Preservation and Maintenance of the Digital Library : A New Challenge

K R Mulla   Shivakumara A S   M Chandrashekara

Abstract

Libraries, archives, and museums play a critical role in organizing, preserving, and providing access to the cultural and historical resources of society. Digital technologies are used increasingly for information production, distribution, and storage. The institutions that have traditionally assumed responsibility for preserving information face technical, organizational, resource, and legal challenges in taking on the preservation of digital holdings. Maintenance will be critical to digital libraries; especially those who promote broad access to diverse, informal materials. If ignored, maintenance issues within the digital library, especially those relating to its materials will threaten its usefulness and even its long-term viability. We perceive the maintenance problem to be both technical and institutional, and this paper considers the preservation and maintenance of the digital library. The paper examines collection maintenance from several vantage points, including software architecture and the type of collection. The paper ends with an examination of potential technical solutions.

Keywords : Digital Library, Preservation.

0.   Introduction

As with any new technology-based idea, there has been considerable controversy over the definition and possibilities of the term "digital library" to the computer science community, the new technical possibilities. However, as traditionalists in the library community might point out, important issues are being ignored. This paper promotes a view of collections and the long-term consequences of their operation, based on the consideration of digital libraries as social institutions.

This runs contrary to the substantial body of digital library research that focuses on creating the initial Preservation, collections and providing access mechanisms. We believe that the problems must be recast to include long-term issues. By centralizing those issues surrounding the maintenance of institutions and their artifacts, especially the library collection, important considerations for the long-term success of digital libraries emerge. To distinguish our concerns from traditional collection management, we call these materially increasing accessibility and content issues, over the long run, collection maintenance. We use "maintenance" to deliberately invoke "software maintenance" and its often ignored importance for software systems. As discussed below, collection maintenance is likely to be a significant problem in the digital library.

This paper begins by discussing the differing notions of the digital library, anchoring the issues in an analysis of institutional needs and practices. We then examine the various types of collections, including those that include dynamic and informal materials. This consideration of collection types and their control lends itself to analyzing the institutional arrangements and resulting maintenance issues for digital libraries. Maintaining collections, that are extensions of traditional collections (with delineated boundaries), not surprisingly require only extensions of traditional methods. Existing institutional arrangements and resources can be modified to handle these requirements, and maintaining collections that include dynamic and informal information will be possible only with new technical solutions.

1.   Digital Library Collections and Preservation

Libraries are about many things. But, collections have always been at the heart of libraries, whether digital, traditional brick and mortar, or hybrid. Collections will retain that role in the future. However, the concept of what constitutes a collection in the networked environment of digital libraries is undergoing a transformation from the age-old concept of library collection signified by ownership. New concept of a digital collection is evolving incorporating adaptations of many old features and standards, and creation of many brands. This conceptual and pragmatic evolution is far from over. What are the digital library collections ? The question looms as a large problem for practice and for research and development.

The concepts and processes of collection development and collection management are undergoing a transformation. New processes and tools for collection development have emerged which are used for development and management of both, traditional and digital collections. The process of collection management became more closely connected than ever before with means, ways, and policies for access, adding an additional dimension. But digital collections also present distinct and serious challenges related to preservation and archiving. Many libraries and other institutions worldwide are concerned with these issues. Libraries started including digital preservation as a vital part of collection management. A number of national and international bodies are developing standards, tools, and practices related to preservation.

a)   Collection Types and Control

Libraries have always managed their collections, selecting and removing items from their shelves. This has been viewed as a critical function of library management. According to the collection management literature, the practices of collection management are dependent on the type of library collection. We will argue here that new types of collections in the digital library will lend themselves to new types of maintenance issues. Table 1 delineates four types of digital collections.

This is not the only method of distinguishing among "digital libraries"; for example, we could have included access methods or network topologies. Additionally, an actual digital library could have elements of any or all of these types. In a traditional, or paper-based, library, there is considerable control over the collection. Library staff can decide what is and what is not in the collection. Maintenance of the collection is within the preview of the institutional members.

b)   Collection Maintenance

The variation in collection control determines the type of institutional and technical maintenance possible. Collections that are closer to those in traditional libraries can use more traditional control and maintenance mechanisms. Digital libraries that incorporate more individualistic, dynamic, and informal information may need to find new maintenance mechanisms.

