which is considered to be a necessary component of emulation.

This method stresses that preserved objects should be self-describing, virtually linking content with all the information required for it to be decoded and understood. The files connected to the digital object would have details of how to understand that object by using logical structures called containers or wrappers to provide a relationship between all information mechanisms that could be used in future advancement of emulators, viewers or converters through machine readable terms. The method of encapsulation is usually applied to collections that will go unused for long periods of time.

Precautions to be undertaken for preserving digital objects:

- There should be a periodical check to ensure that the digital object is not physically disintegrating or becoming compromised.
- Avoid large and rapid fluctuations in temperature and humidity at the place where digital objects are kept.
- Control dust.
- Avoid exposure to fumes.
- Avoid exposure to magnetic fields (*especially for magnetic media*).
- High quality original images should preferably be stored on archival quality.
- Eating, drinking and smoking should be prohibited in media storage area.
- Media should be stored in electrically grounded closed metal cabinets.
- Media should be shelved vertically.
- Media should be stored in their original cases.
- There should be minimum exposure to sunlight in media storage area.
- Prior to the usage, the media should be allowed to acclimate to new temperature and humidity, and it should be returned to the controlled storage soon after its use.

CHALLENGES FOR PRESERVING DIGITAL CONTENTS [9, 10]

A lot of historical information and knowledge about human society exists in several different types of materials like stone, vellum, bamboo, silk, and paper. In the present time, a large amount of information is available in digital

formats, including e-mails, blogs, social networking websites, websites of national elections, web photo albums, and sites which periodically change their contents. Because of the availability of various digital and digitizing media it is quite easy to develop, create and keep up-to-date the content but still there are a lot of technical and economic challenges associated with the access and preservation of these contents. Some of the challenges are:

Technological Obsolescence

With the evolution of new technologies, older ones get obsolete resulting in stoppage of their usage. New media for storing digital information rapidly replace older media and in the same way the devices used to read them also get replaced by the new ones. Moreover, the older or obsolete hardware and software have to be replaced by the compatible versions. Accordingly, the information, which depends on obsolete technologies, becomes inaccessible. Technological obsolescence is regarded as the biggest threat to ensure continued access to the digital material. The speed, with which the changes in technologies are taking place, it appears that the timeframe during which preservation action must be taken is shorter than the paper. Most often, the Technological Obsolescence is also termed as Digital Obsolescence.

Physical Deterioration

An important challenge faced by digital preservation is that the media, on which digital contents are stored, are exposed to deterioration and catastrophic loss. Acid paper also gets deteriorated by way of becoming brittle and yellow. But this deterioration does not become apparent for a very long time. However, if deterioration is observed, it is very much possible to secure the information without any loss. The recording media for digital data deteriorate at a much more rapid pace, and once the deterioration starts, in majority of cases the data gets lost.

Complex and Dynamic Nature of Digital Contents

Digital contents can also pose challenges to their preservation because of their complex and dynamic nature, for example, interactive

Web pages, virtual reality and gaming environments, learning objects, social media sites, etc. In many cases of developing technological advancements, there are a lot of difficulties in maintaining the accuracy, fixity, and veracity of objects over time deriving from the primary issue of experience with that particular digital storage medium and while particular technologies may prove to be strong in terms of storage capacity, there are challenges in providing steps to ensure that the object remains fixed during the period of usage.

Cost Consideration/Economic Challenge

Digital preservation requires right clearance of digital as well as printed materials for digitization or creation of their replica, which besides being time consuming will also be costly and, in some cases, even impossible. Preservation programs require considerable investment to create, along with ongoing costs for data ingest, data management, data storage, and staffing. The factual and the most important challenge with these programmes is the requirement of a lot of funds in the current/present time but their actual benefit will accrue to the coming or future generations.

Machine Dependency

Digital contents are machine dependent. It may not be possible to access the information without appropriate and compatible hardware and related software.

Fragility of the Media

The storage media used for storing digital contents are inherently unstable and highly fragile because of problems inherent to magnetic and optical media that deteriorate at a rapid pace and can quickly fail while exposed to high temperature, moisture, pollutants in the atmosphere, or defective reading and writing equipment. A lot of storage equipment and devices, if not placed under appropriate storage environment and management, are vulnerable of getting quickly damaged or deteriorated without even showing any signs of external or physical damage. This deterioration and damage of the medium of storage may result in corruption of digital files

in such a manner that the corrupted segment of the digital contents may not be identified easily. Furthermore, if the digital contents are not treated properly for their preservation at initial stage, they may become unusable in near future. Besides inadvertent corruptions, digital contents are open to deliberate corruption and exploitation. Since it is quite easy to alter and amend the digital contents, it is necessary to take such steps that ensure the continued veracity, dependability and history of digital contents.

