

OPPORTUNITIES AND CHALLENGES IN CREATING DIGITAL ARCHIVE AND PRESERVATION: AN OVERVIEW

Deepak Yadav

Phd Student

University of Aveiro, Portugal

E-mail: deepakyadav.research@gmail.com

Abstract

This paper discusses the basic introduction of digital archive and preservation and their advantage. It further addresses various basic principles which should be followed in the creation of digital archive and elucidates various significant steps of digital archiving. It also provide a generic discusses on various challenges in creation and maintenance of digital archives for long time preservation along with various issues of digital archive and preservation in developing countries based on selected available limited literature.

Key words: Digital archive, Preservation and Opportunities and challenges.

1. Introduction:

Digital archive is considered as a different form of the traditional archive. Traditional archive practice refers to preserve physical objects (e.g. artifacts, samples, papers, photographs, and microfilms) that carry information. In another way, it may be explained that an archive is often the place in the organization that is required to retain and organize records of the organization that has historical and endure value. The archive serves the purpose of organizing and maintaining records or materials which are accumulated over the course of individual or organization's life span. The archives of an individual may include letters, papers, photographs, computer files, scrapbooks, financial records or diaries, created by the individual in all media or format. The archive of an organization (Such as corporation or government) has a tendency to retain and organize other types of records i.e. administrative file, business records, memos, office correspondence and meeting minutes. Generally, an archive consists of records which have been selected for permanent or long-term preservation for their historical value, enduring research value, operational needs and risk of content loss. Digital archive seeks to preserve the information in digital format regardless of media on which that information is stored (Niven 2011). Digital information or materials are surprisingly fragile. Their sustainability depends on technologies that hastily and frequently change. Changes in technologies will ensure that over relatively short periods of time, both the media and the technical format of old digital materials will become unusable and keeping digital resources usable by future generations requires conscious effort and continually investment (Flicker 2003). There are two main reasons that are responsible for fragility of digital records. One reason is technological obsolescence in which software programs and

other technologies can be superseded by a newer one and consequently old technologies and software become out of use. Once newer technologies become accepted as the norm then it becomes difficult to use any digital object that exists in older format. Another reason is media degradation. Computer disks and other magnetic and optical media on which digital information are stored may degrade and consequently, the digital information stored on such media is buried with media degradation. Digital preservation deals with aforementioned problems. Digital preservation is the management and maintenance of digital objects so they can be accessed and used by future users. The main goal of digital archive ensures the long-term preservation of digital data so that it remains accessible for appropriate use in future. This paper discusses the advantage of digital archive over traditional archive and addresses basic principles which should be followed in the creation of digital archive and elucidates various significant steps of digital archiving and finally discusses various challenges which faced by archivists in creation and maintenance of digital archives.

2. ADVANTAGES OF DIGITAL ARCHIVE OVER TRADITIONAL ARCHIVE.

There are several advantages and benefits of digital archive over traditional archive. These advantages can be taken by archivists or organization by creating a digital archive for their significant resources. Digital archive brings all these advantages and benefits to its users, some of them are as follows:

- Digital archive does not require any new building; information sharing can be enhanced and redundancy of collection can be reduced.
- Digital resources can be accessed through the internet from anywhere, at any time and according to individual choice. It provides faster access to all the form (text, image, graphic, audio and video) of information of onsite and offsite resources on the computer or mobile devices.
- Digital archive can meet simultaneous access request for a document by easily creating multiple copies of the requested document and facilitating storing and printing the resources at users place.
- Archived records or objects in digital format can be managed and administrated easily as compared to a traditional management system.
- Ensuring the long-term preservation and conservation of digital data or materials so that it remains accessible for appropriate use in future.
- Allows easy, wide access and recognition to archived resource throughout the globe.
- Supports the democratic considerations by making public records more widely accessible.

3. THE PRINCIPLES OF ARCHIVING OF DIGITAL RECORDS:

There are some principles or key issues which must be considered by archivists in creation and maintenance of the digital archives. These are as follows:

- Existing digital records should be safeguarded and stored in an appropriate digital archive.

