Digital Preservation with Special Reference to the Open Archival Information System (OAIS) Reference Model: An Overview

Sibsankar Jana **Mrinal Kanti Mondal** **Ujjal Marjit**

Abstract

Libraries are engaged in the creating and marinating digital libraries all over world. A digital library that deals with data those are born digital as well as those that have been digitized from their analogue form. So there is a need to preserve digital resources for future use. Since the nature of digital data or resource is intangible and it cannot see in necked eyes so there is a need to preserve and organized them in such a way that we can get easily access to them. Developing preservation process for digital resources will require the integration of new methods, policies, standards, and technologies. This paper aims to give an overview in digital preservation, highlights some strategy and standards efforts.

Keywords: Digital Preservation, Obsolescence, OAIS

1. Introduction

Digital preservation is the process for maintenance of digital resources over the long-term with a view to ensuring continued accessibility. Digital resource deals with any information processed by a computer and include both that which is "digitized" i.e. reformatted to digital as well as those resources that are "born digital". Digital preservation encompasses a number of organized tasks associated with a variety of technical approaches or strategies for ensuring the digital resources which are not only preserved properly, but also adequately maintained and thus consistent over time. It is based on the management of archive copies of deposited digital resources i.e. copies that is independent for online representation.

2. Digital Preservation: Definition

In general, digital preservation involves a wide rang of organized activities which is designed to expand the usable life of machine readable computer files i.e. digital objects and protect them from media failure, physical loss and obsolescence. Here some established definitions are enumerated:

i). According to the TDR (Trusted Digital Repository) it is the activities into those that promote the long-term maintenance of a bit stream (the 0s & 1s) and those that provide continued accessibility of its contents.

ii). The definition of digital preservation as proposed by the Research Libraries Group (RLG) as follows: Digital preservation is defined as the managed activities necessary: i) for the long–term maintenance of bit stream (including metadata) sufficient to reproduce a suitable facsimile of the original document and ii) for the continued accessibility of the document contents through time and changing technology.

iii). The OCLC/ RLG working Group on Preservation Metadata added the concept of viability to the maintenance of the bit stream, indicating that information must be intact and readable from storage media. They further subdivide the content accessibility need into renderability (viewable by humans and processable by computers) and understandability (interpretable by humans).

iv). A proposed definition of a reliable digital repository in the context of digital preservation also from RLG report is given below: "A reliable digital repository is one whose mission is to provide long-term access to managed digital resources; that accepts responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users; that designs its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it; that establishes methodologies for system evaluation that meet community expectations of trustworthiness; that can be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly; and whose policies, practices, and performance can be audited and measured."

From the above definition it is clear that digital preservation is:

a) The management of digital information resources over time.

b) The usable life of machine-readable computer files and protects them from technological obsolescence.

c) The long-term, error-free storage of digital resources, with means for retrieval and interpretation, for the entire time span that the digital information required.

3. Challenges for digital preservation

In this present situation information created, stored and accessed digitally is at risk for loss in two important ways. These two ways are Digital Obsolescence and Physical threats or damage.

3.1 Digital obsolescence

It is a situation where a digital resource is no longer readable because the physical media, the reader required to read the media, the hardware or the software and file format that runs on it, is no longer available. The rapid growth of different kinds of hardware, modes of digital encoding, operating systems, general or specialized software and file format ensures that digital obsolescence will become a problem in the future. Obsolescence can affect hardware, software and file format, storage media and even the arrangement of the data in stored files. So it is the key challenges in case of digital preservation.

Here some of the key points regarding digital obsolescence are given bellow:

3.1.1. Storage Media

It may be superseded by newer or denser versions of that medium.

3.1.2. Hardware and software

The hardware and software used to store and access digital resources that are constantly upgraded and superseded. Technology obsolescence is generally regarded as the greatest technical threat to ensuring continued access to digital resources.

