DIGITAL PRESERVATION

Babita Garg ***Salek Chand***

Abstract

Digitization technology offers the facilities to preserve documents and make them available in an easily accessible system. Digital technology and high-speed networks are leading to sweeping changes. In the modern electronic era more & more information is being produced in digital format by two means:- a) Converting the existing materials to digital format and b) Increasingly born in digital format. This paper presents the challenges of digital preservation to browser, preservation policies and Infrastructure requirements for digital preservation, current preservation strategies, limitations, present preservation strategies, advantages & limitation of digital preservation.

Keywords : Digitization; Preservation; Digital Technologies; Digital Techniques

1. Introduction

Preservation generally means keeping an object safe from harmful effects such as loss, damage, destruction and the like. In reaction of these challenges, there have been many advances and changes within the field of preservation. Digital technology and high speed networks are leading to sweeping changes throughout the society. In the past few years significant progress has been made to define the terms and outline a research agenda for preserving digital information that was either originally or transformed to digital form. Traditional sources of information is increasing every day and the users expectations are to save the valuable information at safer place. To fulfill the user's requirements for posterity, digital documents and resources should be managed and preserved at a tremendous rate. Digital technologies present a preservation solution for the documents in the libraries with increased access to digitized documents over the electronic networks. Digital technology as well as all others associated with Internet and web technologies is a continuous flux of change.

2. Digital Preservation

Digital Preservation refers to the various methods of keeping digital material alive into the future, typically centers on the choice of interim storage media, the life expectancy of a digital imaging system and the life expectation to migrate the digital files to future systems while maintaining both the future systems, full functionally and the integrity of the original digital system.

It includes everything from electronic publication on CD-ROM to online databases and the collection of experimental databases in digital format maintains the ability to display, retrieve and use digital collections in the face of rapidly changing technological and organizational infrastructures and elements.

2.1 Objectives of Digital Preservation

1. To reduce the effect of deteriorating factors such as temperature, light, humidity, flood, fire, fungus, bacteria, insects, pollution, dust and the important factors called human.
2. Illustrate the combination of developments, events and decisions that led us to where we are today, in regard to technology that pertains to digital preservation.
3. Keep a place for new and emerging technology in the context of digital preservation programs.
4. To maintain the historical value of information.
5. To make it easy to use & handle
6. To make information survive longer.
7. To provide world wide accessibility.

2.2 Infrastructure Required for Digital Preservation

The computer system having the multimedia facility integrates all the media components to a single platform and provides interactivity to the system.

The following IT components are suggested for the preservation of the collection.

1. Central processing unit with High-speed processor.
2. Large memory (RAM).
3. Hard disk with high volume.
4. Floppy drive.
5. CD drive (Read/Write)
6. Modem
7. Sound blaster card.
8. Key board having multimedia keys & mouse scroll
9. Color monitor
10. Color/B&W laser printer.
11. Better digitization tools (Scanner)
12. Digital Camera (still/move)
13. Video camera for capturing analog video and convert into digital
14. Graphic cards

3. Digital Preservation Techniques

Digitization requires certain technologies. These include storage technologies having a variety of devices to store & retrieve information in digital form such as magnetic tapes/cassettes, floppy discs, hard discs, DAT tape, CD-ROM smart cards. Processing technology for creating the system and application software that require for the performance of digital networks, communication technologies primarily to communicate information in digital form display technologies and various output devices.

4. Challenges of Digital Preservation

Digital preservation has raised many challenges of a fundamentally different nature, which are divergent as compared to the problems of preserving traditional format materials. Some of them are in the areas of planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable. The challenges are as follows:

4.1 Durability: The durability of digital content has become problematic for a number of complex and interrelated reasons as most of the documents exist only in encoded form, specific software is required for handling them. The digital components like hardware and software are also often changing their versions or processing capacity. Other aspects of this problem are non-technical issues such as management, funding, staffing, maintenance for digital documents and updating the development of policies for standard techniques and practices to prevent the loss of digital information, need to be well planned.

4.2 Form of material: Digital materials have a very short life as compared to the traditional format. In this situation magnetic and optical media are qualitatively different as these are re-usable media.

4.3 Media Problem: Due to advancement in the computer hardware, storage, and software industries, media obsolescence is a very common fact. When greater storage and processing capacities are available in the market at lower cost, slowly the market for the old product market goes down. The documents created in digital form are equally vulnerable to technological obsolescence. For example, the short-life time media are eight-inch floppy disks, tape cartridges and reels, hard-sectored disks and seven track tapes. Those storage formats are inaccessible and more durable storage media are CD-ROM/DVD-ROM and optical WORM.