Traditional libraries, have developed methods for maintaining their core set of institutional ideals, their community of practice, and their collections of materials. They were based on a constrained collection; i.e., a selection from the bibliographic universe. A traditional library could never cope with much ephemera; it would require too many resources. Going digital, however, changes the cost structure, and collection and maintenance costs need to be revisited. As Figure 2 shows, however, the dynamism and volatility of organizational memories tend to be close to that of traditional libraries. It will be interesting to see how existing maintenance procedures within organizations will adapt.

c)   Preservation of Documents

Distributed bin packing problems and the file allocation problems are known to be NP-hard. This is one reason we have not sought to find an optimal placement for data collections. Moreover, these problems are even harder when the number of sites and sizes of collections are not known in advance. In our model, reliability is more important than performance, configuration like mirrored disks are preferred if possible. However, due to the dynamic nature of the system, with new sites and new collections appearing at any time, it is not possible to statically assign copies to mirrored disk pairs, and a more dynamic data allocation scheme must be devised. For this reason, much work has been done to ensure correctness and consistency for distributed transactions.

Data is replicated so that it can be read despite temporary site failures or network partitions. This is a different goal than long-term reliability, which seeks to preserve data despite permanent site failures or data corruption. Digital library researchers have begun to examine the archiving problem. Some projects have focused on maintaining collection metadata, or on dealing with the data formats. Each of these issues is important, and complements the basic bit-level reliability we seek to provide here. Our trading algorithm could be used in systems like these to place data in the most reliable manner.

2.   Key Components

Digital Library Repository (DLR) is formed by a collection of independent but collaborating sites. Each site manages a collection of digital objects and provides services (to be defined) to other sites. Each site uses one or more computers, and can run different software, as long as it follows certain simple conventions that we describe in this paper. Our architecture is based on following key components.

i)   No Deletions

The deletion is dangerous when sites are managed independently; in particular, it makes it hard to distinguish between a deleted object and one that was corrupted (Morphed into another) and needs to be restored. Ruling out deletions is natural in a digital library, where it is important to keep a historical record. Thus, books are not burnt but removed from circulation.

ii)   Data preservation overview

Archiving sites are autonomous units, managed by different organizations, and thus are not under any centralized control. Each site has a quantity of archival storage. New archival storage may be added to a site at any time. The basic unit of archived data is the data collection. A collection represents a set of related data, for example a group of files, a database table, a set of documents etc. Collections may contain any number of data items, and different collections may be of different sizes. Clients can also read archived data at their own site or at a site storing a copy of the collection. We also need a method for determining reliability. The goal of an archiving system is to reliably protect data, despite the potential for site failures.

Thus, we use the following concepts:

   a) Site reliability : The probability that a site will not fail.

   b) Local data reliability : The probability that the collections owned by a particular site will not be lost.

   c) Global data reliability : The probability that no collection owned by any site will be lost.

iii)   Digital Preservation Responsibilities

Digital preservation is not an isolated problem affecting only large libraries and archives. Some digital materials exist in holdings of other libraries for which the institution assumes preservation responsibility. In fact, the research libraries, which tend to be larger than the archives, museums, and special libraries, are not quite as likely to hold digital materials as the other types of institutions.

iv)   Digital Preservation Policies and Practices

Digital preservation policies and practices are not well developed in academic institutions. One common reason that institutions appear not to develop digital preservation policies is that they have not yet assumed responsibility for preserving materials in digital form. However, taking responsibility for digital preservation does not necessarily mean that institutions use policies to govern their digital preservation activities. Only half of the institutions with digital materials in their holdings have written digital preservation policies.

v)   Preservation Practices

Effective digital preservation requires life-cycle management of digital information from the point of creation through storage, migration, and providing access on a continuing basis. Few institutions have established methods in place for digital preservation. Institutions limit the acceptable formats to flat files while others accept several different formats such as PDF, TIFF, SGML, and word processing formats.

vi)   Problems and Threats to Digital Preservation

The technological obsolescence is the greatest threat to loss of digital materials, followed closely by insufficient resources and an inefficient policy or plan for digital preservation. The lack of resources for digital preservation is the greatest threat in institutions that have policies in place for digital preservation.