3.1.3. File format and Software

File format and software both is closely couple to each other. Computer file is the main target of digital preservation. It is presented according to some pre-defined structural and organizational principles. Those principles generally refer to as file format. Each and every file format has its own format specification, which provides the details necessary to construct a valid file of a particular type to develop software applications that can decode and render such file. Now the question is most of the software upgraded on a regular basis but the files that have not been migrated may not be readable by the latest version of that software, and the older version of the software may no longer be available or may not run on a current computer system. Also we can say due to the complexity and dynamic nature of many file formats, it is extremely difficult to determine whether a file move from one format to another. So it is necessary to know before preserving digital resources in which file format is used. In this context different research initiatives are appearing and trying to define preservation-friendly standard formats as well as strategies for digital resources to be made available over time.

3.2. Physical threats

Like obsolescence, physical threats are a challenge for digital preservation. It may be define as the internal and external forces that can damage or destroy the readability of the digital resources. It includes materials instability i.e. internal storage not suitable for long-term preservation; improper storage environment (such as temperature, humidity, light, dust etc); over use, natural disaster; infrastructure failure; inadequate hardware maintenance; human error i.e. improper handling etc.

4. Digital preservation strategies or techniques

The strategy of digital preservation is a particular technical approach to the preservation of digital resources for marinating and accessing over the long-term. There are various preservation strategies or techniques are in operation, but no one is appropriate for all data types, situation or institutions. Some of these are given below:

4.1. Bit stream Copying

It is generally known as “backing up our data”. It refers to the technique of making an exact duplication of a digital

resource or object. Though it is an important component, yet it is not suitable for long-term maintenance of digital resource because it deals with the question of data loss due to hardware and media failure.

4.2. Refreshing

It means to copy digital resources from one long-term storage medium to another of the same type with no change in bit stream. For example copy from a older DVD to new DVD or from a older DAT tape to a new DAT tape. Although it is an important component of any successful digital preservation, but it is not itself a complete program. It helps both decay and obsolescence Issues related to storage media. Modified refreshing: it is the copying to another medium of a similar type, with on change in bit stream. For example 650 MB disk to 750 MB disk.

4.3. Analog backups

It means the conversion of digital resources or objects into analog form with the use of durable analog media e.g. HD- ROSETTA or the creation of silver halide microfilm from digital image.

4.4. Data migration

It involves the periodic transfer of digital resources form one technology to another or from one hardware or software configuration to another. It is a broader and richer concept than refreshing for identifying the range of option for digital preservation. It focuses on maintaining digital resources in current format. For example a repository that encoded files in SGML, which are migrating to XML since it has become the current standard. This can be time consuming and very expensive. The main aim of data migration is to preserve the integrity of digital resources and to retain the ability for users to retrieve, display and otherwise use them in the situation of constantly changing technology.

4.5. Technology preservation

It is based on preserving the technical environment i.e. original technology that runs the system which includes operating system, original application software, relevant hardware platform, media drive etc. This approach sometimes called "computer museum or museum style" and probably only suitable for short term solution.

4.6. Technology emulation

It seeks to preserve the usability and functionality of a digital project by using new technology to simulate the older technology worked. It requires the creation of emulators (special software) programs that translate the code and instruction from one computing environment, so that it can be properly executed in another. This is extremely complicated, as everything from operating system to scripting language must be emulated. Though it is complicated yet most the digital project choice it as an option.

5. Role of Metadata in digital preservation

Metadata plays an important role in digital preservation. For maintaining a history of digital resources with metadata is a key part of digital preservation.

- **Administrative metadata**, which refers to all information that is relevant for various aspects for digital preservation, including creation, migration, quality control, right information etc. Technical information about data capture, file formats, encoding, and provenance will be useful for digital preservation.
- **Structural metadata** that describes the architecture of digital resources and the relationship between files which helps to display and navigation purpose. It is more helpful for digital preservation.
- **Descriptive metadata**, which refers, to describe or identity of digital resources like cataloguing records, finding aids, specialized indexes, hyperlinked relationship between digital resources etc. These needs fostered the appearance of metadata standards such as Dublin Core Metadata Standard, Encoded Archival Description (EAD) etc which plays an important role in the digital preservation process.

