DIGITAL INFORMATION MANAGEMENT : PRESERVATION ISSUES AND STRATEGIES

K Vinitha **S Kanthimathi** **K Tharani Devi**

Abstract

Transition form print format to electronic format has given rise to various issues in preserving the digital form. The problem of preservation is further complicated by the rapid obsolescence of the hardware and software required to interpret and present digital documents. The paper highlights on necessity in preserving the digital form, strategies related to preservation, technologies and guidelines for digital preservations. Rapid technological obsolescence combined with relatively short-lived media means that collection and preservation is a must in a digital era.

Keywords : Digital Library, Content - Creation Standards, Digital Preservation.

1. INTRODUCTION

Widespread transition of Knowledge from print format to electronic format has also given rise to the problem of its preservation in digital form. The problem of its preservation in digital form. The problem of preservation is further complicated by the rapid obsolescence of the hardware and software required to interpret and present digital documents. Ensuring to interpret and present digital information necessarily involves copying or transforming digital documents to run on current media, software, hardware and operating systems. This paper explores the issues surrounding the preservation of digital information and highlights ways to address them.

2. DIGITAL PRESERVATION

“Digital Preservation” or “digital archiving” essentially aims at taking steps to ensure the longevity of electronic documents. It applies the longevity of electronic documents. It applies to documents that are either “ born digital “ and stored on-line (or on CD-ROM, diskettes or other physical carriers) or to the products of analog-to-digital conversion, if long-term access is intended.

3. PROBLEM

The fundamental problem of preserving electronic documents or “ digital objects” stems from the fact that the objects unlike non-digital formats are accessible only by using combinations of computer hardware and software. Market competition means that this hardware and software can become obsolete in cycles of less than three years. Ensuring ongoing access, therefore, requires currency with technology changes, and moving digital objects from obsolete to current file formats, storage media, operating systems ad so on…Number of other technical, social and legal issues add to the difficulty of the task. These include

- The increasing complexity of digital objects (incorporating text, images, audio, video in various formats) and their increasing software dependence (e.g., storage in databases);
- The rapidly increasing number of digital objects and proliferation of document standards and formats;

- The lack of planning to incorporate preservation needs in systems and lack of availability of off-the shelf products supporting preservation needs;
- The lack of consideration of long-term access requirements when creating digital products;
- The absence of widely-accepted standards which will assure access overtime;
- Copyright/intellectual property rights that may interfere with the ability to preserve digital objects through systematic copying;
- Unstable storage media (e.g., diskettes) whose life span is limited;
- a lack of technical expertise in collection managers and preservation experts;

4. PRESERVATION REQUIREMENTS

Preservation measures ensure that a document digital or otherwise is accessible in usable form overtime. Maintain the accessibility of digital media, however, is much more complex than with such non-digital media as paper. For example, when a book is preserved in its original format, all aspects of the book are preserved. it is practically impossible to extract individual elements e.g., content without layout, because they are inextricably linked. Even reformatting to paper or microfilm does not completely divorce content form layout as page sequences and physical appearance, for instance, can still be captured. Digital objects, in contrast, are easily decomposed into individual elements, and significantly more effort must be made to preserve them as a "whole". For example, one can retain the content of an electronic document, while losing the layout. Further, one can keep its physical presence (i.e., a file) but fail to preserve its readability.

In the digital world, then, the first task is to identify the multiple aspects of a work that must be preserved. Next, to succeed in the preservation of digital objects, preservation measures must ensure that as many of these aspects as possible persist overtime. In preserving digital objects, we aim to:

4.1 Integrity of Objects

A book is a book, or it exists in a set, but what are the boundaries of a document in a hypertext environment? The boundaries of digital objects are less clear, especially if they are compound objects created by assembling different media or by linking to resources from around a network.

4.2 Physical preservation

The physical presence in this case refers to the computer file, the series of "1's and "0" that are the basis of digital objects.

4.3 Content Preservation

This aspect refers to maintaining the ability to access the content at its lowest level, such as ASCII text, without the embellishments of fonts variations and layout features.

4.4 Presentation Format

Content is typically rendered in some presentation format or layout, that includes different font faces and sizes, the use of white space, columns, marginalia, headers, footers, pagination, and so on. In many types of digital documents (e.g., SGML, XML, and some forms of PDF), the layout specifications are separate from the content. To retain the original look of a document, these layout specifications must also be preserved, especially when they contribute significantly to the understanding and interpretation of the content.

4.5 Functionality

Digital objects have a functionality that goes far beyond traditional paper documents. They can contain multimedia components (i.e., text, graphics, audio, and video), exist in hypertext format.

4.6 Authenticity

An individual accessing the object must be able to verify that it is the one s/he wants, and that the transformations to keep it accessible have preserved its original form.

