Preservation of Digital Cultural Heritage Materials

P Lalitha T A V Murthy

Abstract

Cultural Heritage materials are being immensely converting into digital forms. The digital materials are at risk of being lost and their preservation for the benefit of present and future generations is an urgent issue to address. The present paper gives a view on the issues related to digital preservation, approaches for digital preservation. It also emphasizes the importance of metadata in digital preservation.

Keywords : Digital Preservation, Media obsolescence, Encapsulation, XML, Universal Virtual Computer (UVC), Migration, Emulation, OAIS, Metadata.

0. Introduction

Our cultural, scientific and information heritage is increasingly converting to the digital forms. Libraries, archives, museums and research institutes are potentially responsible for preserving digital heritage. Modern digital technologies have made it a reality to exhibit large collections of works from multiple cultures. Most of the cultural and heritage materials are being converted into the digitized forms knowing that permanent access to this heritage will offer broadened opportunities for creation, communication and sharing of knowledge among all peoples, as well as protection of rights and entitlements and support of accountability. Since an enormous amount of historical and cultural materials have been created, both storage and distributions raise many challenges [14]. Digitized cultural artifacts should be useable in future for many possible applications. The data should be therefore being preserved for the long term retaining as much of the information as possible [7].

1. What is Digital Preservation ?

According to Sue Mckemmish, digital preservation means “Enable reliable, authentic, meaningful and accessible records to be carried forward through time within and beyond organizational boundaries for as long as they are needed for the multiple purposes they serve.”[16].

According to Cornell University Library, Digital Preservation encompasses a broad range of activities designed to extend the usable life of machine-readable computer files and protect them from media failure, physical loss, and obsolescence.[3]

Digital preservation consists of the processed aimed at ensuring the continued accessibility of digital materials. [12]

2. Need for Preservation

The long term preservation of the intellectual and cultural record of society has occupied librarians, archivists and museum curators for centuries.[9] The long term future of digital resources must be assured, in order to protect investments in digital collections and to ensure that the scholarly and cultural records are maintained in both its historical continuity and media diversity.[10] It would seem that preservation would be much easily achieved with digitized resources, but there are a number of issues that complicate the maintenance of digital objects over a long period. To address few of them:

a. Media Decay which may be caused due to physical, chemical and magnetic fields, etc.,

b. Media Obsolescence

   i. Physical and logical format incompatibilities.

   ii. Unavailability of suitable “drives” or “controllers”.

c. Changes in applications and operating systems which may cause unavailability of operating systems, input output devices, etc., for required software and unavailability of hardware required to run required software.

A digital artifact is completely software dependent as its structure and content could only be understood by the programme that has created it. That particular programme only runs that software. Through this is evident that all digital documents or artifacts are software dependent. So in the present context of media obsolescence, the threat is very real and insidious and will eat away the future of our cultural heritage, knowledge economies, and information society. [1]

In 1964, the first electronic mail message sent from either the Massachusetts Institute of Technology, or the Carnegie Institute of Technology of Cambridge University. The message does not survive however, and so there is no documentary record to determine which group sent the path breaking message. Satellite observations of Brazil in 1970s, critical for establishing a time-line of changes in the Amazon basin, are also lost on the now obsolete tapes t which they were written. [15]

3. Approaches for Digital Preservation

The below given are some important preservation strategies through which digital preservation can be achieved.

3.1 Technology Preservation

To preserve the technology required to access original records for as long as those records are required. But support for the software and hardware eventually ceases and the parts required to maintain the hardware become more and more scarce as manufacturers discontinue obsolete components. The number of machines available that are capable of recording old files continues to decrease, for computers do not last for ever. The skills required to operate the hardware and software also become rare and eventually disappear.

3.2 Printing to paper

Although this approach is still in practice, printing all records to paper is not a viable preservative method. Printing to paper loses functional or behavioral traits that the records had in them digital form. Certain information may also be lost.

3.3 Encapsulation

The encapsulation approach retains the record in its original form, but encapsulates it with a set of instructions on how the original should be interpreted. This would be needed to be detailed format descriptions of the file format and what the information means. The process can be well understood through Fig.1.