- During creation of a new digital archive for an organization, the existing standards and guidelines should be conformed that how digital records will be structured, preserved and accessed.
- Each digital copy produced for the archival purpose must be checked against the imported file (original file), it must be free from uncorrectable errors or have the lowest possible number of uncorrectable errors. An error status report should also be produced and should be kept for future monitoring.
- Each carrier containing digital records must be checked at regular intervals for data integrity.
- Digital contents must be copied to new carrier whenever the number of errors increases significantly- at any rate before uncorrectable errors occur ("Refreshment").
- Digital content must be copied before the old carriers, formats, and/ or hardware become obsolete ("Migration")
- At least two digital preservation copies of digital record must be kept to access at appropriate time. The preservation copies should be kept in different locations whenever possible.
- Both digital preservation and digitization involve the copying of the content. Therefore, copying to preservation is subjected to copyright legislation. Digital archive must comply with intellectual property rights, other legal and moral rights related to copying, storage, modification of content, and the use of specific digital assets.
- The archive must prioritize items for digital preservation according to their historical significance, operational needs and the risk of content loss.

4. SIGNIFICANT STEPS OF DIGITAL ARCHIVING:

Digital archiving is not only storing the digital information for future but it organized and managed it in a systematic way so that it can be accessed whenever needed. Digital archiving involves various steps in order to organize and manage the digital information, which are as follow:

- Creation
- Acquisition
- Cataloging/identification
- Storage
- Preservation
- Access

4.1. Creation:

Creation of digital object or information is an act of producing the information product. Almost all archival records or objects become available in analog format as a first-hand fact, data, and evidence. Creation of digital archive requires at an initial stage to convert them into digital form for the purpose of long-term archiving and preservation and providing online access on the global scale. For this purpose, digitization process converts the archival resources in digital form. The digital material may be of two types, one that

produced in analog format (e.g. printed books, manuscript etc.) and then converted to digital form. The other one is originally produced in machine-readable digital form (e.g. digital photographs, websites, and multimedia etc.). A wide range of sophisticated technologies and tools are involved in this process like scanning, optical character recognition (OCR), and markup language, audio, video and multimedia technologies, metadata and PDF (Portable Document Format). The preservation and archiving process of digital resources may be made more efficient, if attention is paid to issues of consistency, format, standardization and metadata description in very beginning of the creation of digital documents.

4.2. Acquisition:

Acquisition and collection development are a significant step of digital archiving in which the created digital objects or information are integrated physically or virtually into an archive. A few aspects must be considered in acquisition of digital objects:

i. Collection policies:

Collection development policies must be determined to incorporate the digital objects in a digital archive. That will help the archivist whether digital materials are included under current deposit legislation. Set policies guideline helps to establish the boundaries an unregulated situation. Out of several available materials, all the materials are not important to be archived. In such situation, a guidelines are needed to tailor the general collection practices of the organization.

ii. Selecting what to archive:

A guideline must be prepared for selecting the materials or information to be incorporated into digital archive because various legal and other issues are also associated with the material. In such situation, guideline helps archivists to take the decision of whether what material should be archived or not. National Library of Canada (NLC) and the National Library of Australia (NLA) acknowledge the importance of selection guideline (Kumaravel 2004). Material related to organizational enduring value, operational value, and historical signification should be archived on a selective basis for long-term preservation.

iii. Refreshing the archived contents:

When digital contents are archived then it requires refreshing the archived content after sometimes because changes and updates occur in file format, software, and hardware with time, therefore, time to time contents must be refreshed and made them compatible with accordance to software and hardware. Beside this, a balance must be made between the completeness and currency of archive and the burden on the system resources. Obviously, the burden of refreshing the content increases as the number of sources stored in the archive increases.

iv. Copyright and intellectual property concerns:

Intellectual property remains a key issue in the acquisition, preservation and digitization process. The IPR acts aim a three-level protection viz legal-through legislations like copyright laws; technological through digital rights management system (DRMS); and legal protection to help technological protection through prohibition of acts of avoidance of copyright laws (Maji & Das 2009). Technologies have been developed to protect the content through watermarking, finger printing and tamper proof hardware and software; access control by user ID and password; content use through disabling printing and downloading, copying specified number of times only and restricting copying through originals (master) only. Both digital preservation and digitization involve copying of content. Therefore, copying for preservation is subject to copyright legislation. Digital archive must comply with intellectual property rights and comply with other legal and moral rights related to copying, storage, modification of content, and the use of specific digital assets. Copyright clearances for digital preservation must be secured for existing holdings in an archive (NMA 2012).

