Archiving / Preserving Digital Information

Amritpal Kaur

Abstract

Preservation of digital material has become a major challenge for library and information professionals these days. Although libraries have been concerned with preserving cultural and intellectual heritage for centuries, it is only in recent years that preservation and conservation have become activities in their own right. Preservation of digital material is more complex because of the dependencies that exist between the information object itself and its technical environment- an environment that is generally created using rapidly obsolete and, therefore, fragile technology. This dependence can be at both the software and the hardware levels. In this sense, digital media present more complex problems than any other media libraries have ever had to face. The present article explores various strategies and methodologies for preserving digital materials and also focuses on the pressing need for libraries to develop strategies and practical action plans to maintain the safety and accessibility of the world's historical and cultural heritage.

Keywords : Archiving; Preservation; Digital information; Digital preservation; Digital preservation strategies.

0. Introduction

From ancient times to the present, the physical media that record knowledge has undergone steady revolution with the technological advancements and innovations. Now we are in the digital age and information is recorded, stored retrieved and disseminated in the digital form. Digital media has emerged as a powerful media of information. The libraries as centres of information services have come to rely increasingly on digital information both as supplements to and parallels of print materials. Now libraries are also confronted with the media, which are 'born digital' and have no print equivalent. Digital media, no doubt, have immense capacity to record information enabling the libraries to provide users seamless access to information, but it is very fragile in comparison to print media. Conway rightly remarks, "Today our capacity to record information has increased exponentially overtime while the longevity of the media to store information has decreased equivalently"[1].

Technology changes so rapidly and the lines between the old and the new seem drawn so sharply that Wired magazine can refer to a four-year old network service provider as a "dinosaur" [2]. Magnetic tape may be unreadable just thirty years after its manufacture [3]. It may happen that today's optical storage media may long outlast the life of the computer system that created the information. This is because we depend on machines that rapidly reach obsolescence to create information and then make it readable and intelligible [4].

Preservation of digital material is, therefore, more complex because of the dependencies which exist between the information object itself and its technical environment-an environment that is generally created using rapidly obsolete and therefore fragile technology. This dependence can be at both the software and the hardware levels. In this sense, digital media present more complex problems than any other media libraries have ever had to face [5]. Digital objects are less like artifacts and more like signals- signals that must be continuously refreshed or they disappear [6].

1.   Archiving / Preserving Digital Information

The preservation and archiving of digital information can be seen as one of the greatest challenges for the library and information professionals these days. The greatest dilemma is how to make this information available to future generations. A useful work has been published on archiving/ preserving digital information by the U.S. Commission on Preservation and Access and the Research Libraries Group [7].

The main problem with digital preservation is that digital technology, in comparison to print, is an extremely fragile medium for the cultural memory of the world. The most commonly given example of this fragility is the 1960 United States Census, where raw data stored on magnetic tapes apparently became obsolete and, to all intents and purposes, unreadable by the late nineteen seventies [8]. Another example equally illustrates how readily we can lose our heritage in electronic form when the custodian makes no plans for long-term retention in a changing technical environment. In 1964, the first electronic mail message was sent from either the Massachusetts Institute of Technology, the Carnegie Institute of Technology or Cambridge University. The message does not survive, however, and so there is no documentary record to determine which group sent the path-breaking message [9].

Digital Information has two main weaknesses:

•   Digital storage media, whether magnetic or optical, are subject to relatively rapid decay.

•   Digital information is machine dependent, and to be read accurately it needs specific computer hardware and software. Unfortunately, hardware and software become obsolescent and otherwise unusable [10].

The mix of machinery and media of the digital world increases the need for responsible preservation activity in libraries and information centres.

2.   Strategies for Digital Preservation

Digital preservation is a process by which digital data is preserved in digital form in order to ensure the usability, durability and intellectual integrity of the information contained therein. A more precise definition is: the storage, maintenance, and accessibility of a digital object over the long term, usually as a consequence of applying one or more digital preservation strategies. These strategies may include technology preservation, technology emulation or data migration [11].