vii)   The problem of digital changes and user expectations

Digital technology and high speed networks are leading to sweeping changes throughout society, and moving image production and distribution are in no way immune to either the technological changes or to the social expectations that these changes have induced. In the past completed digital effects were transferred back onto film and inter cut with the rest of a production, but as general moving image production itself becomes increasingly digital, this intermediary transfer to film will become far common. Small-budget independent productions are increasingly being shot and edited in digital form. According to director Mike Figgis, "there is clearly a technical revolution taking place you can edit a film on a laptop, and there is the Internet, the streaming and downloading capabilities. These are the technical elements of the revolution" (Silverman 2000).

viii)   Problems with preserving anything Digital

Information encoded and stored in digital form is fragile, but quite different from film stock. Digital storage shares some characteristics with video storage but it is different from analog storage formats i.e. film and video. But print archivists and special collection librarians, who aggressively pursue print-based collection development in their particular specialty areas, claim that it should be the responsibility of computing staff of their organisation to pursue collection development of material originating in digital form. The costs for handling digital materials diminish and as strategies for long-term maintenance of

digital files become better known, reasons for handling digital material separately will start to fade, and administrators will begin to realize that digital files of moving images have much more in common with film and video than with word-processing files and databases.

ix)   General Approaches to Digital Preservation

This is a brief history of the approaches to preservation of simple types of digital materials. 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 Group formed a task force to explore how significant this problem was really. 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 use threaten to render the life of information in the digital age as, to borrow a phrase from (Task Force 1996)".

The wide body of digital works in Task Force 1996 involves periodically moving a file from one physical storage medium to another to avoid the physical decay or the obsolescence of that medium. Two key approaches have been proposed to deal with the problem of changing file formats (Task Force 1996): migration and emulation.

3.   Conclusion

This paper examined some of the mechanisms required to maintain the digital library. Collection and maintenance is a significant issue in the digital library more than in traditional libraries or current organizational memory repositories. The broadly construed and narrowly construed digital libraries were examined in this paper. The narrowly construed library is analogous to the traditional library where the collection has known boundaries. Because of the possibility of control over the collection in the narrowly construed library, many of the institutional mechanisms for maintaining collections can be assimilated from the traditional library.

The inclusion of dynamic and informal materials in the collection, leads to serious control and long-run maintenance issues. Because of the lack of control over the collections, technical mechanisms will be needed for collection maintenance. While we need several technical possibilities for collection maintenance, we perceive this problem to be both technical and institutional. However, a strictly technical emphasis will not lead to an adequate understanding of the long-run issues in digital library use. The digital library is more than a set of technologies; it is also a social institution with long-term needs and maintenance requirements.

By combining their vast set of skills in handling of analog objects as well as moving to new paradigms provoked by the digital age, moving image archivists can continue to play a critical role in preserving our cultural heritage and ensuring that today's works will last well beyond the life of the team that produces them.

4.   References

1.   Benjamin, Walter (1978). "The Work of Art in the Age of Mechanical Reproduction."   Illuminations. New York: Schocken.

2   Besser, Howard (forthcoming). Longevity of Electronic Art, Proceedings of the International Conference on Hypermedia and Informatics in Museums (September 2001).

3.   Besser, Howard (2000a) Digital Longevity, in Maxine Sitts (ed.) Handbook for Digital Projects: A Management Tool for Preservation and Access, Andover, MA: Northeast Document Conservation Center, pages 155-166.

4.   Besser, Howard (2000b). Bibliography of Moving Image Indexing (website) (http://www.gseis.ucla.edu/~howard/Classes/287-mov-index-bib.html)

5.   Besser, Howard (1997). The Changing Role of Photographic Collections with the Advent of Digitization, in Katherine Jones-Garmil (ed.), The Wired Museum, Washington: American Association of Museums, pages 115-127.

6.   Besser, Howard (1994). Fast Forward: The Future of Moving Image Collections, in Gary Handman (ed.), Video Collection Management and Development: A Multi-type Library Perspective, and Westport, CT: Greenwood, pages 411-426

7.   Besser, Howard (1987). Digital Images for Museums, Museum Studies Journal 3 (1), Fall/Winter, pages 74-81.

8.   Council on Library and Information Resources (2000). Authenticity in a Digital Environment, Washington, D.C.: Council on Library and Information Resources May.

9.   Davis, Ben (2000). Digital Storytelling, Razorfish Reports #24, June 16 (http://reports.razorfish.com/frame.html?rr024_film)

10.   Digital Light Processing (2001). Where is DLP Cinema? (Website) (http://www.dlp.com/dlp/cinema/where.asp)

11.   Goodrum, Abby (1998). Representing moving images: Implications for developers of digital video collections, Proceedings of the 1998 Meeting of the American Society for Information Science.