Formats and Styles

Information contents that were earlier restricted to conventional formats like books, maps, photographs, and sound recordings are getting increasingly available in variety of digital formats. New formats have emerged, such as hypertext, multimedia, dynamic web pages, geographic information systems and interactive video. There are different challenges associated with each format or style with respect to their encryption and compression for the purpose of digital preservation.

Copyright and Intellectual Property Rights (IPR) Issues

Intellectual Property Rights (IPR) have a significant impact on digital preservation. The IPR issues for digital materials and digital contents are much more complicated than for printed materials. IPR issues in digital settings have an impact not only on digital materials and digital contents but also on any software associated with them. For the purpose of long-term preservation, use and access there may be a requirement for migration of digital contents and digital material into new forms or emulation of the original operating environment. This will necessitate appropriate legal permissions from the holder of original rights of the contents and the underlying software. Furthermore, merely stimulating digital materials and digital contents on a new medium, encapsulating contents and software for emulation, or migrating contents to new hardware and software, may result in the violation of IPR in the absence of necessary statutory exemptions or specific permissions obtained from the rights' holders.

Furthermore, since migration and emulation may involve manipulation and changing presentation and functionality to some extent, it is important that these issues are addressed with the copyright holder of the contents during negotiations ensuring preservation of selected items. Some of the additional complexity in IPR issues relates to the fact that digital materials can be copied and distributed easily. Rights holders are, therefore, concerned with controlling access and potential infringements of copyright. Technology developed to address these concerns can also inhibit or prevent actions needed for preservation. These issues of access, violation and preservation are required to be realized by the establishments involved in preserving digital contents and digital materials and examined and reviewed by both the parties while pursuing negotiations regarding rights and procedures for preservation.

Trained Manpower

Digital preservation requires new sets of skills, workflows, and close cooperation across different professions from conventional preservation management to computer science. Competent and skillful manpower having exposure to digital preservation technologies will be required to cope up with the problems involved in the preservation.

DIGITAL PRESERVATION POLICY [11]

It is advisable that a Digital Preservation Policy should be formulated as a first step towards digital preservation of the documents and other objects. The digital preservation policy can be highly useful in assuring many benefits at each institutional level, such as, ensuring digital materials available for current and future use, providing an all-inclusive statement on this subject and planning logical digital preservation programs. Besides, the formulation of a policy allows to deal with difficult subjects as the short life span and small capacities of digital materials, the obsolescence of the hardware required to access them, the obsolescence of software for reading the data and file formats and, finally, the structural

and technical diversity of the different types of digital contents and digital materials.

The digital preservation policy must be structured in several specific and distinctive areas, such as authority and responsibility; conversion and reformatting; appraisal, selection and acquisition; storage and maintenance; access and dissemination; implementation; standards; procedures; quality control, auditing and benchmarking; cooperation; and technical infrastructure. The policy should clearly state and explain in detail the purpose and scope of the programme strictly keeping in view the directives of the repository, possible and anticipated legal pressures, value of the digital object/material, and its expected use in future. A special area should be dedicated to the cooperation between institutions in the policy process; usually, the cooperation regarding archives, libraries, museums or other repositories can be local, national or even international and providing that the work and engagement can be distributed equally or in a different way between participant members. Moreover, the standards pertinent to preservation are of great importance because they assist in cooperation and grasping the knowledge and experience of other initiatives. Therefore, a policy is required to include the objective to stick to appropriate standards. Another essential section concerning with the responsibilities involved particularly with reference to the implementation and the related human resources and tools, such as management, employees, special task force, external advice, resources or models; and in some cases, outcome of internal analyses, risk analysis, are the major actors in the drafting of the policy. A precise list of threats intrinsic in systems that preserve digital materials can help in preparation of a more comprehensive policy on these themes. It is therefore essential to highlight that a digital preservation policy should seek to reduce the risks associated with technological changes and permit other changes. This will greatly help in preserving various objects and materials in digital form to be used and accessed for a very long period of time.

IMPLEMENTING PRESERVATION STRATEGY [12]

Important steps needed to be followed while implementing preservation strategies are given below:

Identification of Material

Identify and select digital materials that require preservation treatments.

Explore Suitable Preservation Strategy

Explore the hardware and software technologies necessary for effective execution of the ideal preservation technique. Different preservation strategies may have different types of requirements for their implementation. For example, in case of Emulation, there may be a requirement of developing a specific software having the ability to recreate the source records within a new computer environment. In the case of migration, there may be a requirement of identifying appropriate migration paths (i.e. software applications with adequate backward compatibility to relocate source records from an obsolete data format to a current data format). In the case of encapsulation, this may involve software with the ability to incorporate metadata or package it with the record.

Testing of Proposed Process

Prior to going ahead with the implementation of a preservation approach, it is highly advisable to carry out proper testing of the technical processes involved in it on the duplicates of the original source records as a necessary precautionary measure to avoid any chance of causing damage to the original records.

