CONSORTIUM

Corresponding author:
Dr. Haroon Idrees PhD, Associate Professor Department of Library and Information Sciences, University of Sargodha, Pakistan. E.:
h.haroonidrees@gmail.com

Review Articles

DIGITAL PRESERVATION: ISSUES AND POSSIBLE SOLUTIONS

Haroon Idrees[1], Rafiq Ahmad[2]

1. Associate Professor, Department of Library and Information Sciences, University of Sargodha, Pakistan.
2. Librarian, Government Degree College, Bannu, Pakistan

Abstract

Digital objects are more vulnerable to destruction or loss as compared to conventional library materials. Issues like format obsolescence, software obsolescence, hardware obsolescence and storage media fault/deterioration are some of the major threats that make the digital resources vulnerable to destruction. Digital resources require special attention and application of strategies for long term retention and accessibility. The purpose of this paper is to outline the basic issues that threaten the retention and accessibility of digital resources in the long term. The paper also attempts to provide an outlook of the possible solutions for overcoming these issues.

Key words: Digital preservation, prospective archives, literature.

1. Introduction

The invention of computer has resulted in the second information revolution. Computer technologies have changed the way people seek, acquire and use information. Most of the information available today is in digital form due to the reason of ease of creation. The world is becoming digital at an astonishing pace and the public is embracing this digital environment. According to Rafiq (2011) "the world is experiencing a paradigm shift from analogue systems to digital ones".Oltmans and Wijngaarden (2004) state that these days results of the scientific research are published in digital form as a rule.

Besides several advantages of digital resources, in comparison to analogue library materials, the preservation of digital resources is a serious problem. The continuous growth and development of technology like changes in software systems, operating systems and hardware systems poses serious problems to the preservation ofavailable digital resources. Huge amount of records have been produced in the last decade than any of the previous decade but most of it less reliable, retrievable or accessible(Duranti, 2010). According to Rothenberg (1998) "the rapid development in different computer technologies ensures that digital obsolescence will become a problem in future". Digital obsolescence is a situation where a digital resource is no longer readable because the physical media, required reader, hardware, or software that runs on it is no longer available. Research Libraries Group through its survey found technological obsolescence as the greatest threat to successful digital preservation (Hedstrom & Montgomery, 1998).Some authors argue that "digital files are constantly under threat – against the forces of time, nature, technology and human error, just to name a few culprits" (Dolan-Mescal, Farwell, Howard, Rozler, & Smith, 2014).The first decade of the 21st century was called as the decade of digitization and the second decade of the 21st century is termed as the era of digital preservation (Trehub, 2011; Trehub & Wilson, 2010). According to Cordeiro(2004)the ever growing literature of digital preservation from the last decade predicts the challenges of durability, accessibility, maintenance, functionality

CONSORTIUM

Additional information is available at the end of the article.

and understandability along with long term retention for digital resources.Bennett (1997) focuses on the risks of digital uncertainty with regards to storage media & unrestrained technological obsolescence imitated in changes in OS, file formats, hardware, and software have been recognized as serious threats to the future of ever growing digital assets.

2.   Digital Preservation

The purpose of preservation is to protect information of enduring value for access by present and future generations (Conway, 1990). According to (Hedstron, 1998) digital preservation is "the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible/usable."

3.   Issues in Digital Preservation

Digital objects are threatened in a number of ways. The continuous changes/developments in formats, software systems, operating systems, and hardware systems are some of the technological threats to digital objects. Besides this storage media fault/failure is also a serious threat to the long term retention of digital objects. Disaster, accidents and human error are also amongst the key issues in the preservation of digital resources.The basic issues in digital preservation are discussed in details in this section.

3.1. Technological obsolescence

Technology is in evolution since its inception. Changes/developments in software systems, formats, operating systems and hardware systems ensure that digital obsolescence will take place where a digital object will become inaccessible due to non-availability of software systems, hardware system and/or readers that are necessary for accessing a specific digital object.

3.2. Storage media fault/failure

Digital storage media are extremely vulnerable to loss/destruction. The fault/failure of digital storage media results in loss of huge amount of digital information. Though the storage media has kept its pace with advancement of technology in terms becoming smaller in size and bigger in storage capacity but the issues of its vulnerability is still unanswered.

3.3 Disasters, accidents and human errors

Disaster and accidents are unpredicted and cause huge loss to human civilization. Digital resources are more vulnerable to disasters and accidents as compared to other assets. Natural disasters like floods, earthquakes and tornados etc. and man-made disasters like fires and wars are serious threats to the retention of digital resources. These disasters and accidents destroy the whole of digital collections and lefts nothing for the future generations.