4.4 Software-dependent problem: The digital documents are generally dependent on application software to make them accessible and meaningful. But the problem is that software is also developing and changing versions. Every software has different kinds of encoding for which every computer needs some specific software to activate the digital documents. A bit stream can be made intelligible only by running the software that created it, or some closely related software that understands it.

4.5 Standards: The absence of established standards, protocols, and established methods for preserving digital information creates problems. The main objectives of digital libraries are to organize the information; maintain the intellectual property rights, and presentation, retrieval and visualization of digital materials. Digital preservation remains largely experimental and replete with the risks associated with unproven methods. Moreover, digital preservation requirements have not been factored into the architecture, resource allocation, or planning for digital libraries.

4.6 Intellectual Property Rights: An intellectual property rights (IPR) is a big barrier for preserving the digital documents. Digitization of documents involves complex methods for resolving the legal and practical questions of migrating intellectual property that includes the creators and owners of intellectual property, managers of digital archives, and potential users of intellectual property.

5. Current Preservation strategy& its limitations

Current methods for preserving digital materials do not fully support achieving their objectives when faced with the responsibility for preserving digital materials. Archives and libraries facing a lot of complex and difficulties based on the format of the original material, the anticipated uses for it, and the technical and financial resources available to invest in preservation initiatives. Some current preservation strategies are need to be reviewed from beginning with the most elementary and established methods and ending with proposals that have not yet been tested.

Probably the most commonly used preservation on strategy is to transfer digital information from less stable magnetic & optical media by printing page image on paper or microfilm. It seems ironic that just as libraries and archives are discovering digital conversion as a cost effective preservation method for certain deteriorating materials, much information that begins its life in electronic forms is printed on paper or microfilm for safe secure long term storage. Yet, high quality acid neutral paper can last a century or longer while archival quality microfilm is projected to last 300 years or more.

Another strategy for digital preservation is to preserve digital information in the simplest possible digital formats in order to minimize the requirements for sophisticated retrieval. Software digital information transferred across. Successful generations of technology are a software independent format as ASCII text files or as flat files with simple uniform structure. Several data archives hold large collections were captured on punch card in the 1950's or 1960's migrated to two or three different magnetic tape formats and now reside on optical media. As new media & storage formats were introduced, the data without any significant change in their logical structure. This approach has the distinct advantage of being universal and easy to implement. It is a cost effective strategy for preserving digital information in those cases where retaining the content is paramount, but display indexing and computational characteristics are not critical. As long as the preservation community lacks more robust and cost effective migration strategies film and preserving flat files will remain the methods of last resort for many institutions and for certain formats of digital information.

6. Mass storage and long term preservation

Current methods fall for short what is required to preserve digital material. All current preservation methods involve trade offs between what is desirable from the stand point of functionally dependability and cost what is possible and affordable and cost what is possible and affordable with current technologies and methods. Consequently, most repositories and copying by employing interim & less than desirable strategies if they are addressing digital preservation issues for example, the simplicity and universality of printing of paper or Microfilm comes at the expense of great losses in the functionality of digital information. Migration strategies that involve reformatting thus lost limiting future analytical potential. Normalization to standard format is not always technically feasible and is usually quite costly.

The present community is only beginning to explore the possible alternatives of storing digital information in software independent from Rothenberg (1995) proposed an approach for maintaining the content of digital materials intact without losing the ability to retrieve meaning rich resources. He recommended retaining the original document in its original format encapsulated in virtual envelope that contains software instructions for retrieval display and processing of the message in the envelope.

The envelopes would contain contextual information and the transformational history of each subject execution of the instruction would rely on an archive of hardware and software emulators or on instructions in the envelope with specification to construct emulators.

Important research is underway to define standards for data interchange that can support electronic commerce and satisfy business requirements in a variety of environments. A major research project at the university of Pittsburgh is defining metadata requirements for evidence that will support the need for integrity, authenticity, reliability and archiving through standards for meta data encapsulated objects archives libraries and other institutions with preservation responsibilities will benefit from its system are built to implement. Such metadata standards wide scale adoption of data and communication standard by the originator of digital information to support current business needs will also facilitate long term preservation.

It is fair to say that the state of development of digital preservation remains largely experimental. Only a few Libraries archives and other institution have established digital preservation programs, while most research and innovation comes from pilot projects and prototypes. Tested methods that have proven effective on small scale in a limited number of repositories are not feasible for preservation of many of the types of digital materials that archives and Libraries their preservation endeavors.

