

---

---

## **PRESERVATION OF DIGITIZED INFORMATION: A NEW CHALLENGES IN THE 21ST CENTURY**

***Subrata Sur***

### **Abstract**

*Digital information preservation is always the thinking of library and information society. Transition from print format to electronic format has given rise to various issues in preserving the digital form. Preservation problem is complicated by the rapid obsolescence of the hardware and software required to interpret and present digital documents. This paper highlights the necessity of preserving the digital form, strategies related to preservation, technologies and UNESCO guidelines for digital preservation of library documents.*

**Keywords:** Preservation; Digital Preservation

### **1. Preamble**

The digital world challenges our notion of presentation. For several decades, specialists have voiced concern about the preservation of the portion of our cultural heritage in e-form. Widespread transition of knowledge from print format to e-format has also given rise to the problem of its preservation in digital form. The major challenge rapid obsolescence of the hardware and software required to interpret and present digital documents has been widely discussed. Ensuring continued access to digital information necessarily involves transforming digital documents to run on current media, software, hardware and operating systems. This paper deals with the issues surrounding the preservation of digital information and highlights work to address them.

### **2. Concept and Definition of Digital Preservation**

#### **2.1 Concept**

Digital preservation is defined as the managed activities necessary:

1. For a long term maintenance of a bit stream (including metadata) sufficient to reproduce a suitable facsimile of the original document.
2. For the continued accessibility of the document contents through time and changing technology.

"Digital Preservation" or "digital archiving" essentially aims at taking steps to ensure the longevity of e-documents. It applies to documents that are either born digital and stored on CD-ROM, diskettes or to the products to analog-to-digital conversion, if long-term access is intended.

During preservation, questions of record context, structure, appearance, and behaviour must also be taken into account. There is a wide range of digital formats available and, to make matters more complicated, different digital objects have different preservation requirements. These can depend on the record is being preserved, how long it needs to be preserved, the context and history of the record, and its original format. It is not always necessary to preserve every aspect of a digital record, and thus research is underway to define the essential aspects of records and their authenticity requirements. In all cases, however, the record must be preserved so that it retains its integrity and is authentic and usable. This presents interesting challenges.

## 2.2 Definition

1. " Digital preservation is the ability to keep digital documents and files available for time periods that can transcend technological advances without concern for alteration or loss of readability."(The Association for Information and Image Management)
2. "Digital preservation refers to a series of managed activities necessary to ensure continued access to and preservation of digital materials."(RLG/OCLC Report) *(Both definitions were cited in Chapman, 2001)*

Preserving digital content entails far more than making backup copies and storing them in disparate location. Digital preservation is to extend the usable life of machine readable files and protect them from media failure, physical loss, and hardware and software obsolescence, the activities include are:

1. Ensuring the long term maintenance of bitstream (the zeros and ones):
  - Backing up files and keeps a copy at an offsite location.
  - Running checks to track the deterioration of storage media, files and bitstreams.
2. Providing continued accessibility of the contents:
  - Viability – making sure that information is readable from the storage media ;
  - Renderability – making sure that information is viewable by humans and able to be processed by computers.
3. Understandability – making sure that information is able to be interpreted by humans.(Hixson,2004)

## 3. Aspects of Digital Preservation

According to Graham (1997), digital preservation problems are associated with three distinct aspects:

1. Medium Preservation- the preservation of the physical media on which the bits and bytes of electronic information reside.
2. Technology Preservation - refreshing technologies from old to new as they become available.
3. Intellectual Preservation - addressing the integrity and authenticity of the information is originally recorded.

---

---

### **3.1 Medium Longevity Issues:**

- Fragile physical media.
- Prone to damage from careless handling.
- The need to copy data regularly.

### **3.2 Technology Preservation Issues**

- Technological Obsolescence.
- Hardware and Software dependence - obsolete technology may cause the dependent digital materials not accessible and useable.
- Digital information has been made available in a variety of different formats.

