Digital Curation Strategies for Information Management in Higher Education Institutions

J Shivarama   **Sheela V**   **A B Deepak**   **K B Agadi**

Abstract

Libraries and archives have been the central institutions that not only house works of value but also focus their efforts for preservation of them for future reuse. The term preservation is thus not unheard of in the traditional library and archival environment where materials in paper, microform, photographic materials are preserved. Preservation involved an array of tools and methodologies developed to reduce the decay of traditional materials and to restore them. The preservation of traditional printed materials however has the advantage that the collections are relatively stable for at least decades if not centuries. The digital information, unfortunately affords no such luxury and are by their very nature ephemeral and fragile, with a lifespan that could be weeks or days if left to themselves. Thus, the digital preservation adds new sets of challenges for libraries and archives. Here we made an experiment on how to manage the information using the digital curation techniques in higher educational institutions.

Keywords: Digital Curation, Information Management, Digital Library, Open Source Software

1. Introduction

Libraries and archives have been the central institutions that not only house works of value but also focus their efforts for preservation of them for future reuse. The term preservation is thus, not unheard of in the traditional library and archival environment where materials in paper, microform, photographic materials are preserved. Preservation involved an array of tools and methodologies developed to reduce the decay of traditional materials and to restore them. Digital preservation involves "the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable" (Hedstrom, Margaret, 1998). Quite unlike the traditional materials the digital materials are "vulnerable to deterioration and catastrophic loss, and even in ideal conditions they are short lived relative to traditional storage media. They deteriorate rapidly, requiring decisions to be taken in very short time for actions to prevent loss in a matter of years, and not decades" (Hedstrom, Margaret, 1998).

2. Digital Preservation: - A Ticking Time Bomb for Digital Libraries.

Hedstrom talked about the problem of obsolescence in retrieval and playback technologies as a cause for concern in digital preservation. While she rightly gauged that the innovations in computer hardware, software, and software technology industries were able to provide faster processing capabilities and greater storage capacities at lower costs but the rate

at which the devices, processes, software meant for recording and storing information were becoming obsolete caused their own set of compatibility issues that could lead to a situation where the information created and stored through them may no longer be usable in future.

Hedstrom also opined that digital works which are created using new or emerging software applications are especially vulnerable to software obsolescence because standards for encoding, representation, retrieval, and other functions take time to develop. New formats also tend to be more complex because they can handle a wider variety of information representations and information processing functions. New multimedia applications have to handle multiple representations of text, data, images, sound, and motion as well as manage the relationships among the components of complex digital objects. Digital preservation is constrained by the absence of established standards, protocols, and proven methods for preserving digital information and by the tendency to consider preservation issues only at the end of the project or after sensational loss.

Hedstrom viewed that the mass storage capabilities and long-term preservation are two terms that embody a contradiction which represents "a time bomb" threatening the long-term viability of digital information. The technologies that are emerging for mass storage are growing, but they lag behind considerably to the growing store of digital information.

3. Digital Curation and Preservation in Indian Context

The digital curation and digital preservation movement have tremendous impact on institutions as well as libraries in India. The scenario in India is very interesting as efforts at both institutional level in their individual capacity as well as the ministries, private organisations, NGOs, also took initiatives towards the capacity building for information organisations and management.

The most notable efforts is the formation of Information and Library Network (INFLIBNET) Centre, an autonomous Inter-University Centre set up by the University Grants Commission. The INFLIBNET has through its various programmes like, the formation of e-journal consortia, web-site hosting, and creation of network resource centres in different universities and research organisations across India enable these organisations to come together to consolidate and share their resources in order to maximise their potential (Chandra & Patkar, 2008).

4. Open Source Software Solutions for Digital Curation

The open source movement has been through a sea change from being the pursued by hobbyists it has developed into a vibrant movement in itself with a dedicated set community of developers and innovators.

The most advantageous reason for using open source software is the flexibility and customization that the availability of source code provides. The software under open-source empowers and encourages, even 'rewards modification, re-use, redistribution and understanding (McHugh, 2005).' Andrew McHugh (2005) discusses that, importance of open-source software along with other reasons lie in their transparent nature. There is nothing hidden or financial agenda behind the software packages and therefore, this ascribes an element of honesty to them.

The Lifecycle approach for digital curation is given a new lease of life by the adoption of open source software. Many organisations working in the field of digital curation or simply facing the challenging tasks of digitizing the collection material have begun to realize the importance of employing open-source software solutions as this has benefits like 'the longevity; the ease of its re-creation or emulation; its adherence to and use of open standards; the level of legal freedom associated with its use; its associated costs; its ubiquity; its support for metadata; and its stability'(McHugh, 2005).