4.3 Identification and cataloging:

Once, digital objects have been acquired by archive. It requires identification and cataloging. The archiving organization is allowed by both identification and cataloging to manage the digital objects over the time. A unique key is provided by identification for finding the object and linking that object to other related objects. Cataloging and identification practices are often related to what is being archived and the resources available for managing the archive. Some important issues are discussed here:

i. Metadata:

Some forms of metadata are used by all archives for description, reuse, administration, and preservation of archive object. The use and accessibility of the digitally archived materials depend on its metadata strategy. Metadata improves discovery and access of archive object. Metadata is the key to ensuring that resources will survive and continue to be accessible into the future. The main purpose of metadata may be expressed in a digital archive as information retrieval, management of information services, documenting ownership and authenticity, interoperability. Digital objects are described, structured, summarized, managed and otherwise manipulated in surrogate form through the use of "metadata" which literally means data about data or information about information (Suresbabu 2005). Mainly metadata is created to facilitate discovery of relevant information. Besides these, it helps to organize electronic resources, facilitates interoperability and legacy resources interrogation, and supports identification, archiving and preservation. Archiving and preservation require special elements to track the lineage of digital objects (Where it came from and how it has changed over time), to details its physical characteristics and to document its behavior in order to emulate it on future technologies.

ii. Digital identification:

The location of a digital object may also be given using a file name, URL or some more persistent identifier such as a persistent URL or the Digital Object Identified (DOI). A

persistent identifier is preferred because file locations change frequently, making the URL invalid when the digital material is not immediately copied into the archive. Digital material moves from server to server or from directory to directory on the network, resulting in a change in the URL. It is a problematic issue in creating a digital archive. The use of the server as the location identifier can result in a lack of persistence over time both for the source object and any linked objects. Despite possible problems, most archives continue to use the URL when referencing the location of the digital object.

4.4 Storage:

Selection of physical storage for digital material in an archive is also a serious issue. Any physical storage medium is completely dependent upon very specific combination of hardware and software for access (Brown 2008). The accessibility of information stored on such media is highly vulnerable in today's rapidly evolving technology environment. Technological obsolescence is also an unavoidable and all media have limited life spans. In the near future, it is required to migrate from old storage system to new storage system which is a costly process. It can be reason of loss of archived data. At the time of selecting storage media for the digital archive, a few points like longevity, capacity, viability, obsolescence, cost, and susceptibility must be taken into consideration by archivists or data creators.

4.5 Preservation:

Preservation is a process by which digital objects is preserved in digital form in order to ensure the usability, durability and integrity of the information contained. The digital preservation is an aspect of archival management which encompasses the policies, strategies, planning, resource allocation and action to ensure access to reformatted and "born digital" content regardless of the challenges of media failure and technological change. Application of preservation methods and technologies are required to ensure the protection of information of enduring value for access by present and future generations. The main purpose of preservation is to protect digital objects for a long term in order to pass down to future generations. A set of strategy has been developed by OCLC (The Online Computer Library Centre 2006) for this purpose which includes following focused points:

- i. Assessing the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software application.
- ii. Evaluating the digital content objects to determine what type and degree of format conversation or other preservation actions should be applied.
- iii. Determining the appropriate metadata needed for each object type and how it is associated with the objects.
- iv. Providing access to the content.

Besides this, some other strategies are also adopted by archivists or contents creators which may be highlighted as a refreshing, migration, replication and emulation etc. Digital preservation mainly faces the problem of digital obsolescence because of lack of established standard, protocols and methods for preserving digital information.

4.6. Access:

Aforementioned all steps are performed for the purpose of ensuring continuous access to archived material. The changes in access mechanisms usually occur with time and Digital Right Management and security requirement over the long term must be considered in successful practice.

i. Access Mechanisms:

Access and presentation of archived items are the dependent on advances in digitization and browser technologies. The changes and advancement in both digitization and browser technologies rapidly occur and it may be possible in the future to enhance the quality of presentation of items from the digital archive.

ii. Digital rights management and security requirements:

DRM systems are designed to enable access and use of digital material and to restrict copying, sharing, reformatting or otherwise changing electronic media. DRM deals issues related to access and use of digital material for digital archiving like what rights does the archive have? What rights do various user groups have? What rights has the owner retained? How will the access mechanism interact with the archive's metadata to ensure that these rights are managed properly? Rights management includes providing or restricting access as appropriate and changing the access rights as the material's copyright and security level changes. Security and version control also impact digital archiving.