Thus, activities to guard authenticity include securing digital objects against unauthorized changes and monitoring digital objects through multiple “copying” cycles to ensure that each copy is an acceptable re edition on the original Establishing authoritative depositories would also help achieve this goal.

4.7 Location and Reference

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 or editions.

4.8 Provenance

Provenance is an archival concept that asserts the origin and chain of custody of an object and contributes to defining it as a whole. Imprint statements and bookplates,

For example, partly fill this role for formally published items. Establishing an object’s origin and history help confirm that the work is authentic and its content are intact.

4.9 Preserve context

Digital objects are partly defined by their hardware and software dependencies, their mode of distribution and linkages to other digital objects. Preserving digital objects may mean weaning them from some technical dependencies, changing the mode of distribution, and deactivating links. In these circumstances, preserving context is a particular challenge.[1]

5. PRESERVATION STRATEGIES

Several strategies attempt to address the primary digital preservation problem of technological obsolescence. 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. Each of these strategies meets some, but not all, preservation goals.

5.1 Migrations and Transfer

Migration is the primary strategy articulated by most organizations that plan to preserve digital objects. It covers a range of activities to periodically copy, convert or transfer digital information from one generation of technology to subsequent ones. 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), Migration certainly preserves the physical presence and the content of a digital object. However, it may not preserve presentation, functionality and context. For example, presentation elements such as bolding and italics may disappear, and the functionality and context provided by links between database entries may be lost because the links break. Successive migrations may eventually result in unacceptable data loss. The focus is on limiting the loss and retaining the content in a usable form. Migration is undeniably an important strategy for preserving digital objects. However, it has yet to be tested and proven as a mechanism for managing complex multimedia objects over the long term.[2]

5.2 Emulation

Emulation refers to creating new software that mimics the operations of older hardware of software in order to reproduce its performance. Thus, not only are physical presence and content preserved, but also digital objects could display original features (e.g., layout) and functionality available with the older software. Emulation has recently attracted attention as a potential strategy to assist preservation, recognizing that some electronic material that is highly dependent on particular hardware and software will not lend itself to migration.

Emulation is used to provide “backward compatibility” for video games, and to model how future systems might run. Emulators exist for some obsolete systems, however, emulation for preserving digital objects over the long term has not been widely tested or priced.

5.3 Output to Paper or Microfilm

Outputting a hard copy of a digital file is a “low tech“ solution that can result in a well-standardized product with a life expectancy of several hundred years. Certainly, this strategy could fix the object as a whole and preserve content and to some extent layout. However, a decreasing number of publications (flat files, printable formats) lend themselves to such methods. For example, output to paper will lead to great functional loss for hypertext documents, and cannot capture multimedia. 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 enhances access and functionality, and microform copy acts as an archival surrogate.

5.4 Preservation Technology

Another method for ensuring ongoing access to digital objects would be to simply keep older technology available for use. Although this would preserve content and enable future generations to view digital objects in their native format with original layout and functionality, creating hardware or software “museums” is prohibitive in cose, space and technical support requirements. At best, this method is an interim measure when migration is not possible.

6. CONTROL

To increase the probability that digital objects will be preserved, organizations need to lay appropriate groundwork. One way is to develop and implement the most effective practices in acquiring, describing and managing digital resources.

7. STANDARDS

To facilitate preservation, the best practice is to adopt a three-part approach:

- Use current standards to create digital objects;
- Monitor standards as they change;
- Migrate to new standards as they are established.

Most digital preservation guidelines advocate collecting digital objects in standard formats. Unfortunately, while 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). In addition, not only do standards change rapidly, but vendors seeking to increase their market share incorporate “value-added” features to accepted standards. Institutions performing the digital preservation activity can also be subject to standardization.

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.

8. DEVELOPING DIGITAL PRESERVATION GUIDELINES

Archives and record-keeping bodies in Europe, North America and Australia have taken the lead in developing best practices and functional requirements that address some preservation issues. Common elements include.

- Recognizing the creator's responsibility to preserve his/her work initially;
- Identifying responsibilities of the archive;
- Adopting appropriate selection guidelines (which emphasize standards when a choice of formats exists);
- Securing archived items from intentional or unintentional alteration;
- Providing contextual documentation including a history of creation, transmission and use, and audit trails;
- Describing digital objects completely.

Some libraries have developed similar guidelines. However, a recent survey concluded that most digital preservation guidelines focus on creation, receipt and capture of digital objects, and do not satisfactorily address their long-term preservation.

8.1 Documenting Resources

A recurrent theme in digital preservation guidelines is documentation and description of electronic resources. The need for such deliberate description stems in part from the fact that digital objects do not carry the visible evidence of creation and use (imprints, bindings, bookplates, marginalia, or Scotch tape) 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 provenance and the context in which it was created.

