
PRESERVING THE INFORMATION CONTENT : ISSUES AND STRATEGIES

Dharmista R. Patel

Abstract

The term preservation and conservation are defined here. Cause for loss of knowledge is explained. Discussed the risk assessment for documents and need for preservation in this paper. Various strategic objectives of preservation and issues of preservation like technological, legal, social, etc. are discussed in detail. Major strategies like migration, emulation, XML, etc. for preservation of documents in original format and in digital format also presented in detail. Digital preservation guidelines like documenting resources, standards, etc. are described. Copyright/ copyright management, water marking is also included in present paper.

Keywords : Preservation, Digital Library, Information Preservation, Digital archiving,

1. INTRODUCTION

In order to understand history, people must study its rural life, agricultural heritage, and economic, social, and cultural values. If people want to study and learn from the past, they must preserve all of their present treasures. Preservation allows them to keep materials alive, whole and available for use so that the material can be authoritatively used for the longest time possible. This will allow materials to remain accessible for their heritage, society, and for the guidance of future generations.

Preserving information content of documents is often considered important. Every organization accumulates documents, which are required to be preserved including drawings, maps, imageries, confidential reports, etc. All these documents have different shapes, sizes and different use patterns.

Everyday valuable information is increasing and to save valuable information and to fulfill the user's requirements for posterity, digital documents should be managed and preserved at a tremendous rate. The goal of digital preservation is to maintain the ability to display, retrieve and use digital collections in rapidly changing technological and organizational infrastructures. Preservation is the creation of digital products worth maintaining over time. It also refers to the various methods of keeping digital materials alive into the future [5].

2. DEFINITION

Preservation and conservation are interrelated processes. Their definitions are as follows:

Preservation : A broader term than conservation. It includes all managerial and financial considerations including storage and accommodation provision, staffing levels, policies, techniques and methods involved in preserving library and archive materials and the information contained therein for access and preservation purposes by future generations[11].

Conservation : Conservation is taken as divert intervention to prevent/make good damage to materials [11].

3. RISK ASSESSMENT FOR DOCUMENTS

The following risks can exist in document collections:

- Natural Routine Risks include relative humidity, temperature, light, pollution, and pests.
- Natural Extreme Risks include fire, flood, and earthquakes.
- Man-made Routine Risks include handling, wear, tear and erasing of material.
- Man-made Extreme Risks include war, civil unrest, and vandalism.

4. NEED FOR DIGITAL PRESERVATION

One of the many functions of any library is to preserve information by collecting the documents of the past and of the present and to store and maintain them so that they are available to users of present and the future.

- Problems generated by deteriorating library collection demanded the preservation of library materials, a considerable amount of time, money and efforts too.
- Some collections that a library holds become rare over a period of time, thus becoming one of few surviving copies. Therefore, preserving it for scarcity and specimen value becomes a critical task.
- The length and breadth of recorded materials, for example books and microforms, have created problems in maintaining them. Hence the need for preservation is even more viable.
- Due to increase in technological know how and information awareness, there seems an increasing concern for the preservation of information materials.
- The primary objective of all libraries, archives, and other custodians, is to meet the user's expectations and requirements. These organizations should preserve all materials in all formats.
- Storage media has different formats, durability and handling systems.

5. DIGITAL PRESERVATION ISSUES

More complex than the preservation of print collections is the preservation of digital collections. There are a number of technical, social and legal issues involved with various preservation strategies for digital objects. The following is an explanation of problems and challenges related to the different issues of digital preservation.

5.1 Technological Issues:

There are problems associated with three aspects of digital information technology; the digital storage medium itself, the hardware configuration and software dependency.

Media Problems: Digital information can be presented using a variety of media that differ in longevity and fragility. Common formats are comprised of floppy disks, magnetic tape, cartridges, worm optical storage,

and CD-ROM disks. Both magnetic and optical media have short lifetimes in comparison with media like paper and microfilm. CD ROM and optical worm are very durable, with lifetimes of 100 yrs.