3.4. Electronic Journal/Databases

In the current scenario libraries subscribe electronic journals/databases for providing access to scholarly communications to their users. Huge amount of funds are spent on subscription of these electronic journals/databases, but these e-journals are usually at risk as the termination of subscription (due to any reason) stops access to the institution. The back issues for which the institution has already paid huge funds and were accessible

previously becomes inaccessible.

4. Digital Preservation Strategies or Possible Solutions

To safeguard the digital resources for future generations, some strategies like migration, emulation, media refreshing, replication, encapsulation and digital archaeology are needed to be adopted. These strategies provide solutions to certain level for keeping the digital resources accessible over time.

4.1. Migration

According to (Garrett et al., 1996), "Migration is the transferring of data to newer system environments". Migration refers to converting the file format to a newer one. According to McCarthy(2008) "Migration is perhaps viewed in some quarters as the more favorable option, with the acceptance that some loss of metadata or context will be inevitable". Hedstrom (1998) states that "better methods for migration of digital materials to new generations of software are much needed for digital preservation regardless of breakthroughs in mass storage technologies."

Migration is a process that involves periodical transfer of digital files from one hardware and software configuration to another or from one generation of technology to modern generationof computer technology for the purpose of preservation/accessibility of digital information for present and future use. Migration also involves change in the configuration of the underlying data, not its intellectual content. This process requires continuous investment for keeping pace with developments in technology to maintain accessto digital information. Besides its several advantages migration also poses some problems for digital objects like authenticity, further it also presumed to be time consuming and expensive method.

4.2. Replication

Replication refers to 'possessing two or more facsimiles of digital objects for the purpose of preservation'. Replication offers a backup support to digital objects from threats like disaster and accidents like malicious attacks, and storage media fault/failure etc.Digital files that exist as a single copy in only one location is presumed to be highly vulnerable to destruction/loss due to unpredicted storage media fault/failure and natural disasters like earthquakes, fires and floods etc. Digital objects are expected to be more secure if they are replicated and kept at different geographical locations. Walters, Bishoff& Gore (2009) suggest "a network of geographically and institutionally distributed digital repositories that are adherent to best practices, to avoid loss of digital resources due to insufficient infrastructure of a single institution that is geographically homogeneous."

4.3. Technology preservation

This proposes that digital data should be preserved on a stable medium with associated copies of the original application software, operating system and relevant hardware devices. This approach actually suggests that the material as well as the technology need to be preserved. This approach doesn't seem to be practical as it is financially unfeasible.

4.4. Emulation

It refers to the preservation of data in original format and offers to create emulator programs which mimic the obsolete hardware and software. It targets the technical atmosphere in which the digital object was created.In other word this is the process of re-creation of hardware and software environment necessary to access a digital resource. This concept mainly emphasizes on software related to data.In simple words emulation does not alter the digital object itself, it only creates an access version for the file through impersonation by using the existing technology. Emulation is being advocated to be a solution for digital preservation(Rothenberg, 1998; J. Rothenberg, 2002). According to Pozo, Long, & Pearson (2010) emulation can be a proper choice as compared to migration.

4.5. Media Refreshing

According to (Zierau, 2012) "the media is the basis for accessing the original stored information, therefore, the media is essential for all preservation actions".Refreshing means the transferring of data from old physical storage media to newer one on periodic basis which provides solution for avoiding the obsolescence of storage media.Sometimes new technology is unable to access the older storage medium so, refreshing is necessary to achieve longevity. Media refreshing also refers to the transfer of data between two types of the same storage medium to avoid deterioration, e.g. transferring the census report from older CDs to newer one. Refreshing provides an ideal solution for avoiding media deterioration.

4.6. Digital Archaeology

Data archaeology is the process of recovery of data later on, using up-graded techniques available in future.This is essential for that digital information which has not been migrated but still vital in respective field.In this process data would be refreshed regularly but no migration would be performed and no software and hardware would be preserved to be emulated at a later storage.

4.7. Encapsulation

It means that objects should be self-describing, in term of requirements for accessibility of specific digital file. The informationsupplemented with the digital filewill contain the details of how to understand that object by using logical structures called "containers" or "wrappers". This method is usually usefulforcollections that goidle for long time.Encapsulation is technically considered as the most suitableoption to cover the requirements of government organizations for preservation of their digital resources(Nelson, Argue, Efron, Denn, & Pattuelli, 2001).

4.8. Metadata attachment

Metadata means data about data. Metadata of digital file contains information about the creation, access rights, restrictions, preservation history, and rights management. Metadata attached to digital files may be affected by file format obsolescence. ASCII is thought to be the best durable format for metadata as it is famous. However, SGML or XML can be used for advanced functionality.

4.9. Archiving of Electronic Journals

The organizations that have subscribed to electronic journals can secure the long term accessibility and retention of these journals through archiving. This will help them in

securing access to the back issues of these journals in case of termination of subscription due to any reason.

