

## Issues and Strategies in Digital Preservation

V Harineeswaran

K Nithyananda

### Abstract

*This paper provides an overview of the major digital preservation activities and preservation of digital records, developed in the library and Information science in order to identify factors that play a key role in ensuring the long term preservation of the records. The paper presents the principles, digital preservation issues and its strategies.*

**Keywords:** Digital Preservation, Longevity, Documents, Records, Issues, Strategies,

### Introduction

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

### What is meant by “Digital Preservation”?

For the purpose of clarity it is worth establishing some basic distinctions between common terms which will occur in this chapter. Even among experts there are misunderstandings based on subtle differences in the way terms are defined. The current lack of accepted terminology in this area makes progress more difficult. Digital preservation involves various stakeholders from different communities; thus clarity of language is critical. For librarians, archivists and technologists a term like “archive” or “metadata” can mean very different things!

Digital preservation is a process by which data is preserved in the digital format for ensuring the usability, durability and intellectual integrity of the information contained therein. A more precise definition is: the storage, maintenance, and accessibility of a digital object over the long term, usually as a consequence of applying one or more digital preservation strategies. These strategies may include technology preservation, technology emulation or data migration. All of which are discussed in more details.

According to Hendley, at a basic level, all digital preservation strategies involve the following tasks:

- ◆ Preserving the digital medium that holds the digital information by storing it in the correct environment and following agreed storage and handling procedures;
- ◆ Copying the digital information onto newer, fresher media before the old media deteriorates past the point where the digital information can be read or becomes so obsolete that we can no longer find a storage device to read it on
- ◆ Preserving the integrity of the digital information during the copying process.

Likewise Jeff Rothenberg suggests that all digital objects have a set of “core digital attributes” which must be retained through any preservation process. For Rothenberg these attributes include the ability to be

- ◆ copied perfectly;
- ◆ accessed without geographic restraint;
- ◆ disseminated at virtually no incremental cost (given the appropriate technical infrastructure);
- ◆ Machine-readable so that they can be accessed, searched and processed by automated mechanisms.

Digital Preservation often refers to different things. For example, *preservation digitization* (sometimes referred to as digital preservation) is a means of reformatting rare or fragile materials to digital form as part of a preservation strategy. In the early 1990s a number of high profile projects were funded in the US and focused on what was termed “digital preservation” at the time but is perhaps more accurately described as *preservation digitization*.

### **Issues**

The fundamental issues of preserving electronic documents or “digital objects” stems from the fact that the object 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 within a short period. Ensuring ongoing access, therefore, requires currency with technology changes, and moving digital objects from obsolete to current file formats, storage media, operating systems and so on. Numbers 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; and
- ◆ a lack of technical expertise in collection managers and preservation experts;

### **Types of Preservation**

**a) Integrity of objects:** a book is a book or it exists in a set, but what are the boundaries of documents 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.

**b) Physical Preservation:** the physical presence in this case refers to the computer file, the series of "1" and "0" that are the basis of digital objects

**c) Contents Preservation:** this aspect refers to maintaining the ability to access the content at its lowest level, such ASCII text, without the embellishments of fonts, variations and layout features.

**d) Presentation Format** : content is typically rendered in some presentation format or layout, that include 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.

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

**f) Authenticity**: an individual accessing the object must be able to verify once authenticity and after accessing the object be intact to 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 reedition on the original establishing authoritative depositories.

**g) 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.

**h) 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. Imprints 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.

**i) 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

## 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.

**a) Migrations and Transfer:** migration is the primary strategy articulated by most organization 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 object. However, it may not preserve presentation, functionally and context. For example, presentation elements such as bolding and italics get disappears and the functionality and context provided by links between database entries get lost when the link breaks. 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.

**b) Emulation:** emulation refers to creating new software that mimics the operations of older hardware and software in order to reproduce its performance. Thus, not only the physical presence and content preserved, but also digital objects 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.

**c) Out to Paper or Microfilm:** outputting a hard copy of a digital file is a “low tech” solution that can result in a well-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.

**d) 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 cost, space and technical support requirements. At best, this method is an interim measure when migration is not possible.

## **Conclusion**

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, Without a continuing dialogue between humanists, preservation, and the endeavors, and it is challenging for those of us outside the scientific community to keep up with and evaluate new product. I hope that the discussion we are beginning at this national seminar will lead to more regulate and formal processes for linking the needs of scholars and preservation with the research agendas and project of scientific.

## **References**

1. **K. Vinitha, S.Kanthimathi and K.Tharani Devi** “Digital Information Management”. Delhi, Mahamaya Publishing House, 2008.pp.241.

2. **Kelly Russell** "Digital Preservation Ensuring Access to Digital Material Information the future" June 1999.
3. **M.Nagarajan, B Ramesh Babu** "strategies for Digital Preservation" Tamilnadu SRM University Publication, 2008, pp-178.
4. **M. Masoom Raza, R.L. Arora** "Digitization, Preservation & Management of Rare Materials in Modern Library System" IASLIC Bulletin, 2002, pp.89-91.
5. [http:// www.leeds.ac.uk/cedars](http://www.leeds.ac.uk/cedars)

#### **About Authors**

**Mr. V Harineeswaran**, Library Trainee, National Institute of Technical and Teacher Training Research Institute

**Mr. K Nithyananda**, Special Officer, University Library, Anna University, Chennai.