For digital media longevity, digital refreshing i.e. the periodic recopying of the data on to a new medium is better. Refreshing lengthens its lifespan if the storage media is independent of any specific technology.

Hardware Obsolescence: Considering technological obsolescence of the equipment which read the media is important. Punch cards, tapes, Pc XT/AT/286, etc are almost extinct today. It would be impractical to use earlier equipment even if it was available because, the storage and handling capacity of new media is much greater.

Software Dependence: Digital information is stored in formats which are dependent upon particular software to interpret them correctly. There are many more formats than the media. As delivery mechanisms and formats change they increase their complexity.

Interoperability strategy for preserving digital information involves the periodic transfer of digital materials from one hardware/software configuration to another. Its purpose is to preserve the integrity of digital objects and to use them in continuing access to information over extended periods of time by changing technology.

5.2. Legal Issues

More challenging than the technical issues related to copyright and ownership of information are the legal ones. Instead of applying copyright laws, publishers and vendors work under the terms of contract law. Libraries have quite a lot of power in deciding how information is utilized, distributed and stored under the terms of "fair dealing". However, there are various agencies that publish and distribute digital objects whose roles are undefined, making their legal responsibility unclear. The paper-based concept of copyright is not applicable in the digital environment because the control of copying is lost. Digital objects are flexible, easily copied, and remotely accessible by multiple users simultaneously.

5.3. Standard

Another challenge is the use of established standards, protocols and proven methods for preserving digital information. e.g. The International Standards Organization (ISO) has produced a reference model (CCSDS 650.0-W-4.0), for an Open Archival Information System (OAIS), NEN-ISO 15489-1:2001, IDT standard, and use of Standard Generalized Markup Language (SGML) and Extensible Markup Language (XML) are some of the emerging standards suggested..

5.6 Organizational Issues

There are two distinct organizational and managerial issues. The first relates to decisions about who should preserve and what should be preserved. The second is preservation role from librarians. Sometimes lack of preservation policies and procedures, lack of technical staff and financial resources. Also, the legislation does not usually cover electronic publications. It may be more useful to physically transfer document information content, using commercial entities like vendors and publishers may also be more cost effective method of preservation.

5.7 Social Issues

The social dilemma to long term archiving concerns the delicate issue of trust. The economic perversities that exist in scholarly development are a breeding ground for mistrust in relationships between publishers and libraries.

6. PRESERVATION REQUIREMENTS

Along with the difficulties in the preservation of digital material, maintaining the accessibility of digital media is also more complicated than any other non-digital media. For example, when a book is preserved in its original formats, all of its content, layout and format are also preserved. On the other hand, though electronic documents can be saved as a whole, it often loses its layout. Digital preservation is storing information in the appropriate environment, following storage and handling procedures that are agreed upon and copying the digital material into newer, fresher media before the old media deteriorates.

For a preservation project to work smoothly, the initial task is to identify the multiple aspects of a work that has to be preserved. Simply preserving the digital information on several copies of a digital medium is not sufficient. Copying the digital information in to newer, fresher media before the old media deteriorates is the best way to ensure preservation of digital information. Make sure that the digital information can be retrieved and processed in future. In preserving a digital object, it is important to do the following:

Fix the object. A document has a fixed boundary but the boundaries of digital objects are less clear, as they are compound objects created by incorporating different media or by linking to resources from around a network.

Preserve the physical presence. The physical presence here is referring to the series of one's and zero's which is the foundation of digital objects in computer files.

Preserve content. This aspect refers to maintaining the ability to access the content in various formats such as ASCII text, without the establishment of font variations and layout features.

Preserve the presentation. In many types of digital documents (e.g., SGML, XML, and PDF), the layout specifications are separate from the content in order to retain the original look of a document; these layout specifications must also be preserved, along with the content.

Preserve functionality. Digital objects have functionality and need to be preserved. They can contain multimedia components (i.e., text, graphics, audio, and video), hypertext format. They can also contain dynamic content generated automatically from databases, or have navigation functions, such as toolbars, keyword search, or interactive tables of contents. Special efforts must be made to preserve the functionality.