5. CHALLENGES IN CREATION AND MAINTENANCE OF DIGITAL ARCHIVE IN DEVELOPING COUNTRIES:

Several benefits and opportunities may be taken by developing countries by creating the digital archives but developing countries have several problems and challenges, in the way of creating of a digital archive. This section of paper discusses various issues and challenges faced by developing countries in creation and maintenance of digital archive based on selected available limited literature (free accessible via internet search engine):

i. Constantly changing software and hardware: Various activities are performed in the maintenance of digital archive that involve digital technologies that enable information to be created, manipulated, disseminated, located and stored with increasing ease, preserving access to this information. All these tasks are performed by certain required combination of hardware and software. Quickly changing digital technologies are one of the most serious challenges associated with digital preservation (Perry 2014). One of greatest issues is faced by developing countries are the longevity of digital collection (Asogwa & Ezema 2012). Digital information is stored in formats which are dependent upon particular software to interpret. There are many more formats than media. As delivery mechanisms and formats change, they increase their complexity (Patel, 2005). The digital material stored on old storage media could be lost because the hardware or software to read them may become obsolete (Arora 2009) and continuously changes in software and hardware technology creates a headache for archivists to make sure that digital content or information is preserved for a long time (Besser 1999).

ii. **Technical expertise:** Successful creation of digital archive depends on the availability of technical expertise. It has usually been seen that less developed countries have inadequate technical expertise. There is still a shortage of IT skilled personnel/human capital in developing countries. Very few archivists hold computer science qualification, who may deal with the digital archive. Asogwa (2011) argues that human resources with appropriate skills, competences and attitude are not readily available to initiate, implement and sustain digital archive project in developing countries. Moghaddam (2011) considered that digital preservation requires highly skilled staff and the ability to employ and develop staff with appropriate skill are made more difficult by the speed of technological change and range of skill needed. Ngulube (2004) pointed out that most trainers, especially in Africa, lack expertise and are ill equipped to train other in the art of electronic preservation and digitization. All these challenges are faced by developing countries and impede the process of creating of a digital archive.

iii. **Funding for digital archive project:** Creation of digital archive is expensive. It requires an enormous finding for performing various tasks in order to create or maintain a digital archive. There are several issues in a digital archive that require high investment and cost such as the development of IT infrastructure for digital archive. It requires high investment and without the appointment of technical experts and skilled employees, creation and maintenance of digital archive is not possible. It also requires a fund to appoint or hire them. Frequent hardware and software up gradation are also required high cost. Training of existing staff also needed to make them compatible to deal with a digital archive or new digital technologies, for this also, high amount of fund is required. Asogwa (2011) pointed out that apart from the inadequate fund to train archivists in Africa, training of archivists in digitization and preservation of electronic format creates a herculean problem. Moghaddam (2010) also argues that cost of digital preservation seem to be much greater than traditional preservation and needs expert staff and considerable expenditure on technological needs.

iv. **Inadequate technology infrastructure:** Almost all developing countries have inadequate technological infrastructure that creates a herculean problem for the creation of the digital archive. Frequent power outage constitute serious bottleneck to the digital archive. It can cause damage to the digital equipments. Most of the countries in Africa (developing countries) don't have the adequate and reliable supply of electricity, such situation consequently make it impossible to maintain a favorable and suitable technological environment for digitization projects or digital archive (Zulu 1994).