### **3.3 Intellectual Preservation Issues**

- The malleability and volatility of digital information.
- Accidentally damaged digital data, e.g.-data loss during transfer, accidents in updating, etc.
- Vulnerable to destruction of digital information through unauthorized tampering or deliberately modifying.

## **4. Digital Preservation Requirements**

In order to preserve digital materials on a scale commensurate with mass storage capabilities and in formats that are accessible and usable, it is necessary to articulate some basic requirements. There are two ways to examine digital preservation requirements from the perspective users' of digital information and other from custodians' perspective.

The potential uses of digital materials are varied, unpredictable, and almost endless. Any generalizations, even if restricted to one community of users such as humanities scholars, run the risk of overlooking and understanding potential user needs. Precise requirements for presentation and analytical tools vary among disciplines. The ability to establish the authenticity and integrity of a source is critical to users. Mechanisms that will enable users to establish authenticity require archives and libraries to store much more than the content of digital documents. Attributes such as formal document structures, metadata that document the maintenance and use history of the document, time and date stamps, and a series of references among documents are essential for determining authenticity and for understanding the provenance of sources and placing them in a larger context (Graham).

Michelson and Rothenberg (1992) argue that networking and access to digital sources will change all dimensions of the scholarly work process, including identifying sources, communicating with colleagues, interpreting and analyzing data, disseminating research findings and teaching.

Preserving digital materials in formats that are reliable and usable, however, will require long-term maintenance of structural characteristics, descriptive metadata, and display, computational, and

---

analytical capabilities that are very demanding of both mass storage and software for retrieval and interpretation.

Archives, libraries and other types of repositories that are struggling to meet escalating user expectations with limited financial and technical resources may express digital preservation requirements differently.

Storage systems should be capable of handling digital information in a wide variety of formats, including text, data, graphics, video, and sound. Conversion from analog to digital formats and migration to new generations of technology will be rapid, accurate, and inexpensive enough to permit very large-scale transfers of heterogeneous materials. Storage space requirements will be minimal and not demand highly sensitive environmental controls. To make digital preservation affordable to the widest possible range of organizations and individuals, equipment, media, and maintenance costs must be modest.

## **5. How Can Digital Information Be Preserved ?**

Digital preservation is both a technical and organizational challenge. In practical terms, the successful preservation of digital information is dependent upon organizations identifying and implementing suitable preservation strategies. Lee et al. (2002) have identified four distinct strategies, based on data migration, preserving technology, software emulation and encapsulation.

Data Migration - The migration of digital information refers to the "periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generations". Migration includes refreshing, that is, copying digital information without changing it.

### **5.1 Developing Migration Strategies**

- Less stable to more stable media
- Highly software dependent formats to less software intensive formats
- Multiplicity of formats to smaller number of common formats
- Development of backward compatibility paths
- Standard development and imposition
- Developments of process centers for migration and reformatting.

### **5.2 Technology Preservation**

This approach intends to retain the needed software and hardware. This approach "may have an important role for the recovery of data from obsolete storage media and platforms, but it is unlikely to become a viable long-term strategy." (Day, 2003,p.180)

This approach may end up with "a museum of ageing and incompatible computer hardware."(Feeney, 1999,p.42).

---

---

### 5.3 Emulation

Emulation refers to the process of mimicking, in software, a piece of hardware and software so that other processes think the original equipment / function is still available in its original form. Emulation is essentially a way of preserving the functionality of and access to digital information, which might otherwise be lost due to technological obsolescence.

One of the benefits of the emulation strategy is that the original data need not be altered in any way. It is the emulation of the computer environment that will change with time. This should help to maintain the integrity.

Another advantage of implementing emulation is its possible efficiency. Once the data is achieved with metadata and software, no other action is required apart from media refreshing until access is desired. One emulator can also be used as a solution for several data objects requiring the same operating environment.

- The emulator involves encapsulating three kinds of information with each digital document:
- Document itself, application software and operating system.
- The specification of an emulator for the document's original hardware platform.
- Explanatory material, labeling information, annotations, metadata about the document and documentation for the software and emulated hardware included in the encapsulation.