Preserve authenticity. It is essential to preserve the authenticity of digital objects against unauthorized changes and monitoring it to ensure that each copy is in an acceptable condition of the original. Establishing authoritative depositories would also help to achieve this goal.

Locate and refer to the digital object over time. Digital objects can be readily altered, copied or moved. An individual must be able to match a citation to a digital object, and to distinguish it from other versions.

Preserve provenance. The digital objects place of origin and history should be saved to help in confirming that the work is original and its content is intact.

Preserve context. Digital objects are partly defined by their hardware and software dependencies, their mode of distribution, and linkages to other digital objects and so preserving context is a particular challenge.

7. PRESERVATION STRATEGIES

Every organization accumulates documents such as drawings, maps, imageries, confidential reports, etc. These documents come in diverse shapes, sizes and use patterns. Unless preservation strategies

are actively employed, the preserved information will rapidly become inaccessible. Choice of strategy will depend upon the nature of the material and what aspects of the document are to be retained, including content and layout of the information, as well as the computer file itself. Ideally all aspects of a digital document would be preserved, but this is not always possible or feasible.

8. STRATEGIES FOR PRESERVATION IN ORIGINAL FORMAT

Tactics for preserving documents need to support all of the values we have come to relate with paper based materials, like their intellectual content and outward appearance. Traditionally these strategies have focused on repair and binding. Since the mid 20th century, chemical processes, such as washing and deacidifying, have also emerged. Four preservation strategies have gained wide acceptance: increasing the use of permanent paper; climate control; collection maintenance; mass deacidification and paper strengthening.

9. STRATEGIES FOR DIGITAL PRESERVATION

Alternative strategies in preserving original formats include transferring it to other formats. The main strategies used are microfilming and photographic techniques. However digitization is increasingly being considered.

Several strategies try to address the initial digital preservation problem of technology becoming obsolete. These include migrating information through successive generations of technology; using software to emulate the behavior of older machines; preserving original hardware and software to run obsolete programs, and creating hard copies (paper or microform) of digital objects. Using these strategies satisfies some, but not all, preservation goals.

The Digital Preservation investigated three main strategies for preserving digital information, namely:

Migration

Migration has been defined as, "A move from one hardware platform or software system to another, usually because the purchaser or lessee believes the new system to be superior. In libraries the most common example is from the catalog or software of one to that of another. Such a change may or may not require the of data one format to another." [9]

The process of migration involves moving files from one file-encoding format to another. It copies, converts and/or transfers files from one hardware/software environment to another; for example from Word 6 to Word 7, or from a Macintosh environment to a Windows environment. Migration may involve copying digital information from a medium that is becoming obsolete or physically deteriorating to a newer one (e.g., floppy disk to CD-ROM), and/or converting from one format to another (e.g., Microsoft Word to ASCII), and/or moving documents from one platform to another (e.g., VAX to UNIX or LINUX).

Migration aids in preserving the physical presence and the content of digital material. Migration, however, may have challenges in preserving presentation, functionality and context. It usually isn't enough to stay on top of technology becoming obsolete. Migration to new operating environments can often mean that the copy is not exactly the same as the original piece of information. Decisions must be made about what parts of the data (e.g. functionality, presentation) to be migrated need to be preserved. Migration is currently the most common preservation strategy for digital records.

Emulation

Emulation has been defined as, “In digital archiving, a preservation technique that employs special software, called an *emulator*, to translate code and instructions from an original archived software program to enable it to run on a newer platform, obviating the need to preserve obsolete hardware and system software”. [9]

One of the benefits of emulation strategy over migration is that it aids in maintaining the integrity and the “look and feel” of the material very efficiently. This allows the original rendering programs for obsolete digital formats (and files) to be run on future computers.

Once the data is archived with appropriate metadata and software, no other action is needed except for media refreshing until access is desired. One emulator can also be utilized as a solution for several data objects requiring the same operating environment. Though this method may be expensive, it may be best for formats that are challenging to convert.

