
Digital Preservation : How can the Process be Designed in New Different Way ?

G Banumathy

D Gnanasekaran

Abstract

The paper intends to focus on the preservation of digital resources and challenges, scope of the problem, methods of preservation. Preservation constitutes one leg of a tripod that supports informational institutions such as libraries and other leg being access and the development and management of corrections. Without preservation, access becomes impossible and collections will delay and disintegrate. Digital preservation means the planning, resource allocation and application of preservation methods and technologies necessary to ensure that digital information of continuing value remain accessible and usable. The purpose of preservation is to ensure protection of information of enduring value for access by present and future generation.

Keywords : Digital Preservation, Digital Resources, Digital Libraries, Storage media, Digital Archiving, Prototype.

0. Introduction

A library system where librarians exploit the facilities of information technology with intension of sharing resources for providing right and budding information to the right user at right time is one of the basic concept of digital library. Preservation has developed into a critically important part of managing a library's most precious asset i.e. its collections.

Preservation of digital documents involves number of requirements. In particular, all such documents possess a unique collection of core digital attributes that must be retained. Normally documents can be easily by automated mechanism, that can be easily modified, reformatted and can perform arbitrary computation.

- Accessed
- Searched
- Processed

Preserving digital records may require substantial new investments and commitments by organisations, institutions and agencies forcing them to adopt new economic and administrative policies for funding and managing the digital preservation.

1. Need for Digital Content Preservation

The use of computers is changing for ever the way information is created, managed and accessed. The ability to generate, amend and copy information in digital form, to search texts and databases, and to transit information rapidly over networks has led to a dramatic growth in the application of digital technologies.

The rate of change in computing technologies is such that information can be rendered inaccessible within a decade. Digital resources will not survive or remain accessible by accident: pro-active preservation is needed.

Not all information or content needs to be preserved in perpetuity but digital preservation is needed to maximize the nation's and institution's investments in digital resources in the medium term, and over the longer term to secure the intellectual and cultural record that increasingly exists only in digital form.

Digital preservation is a global issue, requiring a global solution. Digital preservation is essential within the shifting sands of technological environments..... It is a means of ensuring genuing longevity for digital resources.

2. Requirements for Digital Content Preservation

In order to preserve digital materials with mass storage capabilities that are accessible and usable to the users, some basic requirements are necessary. Digital preservation requirements may be divided into two ways.

1. The perspective of uses of digital materials – from the view of libraries, archives
2. Other custodians who assume responsibility for the maintenance, preservation and dissemination of information.

To understand the requirements of the digital preservation it is useful the to look at the characteristics of an electronic publication. In particular, electronic publication consists of two main elements

1. A formatted bit stream
2. A program to interpret the format

So, if we manage to preserve the authentic structure of the bit stream and to preserve the interpreter's functionality the problem of digital preservation will be solved.

The potential uses of digital materials varied, unpredictable and almost endless[1]. Users will seek the documents that are easily retrieved and manipulated, transmittable and transportable from repository to the sites of research, presentation and teaching. Preservation digital materials in formats that are reliable and usable, will require long term maintenance of structural characteristics, discriptive metadata, and display, computational, analytical capabilities that are very demanding of both mass storage and software for retrieval and interpretation. [2]

Requirements may be expressed differently by libraries, other types of archives, repositories that are struggling to meet the user expectation with limited technical and financial resources. Moreover, storage systems should be capable of handling the digital information in a wide variety of formats, including

- Text
- Data
- Graphics
- Video and
- Sound

To make the digital preservation affordable to the widest possible range of organisation, individuals, equipment, media, and maintenance costs must be modest.

3. Challenges of Digital Preservation

The purpose of preservation is to ensure protection of information of enduring value for access by present and future generations[3]. Libraries and archives have served as the central institutional focus for preservation. Both types of institutions include preservation as one of their core functions. In recent decades, many major libraries and archives have established formal preservation programs for traditional materials which include regular allocation of resources for preservation, preventive measures to arrest deterioration of materials, remedial measures to restore the usability of selected materials, and the incorporation of preservation needs and requirements into overall program planning.

Preservationists within the library and archival community have been instrumental in developing an array of tools and methodologies to reduce the decay of traditional materials and to restore books and documents that have deteriorated to such an extent that their longevity and usability are threatened.

The critical role of digital libraries and archives in ensuring the future accessibility of information with enduring value has taken a back seat to enhancing access to current and actively used materials. As a consequence, digital preservation remains largely experimental and replete with the risks associated with untested methods; and digital preservation requirements have not been factored into the architecture, resource allocation, or planning for digital libraries.

4. Digital Longevity

Information stored in digital format does not get lost forever because of the fragility of digital works. Documents, data, records and informational and cultural artifacts of all kinds are rapidly being converted into digital form. These documents have to be digitalised by powerful incentives including the ability to make perfect copies of digital artifacts and to publish them on a wide range of media and to disseminate them over network.