The main issues of preservation for digital libraries are similar to those of more conventional libraries, although there are two major complications:

•   Access to digital materials depends on the software and the hardware used;

•   The medium upon which digital materials are stored is in and of itself unstable.

Both of these issues mean that preservation decisions, which could have been delayed in the case of paper documents, need to be made quickly for the explosion of material available in digital format. This means that with electronic materials, there is little difference between retention and preservation [12].

In broad terms, there are four main approaches or strategies which are currently advocated for preserving digital resources. These are:

- Intellectual Preservation (IP) : This strategy includes printing of digital material on paper and recording it on microfilm (sometimes referred to as "change media"). Although paper and microfilm have longer life expectancy than digital media, this does not seem to be a viable option for a growing number of digital resources. Many digital objects simply cannot be printed and would lose most of their unique functionality and attributes by printing. To rely on hard copy for preservation purposes is to "sacrifice most of the unique attributes of a digital object and potentially to destroy the interactive and dynamic nature of the object itself" [13].

- Technology Preservation (TP) : This strategy aims at preserving the software and hardware environment that was used to access the resource when it was created. This approach may be the best solution for some digital objects in the short term but it is not a practical approach in the longer term. It is financially unfeasible to maintain hardware and software platforms that support a digital resource.

- Emulation : It refers to creating new software that mimics the operation of older hardware or software in order to reproduce its performance rather than looking backwards in time to the old technology. Jeff Rothenberg, for example, has recently suggested that there may be sufficient demand for the entrepreneurs to create and archive emulators of software and operating systems that would allow the contents of the digital information to be carried forward and used in its original format [14]. Where complex documents or publications are concerned, there may be some future in investigating Rothenberg's concept of 'encapsulating' data together with all application and system software required to access it and a description of the original hardware environment. Text only 'bootstrap standard' metadata would be then attached to the data, which could provide co textual information and an explanation of how to decode the record itself. The future computer system could use this information to emulate the software so that a document can be seen in as close as possible to its original context [15].

The new techniques in metadata thus can be used as a preservation strategy. As libraries and archives experiment with new formats of metadata as a way to catalogue and classify material, they can encode the metadata to contain preservation information [16]. One example of how metadata can be used in preservation is the Universal Preservation Format (UPF). UPF is a "data file mechanism that incorporates metadata that identifies its contents within a registry of standard data types and serves as the source code for mapping or translating binary composition into accessible forms" [17]. The UPF would allow information to be independent of computer applications, and allow a unique identifier that travels with the document through time to help in the intellectual preservation.

- Data Migration (DM) : Migration is a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display and otherwise use them in the face of constantly changing technology [18].

Unfortunately, digital libraries have little market control over the changing technologies, and so must do their best to migrate materials to new formats and actively support international standards to preserve long-term information or digital materials. In data migration, if we use a commercial vendor's database product and the associated software to migrate across versions of the software, we have no assurance that the byte stream has been maintained over many versions of the software. In this case, since the digital object is encapsulated in a proprietary format, techniques such as performing cyclic redundancy checks are not very useful. The preservation architecture should be based on "non-proprietary and open-source technologies in order to achieve long-term sustainability"[19]. Even in the CPA/RIG report 'Preserving Digital Information [20] standards based approaches to digital preservation are identified as an issue of which digital archives must keep abreast and ensure that their own technological infrastructure conforms to widely adopted standards.

All of the methods mentioned above have drawbacks. Saving the old software and hardware is very expensive and not feasible for digital libraries whose future users would no longer have the right equipment to access the information remotely, even constantly updating the information into new formats is also expensive, time consuming, and risks losing information during migration. As digital works have become more complex, choosing to keep a preservation copy in ASCII or printed on paper causes a loss of information as well as of some of the capabilities and multimedia aspects of the materials. To perform digital preservation truly, libraries must retain "the ability to display, retrieve, manipulate and use the digital information in the   face of constantly changing technology" [21].