There are number of open-source software provides a range of tools which are goal oriented and have been customized by developers with specific roles for digital curation professionals.

5. Significant Role of End Users as Stakeholders in Digital Curation Policy

Digital preservation requires to look at the perspectives of the end users and libraries, archives, custodian organisations etc. The users' expectations and requirements may be diverse and libraries may find it difficult to accomplish all their preservation missions due to resource constraints, competing priorities, and lack of technical expertise. But by making preservation requirements explicit from the users and custodians perspectives, libraries and archives will be better able to balance competing demands and to integrate digital preservation into overall planning and resource allocation (Hedstrom, Margaret, 1998).

In a state of divergent need of the users what will be always difficult to fulfil, there will always be some elements that are of common concern to all stakeholders. The ability to establish the authenticity and integrity of a source is critical to all

users. Mechanisms that can enable users to establish the origin, provenance, and authenticity of digital documents require libraries and archives to preserve contextual and descriptive information in addition to the content of digital documents. Over the past decade interests in user study has become an integral part of digital library research with studies that include: usability, interface interaction/design, user interface and accessibility (Chowdhury, Gobinda, 2010).

It is increasingly being understood that digital library is a virtual workspace for scholarly interactions and it can be a valuable asset if such interactions whether in the form of user annotations of digital content, etc., are value additions on the particular digital object. The ultimate goal of any digital library and digital curation, preservation is to make the digital information accessible and usable for user community both present as well as future. However the challenges faced in doing so are considerable ranging from technological concerns to sociological but most importantly for the present paper are the ones related to the users and context.

According to the director of the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP), the estimated life span of a web site is only 44 days (Library of Congress, 2008). This dichotomy in the history of technology for information generation is ironic as with the rise of new media and storage formats there exists both opportunities as well as threats. Newer storage formats and newer generation of media present their own set of challenges with increased storage capacity comes faster processing which may be based on newer technologies that may no longer recognise the data created decades ago using previous software and hardware.

Headstrom (1998) is of the opinion that, it may be cost-effective to preserve digital information in cases where retaining the content is of greater importance than the display, indexing, and computational characteristics. This migration is much simpler for traditional materials where strategies could be designed with printing to archival paper that boasts of a lifespan of 100 to 500 years. However, the same results may not be possible or relevant to the digital information. Libraries and archives have diverse set of collections with larger concerns and few experiments have been conducted like converting digital records from the great multiplicity of formats into smaller easy to manage array of standard formats (Li, Ze-Nian & Drew, Mark S., 2004).

Another significant aspect for digital preservation is compression algorithms that are employed in compressing the data. There are two types of compression algorithms lossy and lossless. Lossless compression is preferred wherever possible but, if one has to use lossy compression it should be done with due care to ensure minimum or virtually no intended loss of data upon compression (Li, Ze-Nian & Drew, Mark S., 2004). The advantages of the above formats are the assurance of a standard protection of digital information to preserve the original bytes of information irrespective of future changes to technology. Another safeguard could be to check whether the software being used for content generation provides utilities for retrospective compatibility for using content created using previous technology and compatibility between complex software systems.

It is beyond the scope of this paper to gauge the compatibilities between different computer technologies but even, the most homogeneous collection holdings or digital libraries with similar set of collections would eventually find themselves at the stage where they will have to migrate to a newer technology. Therefore, the need of the hour is to foresee such possibilities in future and plan now, at present, involving the user community who are the content creators. The above statement can be explained further in the light of the issues that are faced during migration and reformatting of digital materials to simple standards. This is not always feasible as it eliminates the structure of documents and relationships embedded in databases. Computation capabilities, graphic display, indexing, and other features are often difficult to migrate and if attempted may be are lost to the extent of making the information unusable. One of the major reasons for more thoughts to be given at the user end for digital preservation lies at possible future challenges that the libraries can face due to new forms of electronically-enabled discourse, artistic and cultural expressions that may be difficult to be processed without adaptations.

6. Dynamism of Digital Archiving for Social Science Libraries

Needless to say, such preservation possibilities will be of huge benefit to archives, libraries, and other institutions to fulfil their preservation requirements if the information generated by systems that are designed to support long-term preservation needs, such as migration, adequate description, and linkages between the content of digital materials and their larger organisational or intellectual context (Hedstrom, Margaret, 1998). The catalyst for such demand for standardization could come from diverse sectors ranging from the ground level to the top managements of organisations that may realize the

benefits for long-term preservation. Advances in technology could also provide tools in hand for exchange, reuse of digital information at reduced costs that could further trigger libraries and archives to invest time and money in such initiatives. A major challenge faced by libraries and archives with respect to their digital collection is related to users and preservation of context of the information. The real challenge begins once the data and information is created either in the form of born digital form or through digitisation. The immediate concern is how to preserve this information for future use and access 'when the current technology that has been used to create and access the information will not be available any longer' (Chowdhury, Gobinda, 2010).