12.   Hummelen, Ijsbrand and Dionne Sille eds. (1999). Modern art: who cares? : An interdisciplinary research project and international symposium on the conservation of modern and contemporary art, Amsterdam: Foundation for the Conservation of Modern Art and Netherlands Institute for Cultural Heritage.

13.   Laguna Research Partners (2000). Notes from INFOCOMM 2000 (Industry Brief), June 22 (http://www.lrponline.net/infocomm2000_1.PDF)

14.   Laurenson, Pip (2000) Between Cinema and a Hard Place: The Conservation and Documentation of a Video Installation by Gary Hill, Talk to Preservation of Electronic Media session, 28th Annual Meeting of American Institute for Conservation of Historic & Artistic Works, Philadelphia, June 9, 2000.

15.   Lyman, Peter and Howard Besser (1998). Defining the Problem of Our Vanishing Memory: Background, Current Status, Models for Resolution in Margaret MacLean and Ben H. Davis (eds.), Time and Bits: Managing Digital Continuity, Los Angeles: Getty Information Institute and Getty Conservation Institute, pages 11-20.

16.   MacLean, Margaret and Ben H. Davis (eds.) (1998), Time and Bits: Managing Digital Continuity, Los Angeles: Getty Information Institute and Getty Conservation Institute.

17.   Mallinson, John C. (1986). Preserving machine-readable archival records for the millenia, Archivaria 22 (summer), pages 147-152.

18.   National Institute for Standards and Technology (2001). Digital Cinema 2001 Conference, Gaithersburg, MD, January 11-12 (http://digitalcinema.nist.gov/)

19.   Rothenberg, Jeff (2000). An Experiment in Using Emulation to preserve Digital Publications, Den Haag: Koninklijke Bibliotheek, 2000. - (NEDLIB Report series; 1) (http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf).

20.   Rothenberg, Jeff (1999). Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation, Washington, D.C.: Council on Library and Information Resources, January (http://www.clir.org/pubs/abstract/pub77.html)

21.   Rothenberg, Jeff (1995). Ensuring the Longevity of Digital Documents, Scientific American, 272(1): pages 42-7.

22.   Sanders, Terry (1997). Into the Future: Preservation of Information in the Electronic Age, Santa Monica: American Film Foundation (16 mm film, 60 minutes)

23.   Silverman, Jason (2000). Digital Cinema Plays with Form, Wired News, April 12 (http://www.wired.com/news/culture/0,1284,35098,00.html)

24.   Task Force on Archiving of Digital Information (1996). Preserving Digital Information, Commission on Preservation and Access and Research Libraries Group (http://www.rlg.org/ArchTF/tfadi.index.htm)

25.   Turner, James M. (1999). Metadata for moving image documents. Proceedings of the 20th National Online Meeting, New York, May 18-20, 1999, edited by Martha E. Williams. Medford, NJ: Information Today, 477-486.

Annexure - I

1).

	Traditional library	DIGITAL LIBRARY		
		Organizational repository (organizational memory)	World wide Web	Usenet or other computer-mediated communication system
Types of collection	Monographs, Serials, Special collections	Documents and files	Text and multimedia notes	Messages
Authoring agencies	Authors and their publishers	Organizational members or sanctioned individuals	Anyone	Anyone
Collection control	Selection by organizational member (e.g. bibliographers)	Selection by organizational member. May have requirements of official approval	None (or individual)	None (or moderator)

Table 1: Collection control in digital libraries (Roy ???)

2).

	Dynamism of Versions	
Low	Traditional libraries Organizational Memories (Organizational repositories)	WWW
Dynamism of contents on contents	Usenet or other computer mediated communication system	
High		

Figure 2. Dynamism (volatility) in types of digital library collections

About Authors

K R Mulla is working in HKBK College of Engineering, Nagawara, Bangalore.
E-mail : mulla_kamalasab@yahoo.com

Shivakumara A S is working inHKBK College of Engineering, Nagawara, Bangalore.
E-mail : shivaksi@yahoo.com

M Chandrashakara is working as Lecturer in Department of Library Information
Science., University of Mysore, Mysore.
E-mail : chandram5@yahoo.com