The Metadata Encoding and Transmission Standard (METS) is an XML encoded metadata format developed by the Digital Library Federation and maintained by Library of Congress that includes all the above three types of metadata for identifying the digital resources and also supports the management and exchange of digital resources among.

Thus preservation metadata includes wide range of elements for a variety of management purposes, and

also show various levels of detail. Now the question is for any

particular strategy or technique for long-term preservation, the generation and maintenance of this kind of metadata is essential. So there is a need to integrate conceptually the major metadata sets into a common framework or model.

6. The Open Archival Information System Reference Model: OAIS

For preserving and maintaining the access to digital information over long-term, the Open Archival Information System (OAIS) Reference Model has been widely adopted as the foundation for many important digital preservation initiatives. This OAIS Reference Model published by the Consultative Committee on Space Data Systems (CCSDS, 2000) of the National Aeronautics and Space Administration (NASA). It became an ISO standard in 2003. The OAIS Reference Model is a conceptual framework for a generic archival system, which is committed to a dual role of preserving and providing access to information. The reference model includes an OAIS Functional Model that describes the functional components, which collectively fulfill the system's preservation and access activities. The Functional Model can be represented diagrammatically in Figure 1 (CCSDS, 2002).

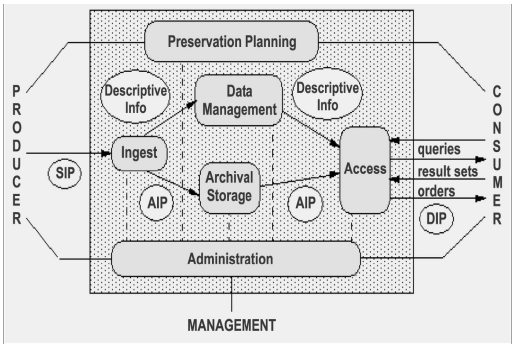

Figure 1: OAIS Functional Model

This model attempts to comprehensively identify the responsibilities and components of an archival system, which includes:

6.1 The people and Institutions

The role of people and institutions that interact in an archive, these appear in this model as**:**

- **Producers:** Producer is the role played by those persons, or client systems, which provide the information to be preserved.
- **Management:** Management is the role played by those who set overall OAIS policy as one component in a broader policy domain. In other words, Management control of the OAIS is only one of Management's responsibilities. Management is not involved in day-to-day archive operations.
- **Consumers:** Consumer is the role played by those persons, or client systems that interact with OAIS services to find and acquire preserved information of interest. A special class of Consumers is the Designated Community. The Designated Community is the set of Consumers who should be able to understand the preserved information.

6.2 Information Packages

The digital resources or objects that are managed by OAIS called Information Packages. The information packages of OAIS Model are the bundles of data objects and metadata about the objects that are the unit of deposit, storage, and distribution by an archive. The model allows transformation to be done as objects move from one type of package to another, which is illustrated in Figure-2. There are three types of Information Packages are in operation in OAIS System. These are:

- **Submission Information Package (SIP)**: An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.
- **Archival Information Package (AIP)**: An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.
- **Dissemination Information Package (DIP)**: The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS.

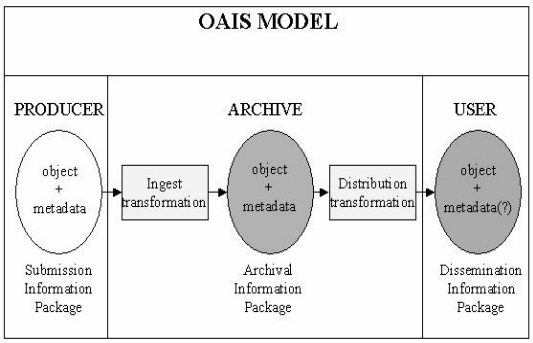

Figure 2: Information Packages in the OAIS Model

6.3 The major functions of OAIS:

The major functions of a fully functioning OAIS System are given below:

- Ingest: It provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. It depend upon the rules established by the organizational side to determine the metadata that must be present, the formats that are acceptable, and the quality checks that must be performed. After all the complex objects are created, the SIPs are moved to Archival Storage as AIPs. When changes are made to the objects in the future, the metadata generating functions will be used again to update the AIPs.
- Archival Storage: It is responsible for the storage, maintenance and retrieval of AIPs. Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfill orders.
- Data Management: It is used for populating, maintaining, and accessing both Descriptive Information, which identifies and documents archive holdings and administrative data used to manage the archive.
- Administration: It provides the services and functions for the overall operation of the archive system including soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software, system engineering functions to monitor and improve archive operations, maintaining archive standards and policies, providing customer support, etc.
- Preservation Planning: This function is responsible for monitoring the environment and evaluating archival content to recommend migrations, software prototyping, changes in archive standards

and policies, etc. Perhaps it is the core function of OAIS.

- Access: It provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the archive, and allowing Consumers to request and receive information products.

This is important to keep in mind that the OAIS Reference Model is a high level conceptual framework. Today this reference model has been used by a variety of organizations for their digital resource preservation aspects.

7. Conclusion

So far the very nature of digital information is concerned, preservation is necessary for long-term access to the digital resources in the archive. Unless and until we adopt the proper preservation strategy for any archive, the digital resources may be at great risk as per access is concerned. Therefore preservation is more vital one than collection, organization and dissemination of digital information in archive.

References

1. **Heslop, H., Davis, S., and Wilson, A**. National Archive Green Paper: An Approach to the Preservation of Digital Records. Available at http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper%.pdf (Accessed on 17-07-2008).

2. **Beagrie, Neil.** A Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) 2002-2005. Available at http://www.jisc.ac.uk/index.cfm?name=pres_continuing. (Accessed on 18-07-2008).

3. **Cornell Tutorial** (2003): The OAIS Reference Model, section 4B in Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems, Cornell University. Available at http://www.library.cornell.edu/iris/tutorial/dpm/foundation/oais/index.html (Accessed on 26-07-2008).

4. **DAY, Michael.** Integrating metadata schema registries with digital preservation systems to support interoperability. *2003 Dublin Core Conference*, Seattle, Washington, USA, 28 September - 2 October 2003. Available at http://www.ukoln.ac.uk/metadata/presentations/dc-2003/day/slides-draft.ppt. (Accessed on 05-08-2008).

5. **Digital Preservation Coalition** (2002), Preservation Management of Digital Materials: A Handbook, Digital Preservation Coalition. Available at http:// www.dpconline.org/graphics/handbook/. (Accessed on 10-08-2008).

6. **JONES, M. and BEAGRIE, N.** Preservation Management of Digital Materials: A Handbook, Digital Preservation Coalition. available at http://www.dpconline.org/graphics/intro/definitions.html. (Accessed on 12-08-2008).

7. **CCSDS-** Consultative Committee for Space Data Systems. Reference model for Open archival Information System (OAIS), CCSDS 650.0-B-1: Blue Book. Available at http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.o-B-1.pdf. (Accessed on 12-08-2008).

8. **RLG-OCLC.** Trusted Digital Repositories: Attributes and Responsibilities – An RLG-

OCLC Report. Available at http://www.rlg.org/legacy/longterm/repositories.pdf. (Accessed on 16-08-2008).

9. **LAVOIE, Brain.** The Open Archival Information System Reference Model: Introductory Guide, (DPC Technology Watch Series Report 04-01), OCLC and DPC, York. Available at http://www.dpconline.org/docs/lavoie_OAIS.pdf. (Accessed on 20-07-2008).

About Authors

Mr. Sibsankar Jana
Assistant Librarian, University of Kalyani, W.B.
E-mail: E-mail : sibs_jana@klyuniv.ac.in, sibs_jana@yahoo.com

Mr. Mrinal Kanti Mondal, Assistant Librarian I, Department of Film Studies, Jadavpur University, W.B. E-mail: mrinal_ju@yahoo.com

Mr. Ujjal Marjit, System-In-Charge, University of Kalyani, W.B.
E-mail: systemsincharge@klyuniv.ac.in