XML

This acronym stands for eXtensible Mark-up Language and is a text-based language for enhancing data with information about structure and meaning. This opens up a window for various solutions for preservation of material over a long span of time. HTML and XML are – both based on SGML, HTML describes how the data should appear. XML lets you know what the data means, Important for data exchange (e-commerce), XML documents can be built using text-editors or XML-editors (e.g., Merlot, XMLSpy). Therefore, it is promising to use XML an open standard as preservation strategy.

XML metadata advantage

An important part of XML is metadata, or data about data. The popularity of using XML for digital preservation is increasing because it protects the digital material by being platform and software independent. The nature of digital objects is best suited to be as structured markup language for customization and reusability.

Platform neutrality – Transformation – XML-based digital objects can be simply changed to alternative formats i.e. all Windows platforms, and all variants of UNIX and Linux.

Validation – complex XML-based digital objects can be easily validated with DTD. Customized presentation – different presentation of the same digital object can be served to different users.

Searchable content – the modular nature of the digital object content can be searched to offer a higher level of search service to the users.

Output to permanent paper or microfilm

This strategy can correct the object as a whole, preserve the content, and to some extent, preserve the layout. However the output to paper will lead to a heavy loss for hypertext documents despite these drawbacks, a “hybrid strategy” of creating both microfilm and digital copies is gaining support as a technique for reformatting paper originals. The digital copy improves access and functionality, and the microform copy acts as an archival surrogate that can be used for later projects, keeping in mind that the original captured files should always be maintained and preserved.

Technology preservation

Yet another strategy for guaranteeing ongoing access to digital objects by keeping older technology available for use is by preserving technology. The strategy of technology preservation includes the preservation of a digital object with both the software and hardware needed to interpret it. It aids in the recovering of data from obsolete storage media and platform, but it requires a lot of space, and can be quite costly. Storage media, such as contemporary floppy discs, DVDs or CD-ROMs, do not ensure access to files over long periods of time. The pre-requisite in achieving this is often the timely transfer of files to new storage media. However, the availability of the hardware and software required to read the specific files must be present. For instance, imagine the disappearance of the 5 ¼ inch floppy drive. The ramifications would be great. Thus, timely transfer of files to new and contemporary storage media is suggested.

8 DEVELOPING DIGITAL PRESERVATION GUIDELINES

In order to preserve digital objects, libraries need to formulate proper selection guidelines for formats and standards. Effective solutions in acquiring, describing and managing long term preservation of digital resources are decided. To facilitate preservation it is essential to use current standards to create digital objects, to monitor changes incorporated in standards, migrate to new standards when they are established.

Standards

Standards, directives, and requirements for preservation systems are all useful and necessary instruments with which digital archives can be well managed and preserved. Topics covered by the standard include: policy, allocating responsibilities; a method for implementing the standard; educating the different processes of information management, quality control, and auditing. Standards play a key role in automating processes and interoperability i.e. in transferring information between hardware and software platforms as new technologies evolve. Standards are well defined for text (e.g., ASCII), images (e.g., JPEG, TIFF) and encoding documents (e.g., SGML, HTML), standards have not emerged for some other types of information (e.g., databases).

The International Standards Organization (ISO) has produced a reference model (CCSDS 650.0-W-4.0), for an Open Archival Information System (OAIS). The model establishes the minimum requirements for a digital archive to ensure long-term preservation of digital information, and provides a framework for describing and comparing archival architecture and operation. Use of Standard Generalized Markup Language (SGML) and Extensible Markup Language (XML) are some of the emerging standards.

Documenting Resources

A digital preservation guideline is documentation and description of electronic resources. The need for such intentional description is for the reason that digital objects do not carry the visible evidence of creation and use of non-electronic formats. Such clues guide preservation decisions. They also help users to establish that the work is whole and intact, and to understand its importance and the context in which it was created.

NEN-ISO 15489-1:2001, IDT standard by NEN, (Nederland Normalisatie Institute) illustrates the steps that must be taken to satisfactorily position information management according to the tasks of the organisation.