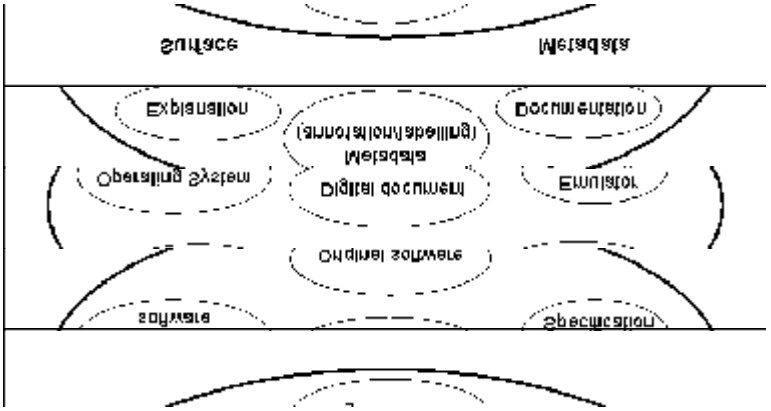

Fig. 1.Encapsulation

3.4 Virtual Machine Software

Raymond Lorie of IBM Almaden has proposed this approach. This addresses the problem of interpreting data files in the future by programming a set if instructions to carry out these interpretations in the machine language of a “Universal Virtual Computer” (UVC). This programme would be written at the time the record was archived and would be preserved together with the record. In order to interpret the record on a future computer, a UVC interpreter would be required and this could be produced fro the specifications of the UVC. With this process the data can be stored in any format and the knowledge required to decode it is encapsulated in the UVC programme.

3.5 XML

Extensible Markup language is a text-based markup language for describing the structure and meaning of data. As it is text based, it is human readable, but desired primarily to be easy to process using the computers. It is a open standard defined by World Wide Web Consortium. Conversions of records to XML format can be seen as a particular type of migration. It is often regarded as a very promising present day data format for archiving and interoperability and so deserves to be considered as an approach in its own right. XML is of the greatest importance for digital preservation, not just because of this widespread uptake, but also because it protects the Achilles’ heel of digital documents: the dependence on obsolete operating systems and application software. It does this by being platform- and software-independent. [5]

3.6 Migration

Migration can be defined as the transfer of files from one hardware configuration or software application to another configuration or application. Problem associated with migration is that the results are often unpredictable, mostly because of a lack of or because the process has not been fully tested. The results of migration are difficult to predict, unless a substantial amount of work is first done regarding the specifications of the source and target formats. Migration can influence the authenticity of a document. Each document that is preserved must be preserved ‘authentically’, otherwise the meaning and validity of the archival record cannot be guaranteed. This has both legal and archivist implications. [6]

3.7 Emulation

The theory behind Emulation is that the only way to ensure the authenticity and integrity or the record over the long term is to continue to provide access to it in its original environment, i.e., its original operating system and software application. This can be only done by preserving not only the record, but also emulator specifications, which contains enough details about the original environment for that environment to be recreated on a future computer whenever necessary. Emulating strategies would involve encapsulating a data object together with the application software and to create or interpret it and a description of the required hardware environment. From the Figure. 2., it is understandable that there are three technical options in emulation.

a. Emulate Applications

b. Emulate Operating Systems

c. Emulate Hardware platforms [8]

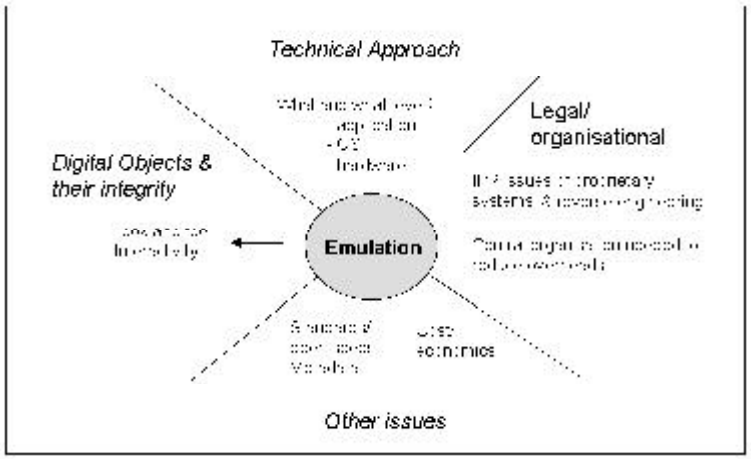

Fig.2. Emulation

It is suggested that these emulator specifications formalisms will require human readable annotations and explanations (metadata). [4] Preservation strategy may be emulation based or migration based, it is that both will have same role that is the long-term preservation of digital information that involves the creation and maintenance of metadata. Within an archive, metadata accompanies and makes reference t each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information. This metadata will also be maintained and will be migrated from format to format and standard to standard, independently of the base object it describes. A Digital object enters a repository as a set of sequence of bits; it is accompanied by a variety of metadata related to that object. With proper storage management, replication and refreshing, this set of sequences of bits can be maintained indefinitely. [11] For example, The Pittsburgh Project, The UBC Project, The SPIRT Record keeping Metadata Project are some research projects and practically-based initiatives have been concerned with the development of record keeping metadata schemes and standards.