Creating Back Up of the Records

It is also necessary to create back up of all the digital records identified for preservation treatment confirming its veracity to serve as master copies to be used if for any reason the preservation process gets failed.

Application of the Preservation Treatment

After the successful testing, the preservation treatment should be applied to all the identified digital records. The preservation treatment will be different for different

preservation strategies. For example, for migration and encapsulation methods, there will be application of preservation treatments to the source records, and thus, changing their formats while in case of emulation, the records identified for preservation will be transferred to the new environment without causing any change in these records.

Reviewing and Testing of Preserved Records

There should be rigorous testing and reviewing of the preserved records for making sure that there is no loss of contents or alteration in their structures or formats. The veracity of all the significant metadata linked with the preserved records should be confirmed. For the purpose of recording the process and procedure of preservation, the metadata is also required to be updated.

Destroying Duplicate Source Records

After the successful completion of the preservation process and verification of the integrity of the preserved records, duplicate source records may be destroyed.

Monitoring of the Preserved Records

There should be periodical monitoring of the integrity, functionality, structure, content and context, and associated metadata of the preserved records after preservation to ensure the stability of the preserved records and to find out whether there is requirement of any further preservation treatment.

Benefits and advantages of digital preservation of library resources:

- Digitization brings the library near to the user. They can get their required information and documents directly on their portable devices like PC, Laptop, Tablet, Smartphone, etc. having an active Internet connection.
- Provision of remote and multiple access of the digital documents is possible 24x7x365.
- Comfortable and improved access, searching and browsing of digital resources.
- Easy sharing of information to everybody by way of placing digital information on a network.

- There is no limitation of time and geographical boundaries in accessing the digital resources.
- Even after rigorous use the digital resources are never checked out, misshelved or stolen.
- Digital preservation supports many features and manipulations, which are not possible in print documents.
- A digital library can meet simultaneous access requests by many users for the same digital resource by easily creating its multiple instances and thereby meeting the requirements of much larger population of the users.

CONCLUSION

There has been a significant change in the library and information environment during the last few years. Digital objects have become dominant the way we create, define and exchange the information. Now, librarians and information scientists are concerned with providing 24x7x365 access to information, as needed by the information society. Information and Communications Technology (ICT) is one of the major components of this drastic change, which is being extensively used in acquiring, processing, storing and disseminating information. The academic society is moving at a great speed into the digital environment and the digital scholarly contents, which are being created nowadays, must be preserved for the future because without assured access to these records for the future use, it will become quite difficult to carry out further research. Conventional preservation procedures are not suitable for digital contents. Preservation in the digital world is a difficult job for the librarians and archivists. However, the various technologies, procedures, techniques, strategies and protocols used and involved in Digital Preservation have now been well defined and understood. Digital Preservation is a very costly activity which is required to be undertaken on regular basis. It will be a very difficult task for the knowledge institutions like libraries, archives, or museums to go ahead with the long-term preservation and storage of research collections without deep and continuing commitment for the same by

the parent institutions. The essential monetary and technological obligations to maintain digital contents for the use of future generations are required to be an organizational commitment. Any negligence towards the well-defined digital preservation problems and strategies may cause loss of valuable digital data and may contribute to cultural and intellectual loss resulting in very high costs for recovery. There are a lot of technical problems involved in the long-term preservation of digital information like, short media life, outdated hardware and software, slow read times of old media, and defunct websites. However, proper planning and careful implementation will be of great help in carrying out preservation process in effectively. The preserved resources will be of highly significant value for the future generations.

REFERENCES

1. Arora, Jagdish. (2006). Digital preservation: an overview. In INFLIBNET Centre. *Proceedings of PLANNER 2006*, Ahmedabad: INFLIBNET Centre, 2006.
2. <http://www.library.cornell.edu/>
3. <http://dpworkshop.org/>

4. <http://www.dpconline.org/>
5. <http://www.ala.org/alcts/resources/preserv/defdigpres0408>
6. <https://discovery.ucl.ac.uk/id/eprint/1854/1/LifeProjMaster.pdf>
7. https://en.wikipedia.org/wiki/Digital_preservation
8. Arora, Jagdish. (2006). Op.cit.
9. https://www.newworldencyclopedia.org/entry/Digital_preservation
10. Arora, Jagdish. (2006). Op.cit.
11. Velmurugan, Chandran. (2013). Digital preservation: issues and challenges on libraries and information resource centres in India. *e-Library Science Research Journal* 1(8). pp. 1-8.
12. Arora, Jagdish. (2006). Op.cit.

Cite this Article

Imran Khan. Digital Preservation in Libraries: An Overview. *Journal of Advancements in Library Sciences*. 2020; 7(2): 40–48p.