ReMANO (Records Management Applicaties voor de Nederlandse Overheid) is a set of functional requirements for information and archives management.

Two worldwide organisations, the European Commission and the U.S. Ministry of Defense, have also created proposals of requirements or specifications, they are MoReq and DoD 5015.2-STD. In addition to these, the DIRKS (Designing and Implementing Record Keeping Systems)

Metadata

A description of a digital object is “data about data,” or metadata. Metadata is the data that describes the content and attributes of any object in a digital library. It is utilised as the solution to resource discovery and use of any document. Metadata can be stored either as an fundamental part of the document it describes (e.g., embedded in an HTML header) or as part of a separate file of information (e.g., a MARC record). A number of standards like MARC, Dublin Core have emerged. The first workshop took place at OCLC headquarters in Dublin, Ohio, consequently the name “Dublin Core.” The Dublin Core defines a set of fifteen metadata elements—much simpler than those used in traditional library cataloguing. They were designed to be simple enough to be used authors, but at the same time, descriptive enough to be useful in resource discovery.

Unique identifiers

Digital objects can be described by allocating it a unique name and persistent identifier. An identifier is a number, like an ISBN, which is used in a conventional library. An identifier for digital objects helps to establish the authenticity of the object by confirming to a user that the resource accessed is the one cited. Some organizations involved in digital preservation are currently using PURLs (Persistent Uniform Resource Locators), URNs (Uniform Resource Names) or modified DOI (Digital Object Identifiers).

Copyright / rights management

The achievements over the last few years of digital technologies, the rapid development of the Internet and of other communications means, have resulted in an ever increasing need for protecting of copyright and ownership of materials i.e. Intellectual Property Rights (IPRs). Due to the major impact of the Web environment on the accessibility to the material, not only the documents digitalized with very high resolution and visual quality have to be protected, but also the documents digitalized with parameters particularly suitable for the Web collaboratory must be protected. The need of data authentication (integrity) is straightforward. Once the archives make available their documents in a Web collaboratory, they need to be sure that the material has not been modified. [14]

Traditional laws on IPR protection do not seem suitable to solve all the problems raised by this technological revolution. . A system called Electronic Copyright Management Systems (ECMSs), i.e. systems able to manage automatically all the issues related to trading works by protecting their IPR. In the design of effective ECMSs, two different approaches (wrapped & watermarking) may help to solve the problem of IPR protection. [5]

Wrapped is an encrypted system and integrated with an application (the reader) allowing to use the work only in a controlled manner; e.g. the images can just be displayed, but they can not be printed, or the audio files can be played but can not be stored in the hard disk of the user. The content can not be accessed without the proper application. [5]

Watermarked through digital watermarking techniques, whose aim is to firmly and robustly embed IPR related information. The objective of the integrity watermarking process is to ensure image content authentication, detecting content-changing manipulations of the documents [5]. It is defined as: Watermarking is the art of hiding information on the internet and still keeps it from public view. This is actual software, or HTML or JAVA can do it. If you keep it hidden, only a certain sequence, code, password' software, all of them or some of them can make it possible to open. It's really complicated, but the internet

is the best and the worst location to hide things. Watermarking embeds a digital signal in text, image, audio or video files, which may enclose information and evidence of rights to a product's owner or publisher.

10. CONCLUSION

The foremost responsibility of digital libraries and archives is to ensure the future accessibility of information and preservation of valuable materials. Objectives of digital libraries consider organizing information, maintaining intellectual property rights, and presenting, retrieving and visualizing digital material.

Research is underway on the questions of integrity, authenticity, and user requirements for digital resources. Authenticity and integrity of digital information have been underlying concerns in digital preservation because of the ease of altering digital objects and the dynamic nature of digital information.

Strategies which discuss authenticity, appraisal and retention, migration acceptable media and formats, metadata, training and awareness for staff, and monitoring of compliance are defined. Information content management needs to define procedures, including design of the record keeping system, the types of electronic records to be captured by the system and the metadata to be kept with the records.