5. Conclusion

As discussed in the beginning, digital resources have several advantages over the conventional sources of information in terms of ease of access and creation, but in term of long term retention and long term accessibility the digital resources have remained a problem due to the reasons like: storage media deterioration/fault/failure, hardware obsolescence, software obsolescence, and accidents/disasters. To secure the digital resources from loss and inaccessibility, they require special care and attention. Digital preservation provides certain strategies like migration, emulation,encapsulation, media refreshing and replication that can secure the digital resources from loss are becoming inaccessible.

6. Recommendations

After review of literature, the following recommendations have been drawn for preservation of digital resources:

1. Digital resources should be migrated to newer version of file format from time to time.
2. Use of open and widely available formats should be preferred. This will help in getting independence from proprietary software and will bring uniformity.
3. The digital resources should be supplemented with information about the software and hardware that is required for accessibility.
4. The storage media should be used that is robust in nature.
5. Multiple backup copies should be kept to avoid loss of digital resources due to storage media fault/failure.
6. The backups should be kept at geographically distinct locations.
7. Digital resources should be archived/stored on a cloud service like googledrive, dropbox, and skydrive etc.
8. The system should be secured from malicious attacks with the help of a suitable antivirus software.
9. The organizations that have subscribed to e-journals should archive these journals to avoid stoppage of access to the back issues in case of cancellation of subscription.

References

Bennett, J. (1997). A Framework of Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material: British Library Reserch and Innovation Centre.

Conway, P. (1990). Archival Preservation in a Nationwide Context. *American Archivist, 53*(2), 204-222.

Cordeiro, M. I. (2004). From rescue to long-term maintenance: preservation as a core function in the management of digital assets. *VINE: The Journal of Information and Knowledge Management Systems, 34*(1), 6-16.

Dolan-Mescal, A., Farwell, M., Howard, S., Rozler, J., & Smith, M. (2014). A digital file inventory of the Queens College Special Collections and Archives: Findings, analysis and

CONSORTIUM

recommendations for long-term preservation. *OCLC Systems & Services: International Digital Library Perspectives, 30*(2), 78-90. doi: 10.1108/OCLC-07-2013-0023

Duranti, L. (2010). Concepts and principles for the management of electronic records, or records management theory is archival diplomatics. *Records Management Journal, 20*(1), 78-95.

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

Hedstrom, M., & Montgomery, S. (1998). Digital Preservation Needs and Requirements in RLG Member Institutions. California: Research Libraries Group.

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

McCarthy, F. (2008). Digital preservation – making the song remain the same. *Records Management Journal, 18*(2), 107-114.

Nelson, M. L., Argue, B., Efron, M., Denn, S., & Pattuelli, M. C. (2001). A Survey of Complex Object Technologies for Digital Libraries (pp. 73). Hampton, Virginia National Aeronautics and Space Administration (NASA)

Langley Research Center.

Oltmans, E., & Wijngaarden, H. v. (2004). Digital preservation in practice: the e-Depot at the Koninklijke Bibliotheek. *VINE: The Journal of Information and Knowledge Management Systems, 34*(1), 21-26.

Pozo, N. d., Long, A. S., & Pearson, D. (2010). "Land of the lost": a discussion of what can be preserved through digital preservation. *Library Hi Tech, 28*(2), 290-300. doi: 10.1108/07378831011047686

Rafiq, M. (2011). *Prospects Of Digitization In University Libraries Of Pakistan.* (Phd Thesis Unpublished), University of the Punjab, Lahore.

Rothenberg. (1998). *Avoiding technological quicksand: finding a viable technical foundation for digital preservation*. CLIR. Washington DC.

Rothenberg, J. (2002). Preservation of the times. *Information Management Journal, 36*(2), 38-43.

Trehub, A. (2011). Applications of Digital Technology to Slavic Librarianship. *SOLANUS: INTERNATIONAL JOURNAL FOR THE STUDY OF THE PRINTED AND WRITTEN WORD IN RUSSIA AND EAST-CENTRAL EUROPE, 22*, 158-170.

Trehub, A., & Wilson, T. C. (2010). Keeping it simple: the Alabama Digital Preservation Network. *Library Hi-Tech, 28*(2), 245-258.

Walters, T. O., Bishoff, L., & Gore, E. B. (2009). *Distributed Digital Preservation: Technical, Sustainability, and Organizational Developments*. Paper presented at the iPRESS 2009: the Sixth International Conference on Preservation of Digital Objects.

Zierau, E. (2012). A holistic approach to bit preservation. *Library Hi Tech, 30*(3), 472-489.

CONSORTIUM

doi: 10.1108/07378831211266618