3.   Reponsibility for Digital Preservation

Responsibility for preservation is closely linked with the ownership. But in case of digital material, the ownership is diffused. Libraries often pay license fee for accessing the material, which is hosted at the server of the publisher. Unlike the book scenario, libraries do not always hold copies of the material making responsibility for long-term preservation unclear. Though most of the electronic resources to which libraries subscribe or have access to include permission to make back up copies of the resource just in case there is damage to the original. But digital preservation is more than this. To archive a copy of the digital resource, the libraries will have to take permission from the right owner whether the author himself or the publisher or any other intermediary. This may include routine migration or software re-engineering to allow for emulation. Such activities need to be covered under legislation.

4.   Preserving the Integrity of Digital Information

The digital information objects with long-term cultural and intellectual value should be preserved with their integrity. Whatever preservation method is applied, the central goal must be to preserve information integrity. In the digital environment the features that determine information integrity and deserve special attention for archival purpose include 'content, fixity, reference, provenance and content' [22].

The integrity of the information objects in the digital environment is so fragile that as stakeholders disseminate, use, reuse information they can easily destroy valuable information, corrupt the cultural record and ultimately thwart the pursuit of knowledge. One of the greatest dangers to long life of the digital information is the ease with which it can be abandoned and then deliberately or inadvertently destroyed. One of the mechanisms against this danger might rest in enacting laws covering preservation of digital material. A depository system might well serve the purpose. Under such a system, publishers could be legally bound to place with a certified national library/ archive a copy of their published digital works in a standard archival format which in turn would put responsibility on the national library/archive to protect the integrity of the digital material over the long term and retain them in an accessible form for future use. Recently, most of the countries in Europe have extended legal deposit to electronic publications e.g. in early 1996, the British Library proposed that legal deposit should be extended to include non-print materials, even allowing for the possibility of collecting networked (on-line) publications if this becomes technically and economically feasible [23]. This implies a commitment by national library at least to the long-term preservation of digital publications.

However, there are as yet no meaningful laws regarding deposit of digital material. Moreover, the presence of legislation for legal deposit will not solve the problem for libraries and archives. They will still need to understand the resources within their collection for which they take primary preservation responsibilities. They will have to transform the traditional preservation concepts from protecting the physical integrity of the object in specifying the creation and maintenance of the object whose intellectual integrity is the prime concern. They will have to integrate the issue of digital preservation into existing management policies and access agreements.

5.   Coordinated Efforts for Digital Preservation

The librarians and archivists must join hands to develop strategies and policies for preserving digital materials. Coordinated efforts are a must for developing broader solutions to digital preservation. National efforts such as those funded by the Research Libraries Group and Council for Library and Information resources in the U.S. [24] or at a national library level at the National Library of Australia [25] all emphasize the benefit of coordinated activities. Hence new approaches to creating and making digital files cooperatively must be developed. The acceptance of shared responsibility and shared funding commitments will ensure that no valued digital information is lost to future generations.

6.   Conclusion

Preservation of digital material is indeed a very challenging task. The variety and complexity of digital objects are at times overwhelming and will be increasingly so as we develop new technologies for delivering information. Digital preservation and archiving are natural extension of the traditional roles of the libraries with regard to the preservation of non-digital content. As an institution, a library offers a degree of permanence to the material in its collection not easily provided by commercial information organizations. Although considerable change will be required in roles, processes and policies, it seems appropriate that libraries take on this challenge. However, organized and coordinated efforts will be required to find solutions for digital preservation and ensuring access to digital material into the future.

7.   References

1. Conway, Paul. "Preservation in the Digital World" <www.clir.org> (Accessed on 7th June 2003).

2. Nollinger, Mark.   "America Online!" Wired, September, 1995, p. 204.

3. Van Bogart, J.W.   "Magnetic Tape Storage and Handling: A Guide for Libraries and Archives". Washington: Commission on Preservation and Access, 1995, p. 11.