### 5.4 Encapsulation

It is a technique of grouping together a digital object and anything else necessary to provide access to that object. This technique aims to overcome the problems of technological obsolescence of file formats because the details of how to interpret the digital bits in the object can be part of the encapsulated information. Encapsulation can be achieved by using physical or logical structures called "containers" or "wrappers" to provide a relationship between all information components, such as the digital object and other supporting information such as persistent identifier, metadata, software specifications for emulation.

### 5.5 Assuming 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 affective practices in acquiring, describing and managing digital resources.

### 5.6 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 which are well defined for text (e.g.-ASCII), images (e.g.-JPEG, TIFF) and encoding documents (e.g.-SGML, HTML).The ISO has produced a reference model(CCS DS 650.0-W-4.0), for an Open Archival Information System(OAIS) which establishes the minimum requirements for a digital archive to ensure long-term preservation and provides a framework for describing and comparing archival architecture and operation.

## **6. UNESCO'S Guidelines -**

In the UNESCO'S reports, a number of recommendations are given:

- Not all digital materials need to be kept, only those that are judged to have ongoing value: these form of the digital heritage.
- Materials cannot be said to be preserved if access is lost. The purpose of preservation is to maintain the ability to present the essential elements of authentic digital materials.
- Digital preservation must address threats to all layers of the digital objects: physical, logical, conceptual, and essential.
- Authenticity is best protected by measures that integrity of data is not compromised and by documentation that maintains the clear identity of the material.

## **7. Documenting Resources**

### **7.1 Metadata**

- It is a description of a digital object.
- Metadata elements useful in preservation might include:
  - identifiers;
  - hardware, operating system and software required to access a document;
  - encoding standard and version;
  - migration history and its access;
  - authenticity determination; etc.
- Conversion projects would employ additional metadata elements such as:
  - capture device
  - resolution
  - compression
  - source material
  - producer

### **7.2 Identifier**

- It is a number, associated with a particular instance of a digital object.
- Establish the authenticity of the object by confirming to a user that the resource is accessing is the one cited.
- It could also help to establish the relationship between copies or versions of digital objects.

---

---

### 7.3 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. One way of linking metadata and the digital object is to package them together. Universal Preservation Format (UPF), a data file mechanism that uses a container structure to incorporate metadata into digital media objects.

### 8. Infrastructure for digital Archives -

Already in 1996, it was proposed to build a 'deep infrastructure capable of supporting a distributed system of digital archives'. Such a system would depend on trusted organizations capable of keeping materials alive for the long term and making them available to users as agreed with the depositor.

Digital archives need to be trusted organizations. Those who transfer materials for preservation have to be certain that integrity and authentic are ensured, that technical measures are taken in time, and that rights and restrictions for access will be observed.

### 9. Examples of Current Initials and Projects

#### Organizations/Cooperatives:

- Digital Library Federation (DLF) <http://www.diglib.org/dlfhomepage.html>
- Electronic Media Group (EMG) <http://aic.stanford.edu/sg/emg/>
- Electronic Resource Preservation and access Network (ERPANET) <http://www.erpnet.org/>

### 10. Conclusion

Today's digital media should be handled with care, but most likely will far outlast the capability of systems to retrieve and interpret the stored data. We can never know for certain when a system has become obsolete. Archives must prepare to migrate valuable digitized data, indexes, and software from one generation of computer technology to a subsequent generation. The use of digital technologies from a preservation perspective requires a deep and long standing institutional commitment to long term access, the full integration of the technology into information management procedures and processes, and significant leadership in developing standards for digital preservation.

#### References:

1. Graham, Peter S. -Long-Term Intellectual Preservation. -Collection Management. -Vol.22, no.3/4(1998). -P.81-98
2. Hedstrom, Margaret. -Digital Preservation: a time bomb for Digital Libraries. -(n.d.).
3. <http://www.diglib.org/preserve/criteria.htm>
4. <http://www.oclc.org/education/conferences/presentations/2001/preservation/chapman.htm>.