There are replication, adaptation and redundancy of hardware, software and data formats which implies that what is readable or interpretable today will be usable long into the future.

The life of data written even to optical media such as CD-ROM or DVD may be measured in years rather than decades. Even today, we have not achieved much stability in data storage technology.

In addition to various problems, there are administrative, procedural, organisational and policy issues surrounding the management of digital material. Digital documents are different from traditional paper documents in the means by which they are

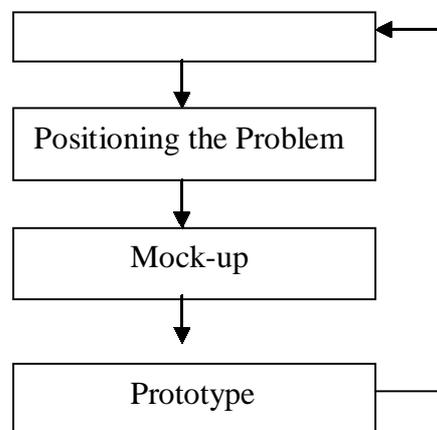
- Generated
- Captured
- Transmitted
- Stored
- Maintained, accessed and managed

Efforts should be made for survival of digital data for longer duration right from its birth. This is known as retention-intention and is essential, as data is very bad at self-preservation. e.g. of born-digital includes theses and dissertations submitted at college and University levels, Government documents, e-journals, etc.

Digital preservation remains largely experimental and replete with the risks associated with untested methods; and digital preservation requirements have not been factored into the architecture, resource allocation or planning for digital libraries.

5. Model for Solving Digital Preservation Problem

According to Simon's decision making process (1977) our problem deals "Intelligence phase" which consists on problem identification and formulation. Therefore, when the decision support supports problem structuring, it is more important that is support structuring information. The information structuring consists of integrating and transforming pieces of insignificant collected information into the appropriate one for useful action.[4]



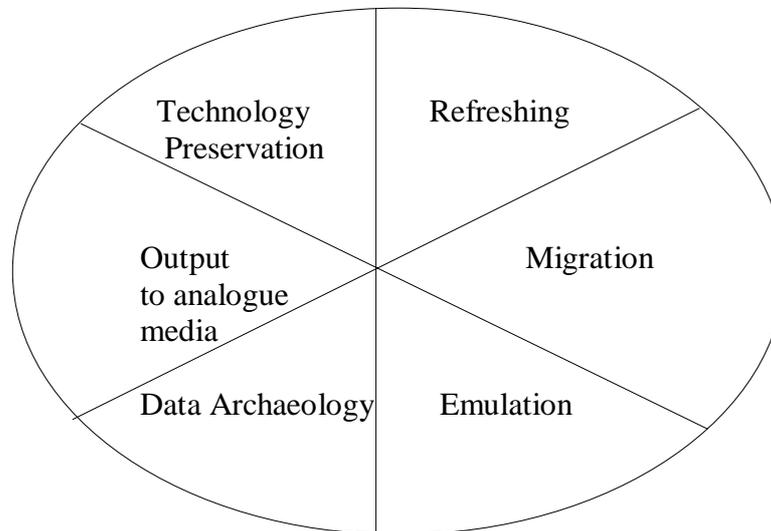
We can therefore be formulated in several ways but this must be consistent with the process chosen to represent the problem, then elaborate a mock-up <<surface>> approach which will be implemented as a prototype.

Bulk of existing digital documents may be textual, multimedia and hypermedia records are likely to become ever more popular and may well become dominant in the near future. Any solution to digital preservation that is limited to text will therefore quickly become obsolete.

Preservation and management of digital records involves interrelated technical, administrative, procedural, organisational and policy issues.

6. Prototyping Methods for Digital Preservation

Prototyping is the process that enables the developer to create a model of the software that must be built. The sequence of events for prototyping paradigm is illustrated below[5].



1. technology preservation
2. refreshing
3. migration and reformatting
4. emulation
5. data archaeology
6. output to analogue media

- **Technology preservation** : It is not a practical approach for data preservation, because is financially unfeasible to go for a regular cycle of media i.e. refreshing maintenance of the hardware and software platforms that support a digital resource.
- **Refreshing** : This is very important in order to achieve longevity. Refreshing copies that bit streams exactly as it is in the document that makes no changes to the underlying data. Refreshing involves periodically moving one file from one physical storage medium to another storage medium in order to avoid physical decay or obsolescence.
- **Migration and Reformatting** : It involves periodically moving files from one file encoding format to another that is very usable in modern computing environment. Moving a wordstar file to word perfect, then to word 5.0 etc. Migration involves change in the configuration of the underlying data, without change in the intellectual content.
The purpose of migration is to preserve the integrity of digital objects and to retained the ability for clients to retrieve the information.
- **Emulation** : Emulation process is the process of re-creation of the hardware and software environment required to access a resource. Its approach is mainly focussed on the application of software rather than the files containing information.
- **Data Archaeology** : In this model, data would be refreshed regularly, but no migration would be performed, and no programmes would preserved to be emulated at a latest stage. It involves recovery of data by using better techniques available in a future. It is last resort for valuable data and programmes.