4. Open Archival Information System (OAIS)

This model has been developed by CCSDS: Consultative Committee for Space Data Systems (NASA) – ISO: 2002. This Reference Model:

- Provides a frame work for the understanding and increased awareness of archival concepts needed for long term preservation and access;
- Provides the concepts needed by non-archival organizations to be effective participant in the preservation process;
- Provides a framework for describing and comparing different long term preservation strategies and techniques;
- Expands consumers on the elements and processes for long-term digital information, preservation and access, and promotes a larger market which vendors can support.

The reference model defined common terminologies like :

- AIP (Archival Information Package)
- SIP (Submission Information Package)
- DIP (Dissemination Information Package)
- PDI (Preservation Description Information)

It has also discussed many important issues like

- Ingest formats and processing
- Use of standards
- Metadata
- Existing Records (bibliographic record in world Catalog).

As the present paper emphasizes on the digital preservation, Preservation Description part of the OAIS information model has been briefly described.

The OAIS information model divides Preservation Description Information into four categories:

i. Reference Information : It describes identification systems, and the mechanisms for providing assigned identifiers, used to unambiguously identify the Content Information both internally and externally to the archive in which it resides.

ii. Context Information : It documents relationships of the Content Information with its environment, including the reasons for its creation and relationships to other Content Information objects.

iii. Provenance Information : It documents the history of the Content Information, including its origin, changes to the object or its content over time, and its chain of custody.

iv. Fixity Information : It provides the Data Integrity checks or Validation/Verification keys used to ensure that the particular Content Information object has not been altered in an undocumented manner.

In a nutshell, Preservation Description Information records the identity, relationships, history and integrity of the archived Content Data Object. With this it is understandable that Effective Metadata is a necessary condition for effective digital preservation. The elucidation and maintenance of Preservation Description Information, however, is the keystone to building an information infrastructure to support the processes associated with digital preservation. [13]

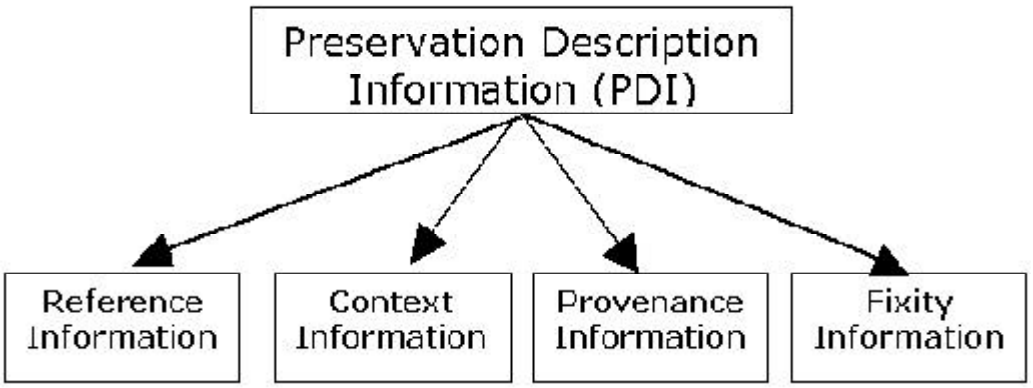

Fig. 3. PDI

In OAIS model the Digital Migration is defined to be the transfer of digital information, while intending to preserve it, within the OAIS. It is distinguished from transfers in general by three attributes:

- a focus on the preservation of the full information content,
- a perspective that the new archival implementations of the information is a replacement of the old; and
- full control and responsibility over all aspects of the transfer resides the OAIS.

But recognized that: the Digital Migrations are time consuming, costly, and expose the OAIS to greatly increased probabilities of information loss. Therefore, an OAIS has a strong incentive to consider Digital Migration issues and approaches. [2]

5. Conclusion

Even though it is difficult to decide which digital materials are to be preserved, there should be a strict plan for preservation according to the organizational agendas. In possible environments, it is always better to use the standard formats for preservation. Although the Preservation Community is at its disposal for providing solutions for digital preservation, till date there are no scalable solutions for the general problem of digital preservation.