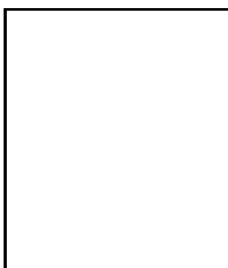
Develop written guidelines on the criteria required for emulation and assessing the migration process and results. The Library should ensure that it has the necessary hardware and software to provide access to the digital collections. A programme of the Commission on Preservation and Access Council on Library and Information Resources supports the efforts of libraries to preserve their paper-based collections and to meet the new preservation challenges of the digital environment. We need to pay great attention to the longtime preservation and use of digital data. Trustworthy models, information systems and environments for copyright management are necessary. Technological development, good practices and successful implementations provide access to their valuable content. · A widely adopted trust model for an open IPR protection and management system is still an open issue.

11. REFERENCES

1. Bhargava, Vijay and Vijayavergiya, Ashok, "Preserving the information content of old documents—some issues.", *Ann. of Lib. Sc. & Doc.*, V.47, No.1, 2000, pp.1-4.
2. Chapman, Stephen and Kenney, Anne R., "Digital conversion of research library materials." *D-Lib Magazine*, Oct. 1996.
3. Conservation and Preservation of Library Materials, Clarion University of Pennsylvania -Department of Library Science-Syllabus <http://www.clarion.edu/libsci/ls588syl011.htm>
4. Digital Information Preservation, by Jaba Das, "Workshop on Digital Libraries: Theory and Practice, DRTC, Bangalore, March 2003
5. Digital Watermarking: a solution to Electronic Copyright Management Systems Requirements, By F. Bartolini, R. Caldelli, V. Cappellini, A. DeRosa, APiva http://www.medicif.org/dig_library/StateArt/lpr/Piva/piva_doc.html
6. Handbook for Digital Projects: A Management Tool for Preservation and access from <http://www.nedcc.org/digital/II.htm>
7. Masoom Raza, M., "Digitization, Preservation and Management of Rare Materials in Modern Library System." *IASLIC Bulletin*, v.49, No. 2, June 2004.

8. Nishta Anilkumar, "Preservation and maintenance of E-Products." Seminar organized by ADINET, INFLIBNET, AMA, 28/8/04, pp.49-52.
9. ODLIS —Online Dictionary for Library and Information Science, by Joan M. Reitz
10. PADI : Preservation Access to Digital Information. <http://www.nla.gov.au/padi/topics/1.html>
11. Preservation Conservation of Information Materials, Univ. of Madras, VIII/1, 73p.
12. Preservation policy: <http://www.ucl.ac.uk/library/preserve.shtml>
13. Preservation Strategies for Digital Objects. <http://www.ischool.utexas.edu/~sgurka/hw6.html>
14. Preservation Strategies for Public Libraries, by Brian J. Baird. <http://clrc.org/progs/baird.shtml>
15. Protecting property rights in digital libraries:advanced methods of robust digital watermarking., by Fraunholder <http://www.collate.de/Workshop->
16. Romano Frank, "E-books and the challenge of preservation", Book titled: Building a National Strategy for Digital Preservation,14p.
17. What is preservation planning?, by Sherelyn Ogden; <http://www.nedcc.org/palms/tleaf11.htm>

About Author



Dharmista R. Patel have been serving this profession for the last 31 years. I have worked in almost all the sections of library. Initiated and implemented various services using Libsys software in a multi-user environment. Team Leader for implementing total error free Bar-coding at SAC library along with Physical Stock Verification. Working with Information Services, providing CAS, SDI service using SAC articles database and various other databases. Literature Search service using Internet and providing Project Literature Search through Intranet to SAC scientists. Preparing specialized bibliographies on demand. Presently developing digital collection – "SAC Digital Collection Repository" consisting of daily Newspaper Clippings, E-journals, E-contents, and papers/reports of SAC scientists. Contributed about 22 papers in various journals, conference proceedings, books, etc. Academic Counselor for M.L.I.S. Course of IGNOU.

Email : drp@sac.isro.gov.in and dharmista52@yahoo.com