8.2 Metadata

A description of a digital object is "data about data," or metadata. Such descriptive data should include the contextual information crucial to the long-term management of electronic information. Metadata elements useful in preservation might include.

- Identifiers;
- Hardware, operating system and software required to access a document;
- Physical details of tangible format publications such as CD-ROM, floppy disks;
- Encoding standard and version;
- Migration history and its success;
- Data to assist determining authenticity;
- Rights management information;
- Versions and dates.

Conversion projects would employ additional metadata elements such as the capture device, resolution, compression, source material, and producer (of the digital document). Existing metadata schemes (e.g., MARC, Dublin Core) provide for some such information capture, but there is no consensus on which approach will work best for preservation purposes. In MARC, for example, some of the necessary data is captured in note' fields that may not use sufficiently detailed or consistent language for subsequent search and retrieval.

8.3 Unique identifiers

One element of describing a digital object is to assign it a unique and persistent identifier. An identifier is a number, like an ISBN, which is associated with a particular instance of a digital object. Unlike an URL, it is independent of the location of that object. A unique, widely supported identifier for digital objects would help establish the authenticity of the object by confirming to a user that the resource s/he is accessing the one cited. It could also help establish the relationship between copies or versions of digital objects, as any modification of the original would result in a modification of its identifier. Some organizations involved in digital preservation are currently using PURL (Persistent Uniform Resource Locators), URN (Uniform Resource Names) or modified Digital Object Identifiers. As yet, no single digital identifier has achieved widespread, international acceptance.

8.4 Linking Metadata with Content

Metadata can be stored either as an integral part of the document it describes (e.g., embedded in an HTML header) or as part of a separate file of information (e.g., MARC record). One way of linking metadata and the digital object is to package them together. To this end, the aforementioned Reference Model for an Open Archival Information System (OAIS) proposes an "information package" comprised of "content information" and "preservation descriptive information". Along similar lines a working group of the Society of Motion Picture and Television Engineers has developed a Universal Preservation Format (UPF), a data-file mechanism that uses a container structure to incorporate metadata into digital media objects. Although primarily developed for audio-visual data, the principles underlying the UPF may have wider applicability.

8.5 Establishing Infrastructure

A final way in which libraries are gaining control over the preservation of digital objects is by building preservation requirements into systems development. The Australian National Library, for example has started a Digital Services Project to develop systems to manage its digital collections and to support cooperative, shared access."

The next steps remain ill defined. To reduce the uncertainty surrounding digital preservation, we can encourage creators to use practices that will aid long-term preservation, such as creating valid documents in standard document formats and address the question of which preservation strategy will work best by testing migration and emulation and identifying associated costs and tradeoffs in each case.

9. CONCLUSION

Rapid technological obsolescence combined with relatively short-lived media means that collections must be actively managed. Simply collecting and "shelving" important works, a passive strategy that works to some extent for paper-based publications, is insufficient to ensure digital objects will survive in perpetuity.

10. REFERENCE

1. Colin Webb, "Migration Trials: Migrating publications on floppy disk to CD-R". October 1997. http://www.nla.gov.au/nla/staffpaper/cwebb7.html (Jan.20,1999)

2. http://www.leeds.ac.uk/cedars/

3. http://www.news.cornell.edu/releases/Nov98/preserving.digital.bs.html

4. UPF Homepage: http://info.wgbh.org/upf/

5. Consultative Committee for Space Data Systems. 1998. Reference Model for an Open Archival Information System (OAIS), ed. L.Reich and D. Sawyer. CCSDS 650.0-W-3.0. White Book, 17 September, 1998,p.17

6. The full text of this report, titled *"Avoiding Technological Quicksand*: Finding a Viable Technological Foundation for Digital Preservation", is available at http://www.clir.org/pubs/reports/reports.html

7. Preserving digital information: report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group. Washington, E.C.: Commission of Preservation and Access,1996.http://www.rlg.org/ArchTF/

About Authors

Ms. K. Vinitha is working as Librarian in National Engineering College, Kovilpatti, Tamil Nadu.

E-mail : vinithakaran@yahoo.com

Dr. S Kanthimathi is working as Librarian (Senior Grade) at Rani Anna Constituent College for Women, Tirunelveli, Tamil Nadu. She has done MLIS and PhD. She is a Research Guide for MPhil and PhD scholars at Manaonmanium Sundaranar University, Tirunelveli. She is a life member of various professional associations. She has contributed 15 papers in seminars, conferences & journals.

K. Tharani Devi is a Research Scholar in Alagappa University, Karaikudi, TN.

E-mail : tharani_ks@yahoo.com