- **Output to Analogue Media :** This provides a preservation copy in an analogue format. It is fact that microfilm cannot capture all of the features of original digital object that satisfy the access needs of the majority of users. The method employed in this process is termed as "Computer Output to Microfilm" or "COM".

7. Preservation Strategies and their Limitations

While digital technologies are enabling information to be created, manipulated, disseminated, located and stored with increasing ease, preserving access to this information poses a significant challenge. Unless preservation strategies are actively employed, this information will rapidly become inaccessible. Choice of strategy will depend upon the nature of the material and what aspects are to be retained.

Refreshing, that is copying information without changing it, offers a short-term solution for preserving access to digital material by ensuring that information is stored on newer media before the old media deteriorates beyond the point at which the information can be retrieved.

Most librarians and archivists have accepted the basic goal for now atleast-digital preservation depends upon only copying not on the survival of physical media.[6]. Here copying referred as "Refreshing is more complex than simply transferring a stream of bits from old to new media"

Current methods for preserving do not fully support to achieve these objectives. Complex and expensive transformations of digital objects are needed to preserve digital materials in order to remain authentic representations of the original versions and useful souces for analyses and research. [7]

The most commonly used preservation strategy is to transfer digital information from less stable magnetic and optical media by printing page images on paper or microfilm. Archival quality microfilm is projected to last 300 years are more. It have the additional advantage of requiring no special software or hardware for retrieval or viewing.

Another strategy is to preserve digital information in the simplest possible digital formats in order to minimize the requirements for sophisticated retrieval storage. As new media and storage formats were introduced, the date were migrated without any significant change in the structure.

It has the advantage of preserving more of the display, dissemination and computational characteristics of the original messages, while reducing the large variety of customised transformations.

It is fair to say that the state of development in digital preservation remains largely experimental. Only a few libraries, archives, and other institutions have established digital preservation programs, while most research and innovation comes from pilot projects and prototypes.

8. Conclusion

The preservation community has at its disposal a variety of tactics for digital preservation that appear to work effectively for certain types of materials in certain restricted environments. But we have not yet developed solutions that are scalable to the general problem of digital preservation. This is not to suggest that there is or should be a single solution to digital preservation. The methods used will vary depending upon the complexity of the original data objects, the extent to which the functionality for computation, display, indexing, and authentication must be maintained, and the requirements of current or anticipated users. But any solution must be scalable from the laboratory, prototype, or pilot project to the wide range of individuals and institutions who have a need to make digital materials last longer than the current generation of technology permits.

Finally, it would be beneficial to both the preservation community and to those conducting research on issues of longevity, migration, and conversion if there were more venues for exchange of ideas, requirements, and recent developments. I hope that the discussion we are beginning at this convention will lead to more regular and formal processes for linking the needs of scholars and preservationists with the research agendas and projects of scientists.

The future of library and information services is closely linked to the digital preservation and the new technologies will create, collect, store, process and retrieve the information and deliver across the globe.

9. References

1. Gould, Constance.(1988). Information Needs in the Humanities: An Assessment, Stanford, CA: The Research Libraries Group.
2. Michelson, Avra and Jeff Rothenberg. (1992). "Scholarly Communications and Information Technology: Exploring the Impact of Changes in the Research Process on Archives," American Archivist. 55, No. 2: 236-315. Also available at [<http://www.uky.edu/~kiernan/DL/hedstrom.html>]
3. Conway, Paul.(1990). "Archival Preservation in a Nationwide Context," American Archivist, 53, No.2:204-22.
4. Scheer, A.W(1993). "An New Approach to Business Processes". IBM Systems Journal, 32 (1993) pp80-98.
5. Rupak Chakravarthy (2003) "Digital Libraries: Issues and Future", SIS2003 21st Annual Convention and Conference.
6. Lesk, Michael.(1992). Preservation of New Technology: A Report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access. Washington, D.C., Commission on Preservation and Access.
7. Waters, Donald J. (1996) "Realizing Benefits from Inter-Institutional Agreements: The Implications of the Draft Report of the Task Force on Archiving of Digital Information" Washington, D.C.: The Commission on Preservation and Access. Also available at [[URL:http://arl.cni.org/arl/proceedings/127/waters.html](http://arl.cni.org/arl/proceedings/127/waters.html)].

About Authors



Mrs. G. Banumathy is Library Assistant at Anna University, Chennai –25, India and holds M.Com and M.L.I.S.
E-mail: banusundar32@yahoo.co.in



Mr. D. Gnanasekaran is Project Assistant at Anna University, Chennai-25, India in the project "Creation of Digital Library". He holds M.A and M.L.I.S.
E-mail: dsekarg@rediffmail.com