6. References

1. Beagrie, Neil., 2004. “The Continuing Access and Digital Preservation Strategy for the UK Joint Information Systems Committee (JISC)”, *D-Lib Magazine*, 10(7/8). http://www.dlib.org/dlib/july04/beagrie/07beagrie.html Accessed on 14.09.04.

2. CCSDS: Consultative Committee for Space Data Systems (NASA), 2002. http://www.ccsds.org/documents.650xob1.pdf Accessed on 20.9.04.

3. Cornell University Library, 2003.Digital Preservation Management: Implementing Short-term strategies for Long-term Problems. http://www.library.cornell.edu/iris/tutorial/dpm/terminology/preservation.html Accessed on 06.03.04

4. Day, Michael., 1999. "Metadata for Digital Preservation: an Update", *Ariadne*. 22. http://www.ariadne.ac.uk/issue22/metadata/intro.html Accessed on 06.03.04.

5. Digital Bewarning, 2002. Digital Preservation Test bed White Paper, XML and Digital Preservation http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white_paper_xml_en.pdf Accessed on 21.10.04.

6. Digital Bewarning, 2003. Digital Preservation Test bed White Paper Emulation: Context and Current Status. http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white_paper_emulatie_EN.pdf Accessed on 21.11.04.

7. European Commission on Preservation and Access, Amsterdam, 1997. Digitization as a Means of Preservation?: Recommendations for the digitization of microfilm, http://www.clir.prg/pubs/reports/digpres/digpres3.html Accessed on 12.10.04

8. Granger, Stewart; 2000. "Emulation as a Digital Preservation Strategy", *D-Lib Magazine,* 6(10). http://www.dlib.org.october00/granger.html Accessed on 16/8/04

9. Hedstrom, Margaret., Digital Preservation: a time bomb for Digital Libraries.

   http://www.uky.edu/~kiernan/DL/hedstrom.html Accessed on 06.03.04.

10. Lavoie, Brian; Dempsey, Lorcan., 2004. "Thirteen Ways of Looking at ...Digital Preservation", *D-Lib Magazine*, 10(7/8). http://www.dlib.org/dlib/july04/lavoie/07lavoie.html Accessed on 14.09.04.

11. Lynch,Clifford., 1999. "Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information". *D-Lib Magazine*, 5(9). http://www.dlib.org/dlib/september99/09lynch.html Accessed on 06.03.04.

12. National Library of Australia, 2003. Guidelines for the Preservation of Digital Heritage. http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicities/unesco_guidelines.pdf. Accessed on 04.09.04.

13. OCLC/RLG Working Group on Preservation Metadata, 2002. Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects. http://www.oclc.org/research/pmwg/ Accessed on 03.04.04.

14. Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials. http://www.delosnsf-imaging.unifi.it/assets/wg-final_report.doc Accessed on 16.09.04

15. Research Libraries Group, Inc., May 1, 1996. Preserving Digital Information: Report of the Task Force on Archiving of Digital Information, Commission on Preservation and Access http://lyra2.rlg.org/ArchTF/tfadi.index.htm Accessed on 01.04.04

16. Rothenberg, Jeff., 2002. Digital Preservation: The State of the Art. http://www.kb.nl/kb/hrd/dd/dd-links_en_publicaties/workshop2002/ rothenberg.pdf Accessed on 10.04.04

About Authors

Ms. P Lalitha is the Librarian at Kesar SAL Medical College & Research Institute. She holds MLISc and pursuing Ph.D. in LIS from University of Bundelkhand , Jhansi. She is also teaching LIS students of IGNOU at Ahmedabad Regional Centre. She has published three papers, and her area of interest is Digitization of Artifacts related to Indian Heritage and Culture.
E-mail : lalithapoluru@yahoo.com.

Dr. T A V Murthy is Director of INFLIBNET Centre, Ahmedabad and holds B Sc, M L I Sc, M S L S (USA) and Ph.D. He is President and Fellow of SIS-India, Hon. Director of E.M.R.C., Gujarat University and Member Secretary of ADINET, Ahmedabad. He carries with him a rich experience and expertise of having worked in managerial level at a number of libraries in many prestigious institutions in India including National Library, IGNCA, IARI, Univ of Hyderabad, ASC, CIEFL etc and Catholic Univ and Case western Reserve Univ in USA. He has been associated with number of universities and has guided number of Ph.Ds and actively associated with the national and international professional associations, expert committees and has published good number of research papers. He visited several countries and organized several national and international conferences and programmes.
E-mail : tav@inflibnet.ac.in