v. **Technological Obsolescence:** Technology is evolving at very high speed and it changes very quickly. Frequent changes occur in computer software and hardware and storage media. It also creates technologies obsolescence which is a hazard to digital archive in developing countries. Technological obsolescence is considered one of the greatest threats to successful digital preservation (Moghaddam 2010 and RLG). This technological obsolescence is caused by the continuous up gradation of operating system, programming language application, and storage device. A survey of 54 institutions found that the technology obsolescence as the greatest threat to sustaining the continuous access to digital resources (Hedstrom & Montgomery 1998). The greatest example of technological obsolescence is 5.25 Inch floppy disks have been superseded by 3.5 Inch

and now floppies have become obsolete with the arrival of filching memories/data storage. The obsolescence applies to software as well. For instance, several upgraded versions of window operating system have come since the time, it was introduced. Converting software from earlier version to the current version is not always easy and it may be a caused by loss of access to digital resource or information in the digital archive. Lin, et al. (2003) reported that information technologies become obsolete within 18 months. This is true because very often records created and maintained by one generation of software and hardware cannot be accessed by later generations. Digital archiving requires relatively frequent investments to overcome rapid obsolescence introduced by galloping technological change (Feeney 1999). Alegbeleye (2009) suggested that digital archives should be transcribed every ten to twenty years to ensure that they will not become technologically obsolete.

vi. **Refreshing:** This is a way of avoiding the obsolescence or degradation of the storage medium by periodically moving a file from one physical medium to another. In another word, we can say that it is a way of transforming digital files periodically to new physical storage media. The main intention is to refresh the materials and protect materials from physical decay and obsolescence of the medium. Loss of format is a troubling issue in refreshing because as the information is transferred from program to program, information is lost when an analog material is digitized and there is a possibility of losing the information when digital resources are refreshed or migrated to modern computing environments.

vii. **Continuous migration:** The constant changing technological environment requires continuous migration process to keep on preserving the intellectual contents of digital objects. It also maintains the ability of users to continuously use them in the face of constant changing technology is also a challenging task for archivists in developing countries. Migration is seen as a means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. Lin et al (2003) report that migration of data to the latest media and software version can be done on a two to three years cycle but will require a significant monetary investment for each conversion, constant human attention and personnel training which may be extra burden for institute. The migration process arises several problems in front of archivists in developing countries and requires the fund to solve all these problems.

viii. **The scramble problem:** Data is often compressed or "scrambled" to assist in its storage and protect its intellectual content in a digital archive. These compression and encryption algorithms are often developed by private organizations who may one day cease to support them. In such case, data is stuck between a rock and a hard place. If you don't want to get in legal trouble you are no longer able to read your data (Dharini 2009).

ix. **Lack of legislation/policy:** There is the lack of legislation/policy in developing countries regard to digitization and digital preservation. Even legislators in developing countries are neither aware of, nor familiar with the requirement of digital preservation therefore; they either ignore or inadequately cover digital preservation issues. Internet link has also been a challenge to digitization owing to copyrights legislation. The right to copy for preservation was not adequately articulated in most national legislation in developing countries, and if permission can't be obtained for digitization, digitization of

such materials should not proceed. The intellectual property rights issues in digital materials are more complex and important than the print media and if it is not addressed, it can obstruct or prevent preservation activities (Uluocho 2014).

x. **Deterioration of digital media:** Deterioration of digital media is also a challenge for developing countries. Moghaddam (2010) pointed out that the media on which digital materials are stored is inherently unstable and without suitable storage conditions and management, it can deteriorate very quickly, even though they may not appear to be damaged externally. Technology is changed very quickly if the media is preserved in stable condition then also, it may not be possible to access information, it contains yet because new technology makes present technology obsolesces. The main problem with digital media as it deteriorates or perishes within few years after digitization. Another challenge in front of archivists (in developing countries) is the absence of inadequate organizational plans to manage digital-records. This can also be a cause of losing digital media during disaster or virus attacks. All these harsh situations increase the chance of degradation of electronic equipment and demands for re-digitization.

6. CONCLUSION:

The digital archive and preservation is advantageous in comparison to traditional archive. For successful creation and maintenance of digital archive and preservation various basic principles should be followed. There are several technological, legal and policies related challenges which have to solve for successful implementation of digital archive and preservation. The developing countries have started taking interest in digital archive and preservation but facing several challenges on the way. Lack of funds is also one of the biggest challenges in front of developing countries for creating digital archive and preservation.