4. Dollar, Charles M.   "Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods". Macerata: University of Macerata Press , 1992.

5. Russell, Kelly. "Digital Preservation: Ensuring Access to Digital Materials Into the Future"<www.leeds.ac.uk/cedars/chapter.html>   (Accessed on 7th June 2003)

6. Lyman, P. and Besser, H.   "Time and Bits Conference Background Paper." Feb. 1998. <http:// .www.ahip.gelty.edu/time and bits/paper.html> (Accessed on 10th July 2003)

7. Task Force on the Archiving of Digital Information. "Preserving Digital Information: Report of the Task Force on Archiving of Digital Information" commissioned by the commission on Preservation and Access and the Research Library Group. Washington, D.C.: Commission on Preservation and Access, 1996 <http://www.rlg.org/Arch TF/> (Accessed on 9th September 2003

8. Ibid.

9. Ibid.

10. Mallinson, J.C. "On the Preservation of Human and Machine Readable Records. Information Technology and Libraries, 7, 1988, pp. 19-23.

11. Russell, Kelly, op. cit.

12. 12. Foot, Mirijam   "A Preservation Policy for Digital Material: A Librarian's Point of View" speech given at the JISc/British Library Workshop. 27-28 November 1995 <www.ukoln.ac.uk> (Accessed on 18 June 2003)

13. Russell, Kelly, op. cit.

14. Rothenberg, J. "Metadata to Support Data Quality and Longevity" <http://ww.computer.org/conference/ meta96/rothenberg paper/ieee.data-quality.html> (Accessed on 16 June 2003)

15. Day, Michael "Extending Metadata for Digital Preservation" <www.ariadne.ac.uk> (Accessed on 16 June 2003)

16. Ibid.

17. Shepard,Thom and MacCam, Dave. "The Universal Preservation Format" <www.info.wgbh.org> (Accessed on 16 June 2003)

18. Task Force on the Archiving of Digital Information. "Preserving Digital Information", op. cit.

19. Jantz, Ronald. "Public Opinion Polls and Digital Preservation: An Application of the Fedara Digital Object Repository System". <http://www.dlib.org/dlib/november03/jantz/11jantzhtml> (Accessed on 24 November 2003)

20. Task Force on the Archiving of Digital Information. "Preserving Digital Information", op. cit.

21. Headstrom, Margaret. "Preserving Digital Information"<www.ukoln.ac.uk> (Accessed on 29 september 2003

22. Task Force on the Archiving of Digital Information. "Preserving Digital Information", op. cit.

23. British Library Research and Innovation Centre. "Proposal for legal Deposit of Non-print Publications" (London: British Library, 1996, accessed 12 June 2003) <http://www.portico.bl.uk/ric/legal/ legalpro.html>

24. The Netlib Project involves national libraries from a number of European countries and focuses on the creation of an infrastructure for distributed digital archives. <http://www.knolib.nl/coop/netlib> (Accessed on 12 June 2003)

25. The National Library of Australia is involved in a number of key digital preservation activities including Pandora "Preserving Access to Networked Documentary Resources of Australia" <http:// www.nia.gov.au/pandora> (Accessed on 7 June 2003)

About Author

Dr. Amritpal Kaur is Reader and Chairperson in DLIS at Guru Nanak Dev University, Amritsar, India. She holds Master Degree in Philosophy and Master and Ph D Degree in LIS. She has published 40 research articles and participated in many national and international seminars, conferences, and workshops. She is guiding 4 Ph.D. scholars. She has won ILA-PV Verghese Award for the Best Paper published in ILA Bulletin and IATLIS-Smt. Ratna Rao Best Woman Teacher Award.
E-mail: amrit_lisc@yahoo.co.in