7. General approach to digital preservation

In the mid 1990s the library community began to worry about the fragility of works stored in digital form. The commission on preservation and access and the research libraries groups formed a task force to explore this problem. The task force report sounded an alarm. Rapid changes in the means of recording information, in the formats for storage, and in the technologies for user threaten to sender the life of information in the digital age, to borrow a phrase from Hobbes nasty British and short (task force 1996) As the problem of digital longevity had repercussions with in the arts community as well, the Getty conversation institute and Getty information institute collaborated with leading technologists to put together a conference and book trying to broadly outline and bring attention to the problem (MacLean and David 1998) Both of these seminal works identified the depth of the digital longevity problem but only pointed out very general approaches that might possible lead to solutions. Two key approaches have been proposed to deal with the problem of changing file formats (task force 1996)

7.1 Migration and Emulation

These are seen as alternatives to one another but both approaches are supposed to be used in conjunction and refreshing.

Migration is an approach that involves periodically files from one file encoding format to another that is useable in more modern computing environment (An example would be moving a word star file to word perfect, then to word perfect, then to word 3.0 then to word 5.0 then to word 97) Migration seeks to limit the problem of files encoded in a wide variety of file formats that have existed ever time by gradually bringing all former formats into a limited number of contemporary formats.

Emulation seeks to solve a similar problem that migration addresses, but its approach is to focus on the application software rather than on the files containing information. Emulation backers want to build software. Every application that has ever been written for every type of file format and run on whatever the current computing environment is (sp, with the proper emulators, applications like word star and word 3.0 could effectively run on today's machines) Emulation is most closely associated with the writings of R and scientist Jeff Rotherberg (Rothenberg 1995, 1999, 2000).

Advantages

- Space Saving: Such preservation occupies very limited space as compared to space covered by printed material in library.
- Easy access: All information preserved in electronic form is easily accessible in different ways and with in no time.
- Easy to handle: Information retrieval software are user friendly and easy to use.
- Easy to transmit: Information stored on a CD or in the computer can be transmitted easily from one place to another by email.
- Time saving task: Worldwide information can be retrieved with in a few seconds.
- Low cost: Very low cost is involved in developing an electronic document in comparison to printed document.
- Less manpower involved: To generate electronic information, less manpower is involved.

Disadvantages

- Process of installation is complex and costly: In the initial stage, it is very complicated and long process. It involves preparing images, editing, proofing assigning keywords generating indexes converting into makeup language etc.
- Lack of infrastructure: In developing countries like India, it is very tedious job to provide all infrastructure facilities at one platform as and when it is required.
- Insufficient budget: The Indian labors are not financially strong. They have very limited budget and to procure the recommended system is not affordable.
- Shortage of expertise in India: Lack of expertise in the field of multimedia is the greatest disadvantage.

8. Conclusion

Storage systems should be capable of handling digital information in a wide variety of formats, including text, data, graphics, video & sound conversion from digital formats and migration to new generation of technology will be rapid, accurate and in expensive enough to permit very large scale transfer to heterogeneous materials storage space requirements will be minimal and not demand highly sensitive environmental control. To make digital preservation affordable to the widest possible range of organizations and individuals equipment, media & maintenance cost must be modest with enough coordination between library professionals and co-operation and support of higher or competent authority, the goal of preserving all of the important literature can be accomplished. It is

also suggested that the libraries may consider developing a separate unit where the preservations work can be done regular basis.

References

1. Rothenberg, Jeff. 1995. "Ensuring the Longevity of Digital Documents.", Scientific American 272(1): 24-29
2. Conway, Paul. 1996. Preservation in the Digital World. Washington, D.C.: Commission on 3. Preservation and Access. Available at http://www.clir.org/pubs/reports/conway2/index.html.
4. Dubin, David, Futrelle, Joe and Plutchak, Joel 'Metadata Enrichment for Digital Preservation'; Proceeding of Extreme Markup Language accessed in August 2006 http://www.idealliance.org/papers/extreme/Proceedings/html/2006/Dubin01/EML2006Dubin01.html
5. Adrian Brown, Digital Preservation Guidance Note 3: Care, handling, and storage of removable media, June 2003.
6. John Van Bogart, Magnetic Tape Storage and Handling, A Guide for Libraries and Archives, National Media Laboratory, June 1995.
7. Brian Cooper, Arturo Crespo, and Hector Garcia-Molina. Implementing a reliable digital object archive, 1999. Submitted for publication to ACM DL 2000