REFERENCES:

- Alegbeleye, GO 2009. 'Avoiding Technological Quicksand: Coming to Grips with the Preservation of Digital Information in Nigeria'. Paper Presented at the 47th National Conference and Annual General Meeting of the Nigerian Library
- Arora, J 2009. 'Digital Preservation: An Overview. Open Access to Textual and Multimedia Content: Bridging the Digital Divide, January 29-30, 2009, pp. 106–148.
- Asogwa, BE 2011. 'Digitization of Archival Collections in Africa for Scholarly Communication: Issues, Strategies, and Challenges'. *Library Philosophy & Practice*, pp. 1–14.
- Asogwa, EB and Ezema, JI 2012. 'The challenges of preservation of archives and records in the electronic age', *PNLA Quarterly*, vol 76, no 3
- Besser, H 1999. 'Implications in digitizing Special Collections Materials: the institution, scholarship, interoperability, longebility'. Available from: : <http://sunsite.berkeley.edu/imaging/Databases/Conservation/RBMS99/ppframe.html> [10 February 2016]
- Brown, A 2008. 'Digital preservation Guidance Note 2: Selection Storage Media for Long –Term Preservation', *The National Archive*. Available from:

- <https://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf> [05 January 2015]
- Dharini, A 2009. 'Digital Preservation and Persistence of Digital Collection'. *7th International CALIBER-2009 Pondicherry University, Pondicherry, February 25-27, 2009 @ INFLIBNET Centre, Ahmedabad*, pp. 77–86.
 - Feeney, M (ed). 'Digital Culture: Maximising the Nation's Investment. London, the National Preservation Office, 1999. pp.11.
 - Flecker, D 2003. 'Digital Archiving: What is involved?' Available from <https://net.educause.edu/ir/library/pdf/erm0316.pdf> [: 20 January 2016 from]
 - Hedstrom, M and Montgomery, S 1998, 'Digital preservation needs and requirements in RLG member institutions', *Research Library Group, Mountain View, CA*, working paper, December.
 - Kumaravel, JPS 2004. 'Is Digital Archiving Simple ?' *2nd Convention PLANNER. Manipur University, Imphal 4-5 November, 2004*, pp. 37–43.
 - Lin, LS, Ramaiah, CK, & Wal, PK 2003. 'Problems in the preservation of electronic records'. *Library Review*, vol. 52, no. 3, pp. 117–125.
 - Maji, DC & Das, P 2009. 'Digital Copyright Management : Need for Managerial Aspects'. *7th International CALIBER-2009, Pondicherry University, Puducherry, February 25-27, 2009 @ INFLIBNET Centre, Ahmedabad*, pp. 449–454.
 - Moghaddam, GG 2010. 'Preserving digital resources: issues and concerns from a view of librarians'. *Collection Building*, vol. 29 no. 2, pp. 65–69.
 - National Museum Australia. 2012. 'Digital preservation and digitization Policy. Original approved by Council April 2009; this version approved by Executive 21 May 2012'. Available from: http://www.nma.gov.au/_data/assets/pdf_file/0013/1453/POL-C-028_Digital_preservation_and_digitisation-2.2_public.pdf [27 December 2015]
 - Ngulube, P 2004. 'Fostering accountability: opportunities and challenges faced by records managers in changing societies. Paper read at Access Information Management Services' Annual Records Management Conference in Kruger National Park, South Africa, 18-21 May 2004.
 - Niven, K 2014. 'What is Digital Archiving? Archaeology Data Service / Digital Antiquity (2011) Guides to Good Practice'. Available from: http://guides.archaeologydataservice.ac.uk/g2gp/ArchivalStrat_1-0 [12 January 2016]
 - OCLC. 2006. 'OCLC Digital Archive Preservation Policy and Supporting Documentation'. *Online Computer Library Centre, Inc Dublin, Ohio, USA, 2006*. Available from <http://wiki.lib.sun.ac.za/images/e/e5/Oclc-digital-preservation-policy.pdf> [06 January 2016]
 - Patel, DR 2005. 'Preserving the information content: issues and strategies'. *PLANNER*, pp. 218–227.
 - Perry, SR 2014. 'Digitization and Digital Preservation : A Review of the Literature'. *iSchool Student Research Journal*, vol. 4, no. 1, pp. 1–12. Available from: <http://scholarworks.sjsu.edu/slissrj/vol4/iss1/4>
 - Sureshbabu, R 2005. 'Metadata at Digital-Informative Era'. *3rd Convention PLANNER-2005, Assam University, Silchar, 10-11 Nov., 2005 @ Ahmedabad*, pp. 159–165.