It may be noted here that digital preservation research has always focused on meeting one main objective: to make sure that the information can be used in future. But use by whom and in what context? In the digital age this is a major question. A given content (information) may have different types of potential users, each with a different characteristic, need and expectation, and the same content may be viewed and used by different types of users differently. Libraries have always played a key role in handling this sort of problem in the printed world by acting as an intermediary between content producers/providers and users, and adding value in the process based on an understanding of the user community as well as the context. For example, with knowledge of the university and its missions, etc. in case of an academic library, or nature and composition of the society vis-a'-vis the mission and targets of the government and the local community in case of a public library, and so on. How can this role be simulated in a digital library

environment, and how this can be passed on to the future generation of users through the preservation system, remain key questions that perplex digital curators (Chowdhury, Gobinda, 2010).

7. Digital Preservation and Policy Framework

Preservation policies are guidelines on how to manage digital assets in a repository in order to avert the risk of content loss. They define and specify, amongst other things, data storage requirements, preservation actions, roles and responsibilities. 'A preservation policy specifies digital preservation goals to ensure that:

❖ Digital content is within the physical control of the repository;

❖ Digital content can be uniquely and persistently identified and retrieved in the future;

❖ All information is available so that digital content can be understood by its designated user community;

❖ Significant characteristics of the digital assets are preserved even as data carriers or physical representations change; physical media are cared for; digital objects remain render able or executable;

❖ Digital objects remain whole and unimpaired and that it is clear how all the parts relate to each other; and digital objects are what they purport to be' (Dappert, Angela & Enders, Markus, 2010); and

❖ The digital preservation activities at The National Archives of UK (TNA) is based on two sets of activities: "passive preservation, which provides secure storage, and active preservation, which ensures the continued

accessibility of the stored records over time, and across changing technologies"(Brown, A., 2007, p. 5).

8.   Metadata Standards for Digital Archiving and Digital Curation Models

The preservation goals of institutional repositories depend on the availability of preservation metadata – in other words it is the set of information that describes the digital content in the repository to ensure the content has a longer lifespan and remain available for future reuse.

The specific metadata needed for long-term preservation falls into four categories based on basic preservation functional groupings:

- ❖ **Descriptive Metadata:** 'It describes the intellectual entity through properties such as author and title, and supports discovery and delivery of digital content. It may also provide an historic context, by, for example, specifying which print-based material was the original source for a digital derivative.

- ❖ **Structural Metadata:** Captures physical structural relationships, such as which image is embedded within which website, as well as logical structural relationships, such as which page follows which in a digitized book.

- ❖ **Technical Metadata:** It is required for physical files and includes technical information that applies to any file type, such as information about the software and hardware on which the digital object can be rendered or executed, or checksums and digital signatures to ensure fixity and authenticity. It also includes content type-specific technical information, such as

image width for an image or elapsed time for an audio file (Dappert, Angela & Enders, Markus, 2010).

- ❖ **Administrative Metadata:** Includes provenance information of who has cared for the digital object and what preservation actions have been performed on it, as well as rights and permission information that specifies, for example, access to the digital object, including which preservation actions are permissible' (Dappert, Angela & Enders, Markus, 2010).

9.   The Open Archival Information System (OAIS)

The OAIS reference model defines a framework with a common vocabulary and provides a functional and information model for the preservation community, it does not define which specific metadata should be collected or how it should be implemented in order to support preservation goals. The OAIS has defined the field of digital preservation systems. It identifies the participants, describes their roles and responsibilities, and classifies the types of information they exchange. One of the systems based on OAIS is LOCKSS system (Lots of Copies Keep Stuff Safe), developed by Stanford University Libraries. It works closely with publishers and enables participating libraries to own copies of the e-journals that they subscribe to as well as content published in open-access e-journals.

10. Digital Curation Centre-The DCC Lifecycle Model

The Draft DCC Curation Lifecycle Model was first published in the International Journal of Digital Curation in December 2007. The DCC Curation Lifecycle Model (Fig. 1) provides a graphical high-level visualization of the stages of information's life

as required for successful curation. The model presents a generalized depiction and is not specific to diverse information requirements or case. The remarkable feature of the model is its employability as a planning tool at (Table. 1) organisational level where it is flexible enough to be adapted to planning, designing and allocating curation and preservation activities at different levels of granularity.

to be taken or discarded as the case may be in order to achieve the results. To begin with, the model deals with the definition of digital data that is inclusive of digital objects as well as databases. Data that is information in binary digital form and could range
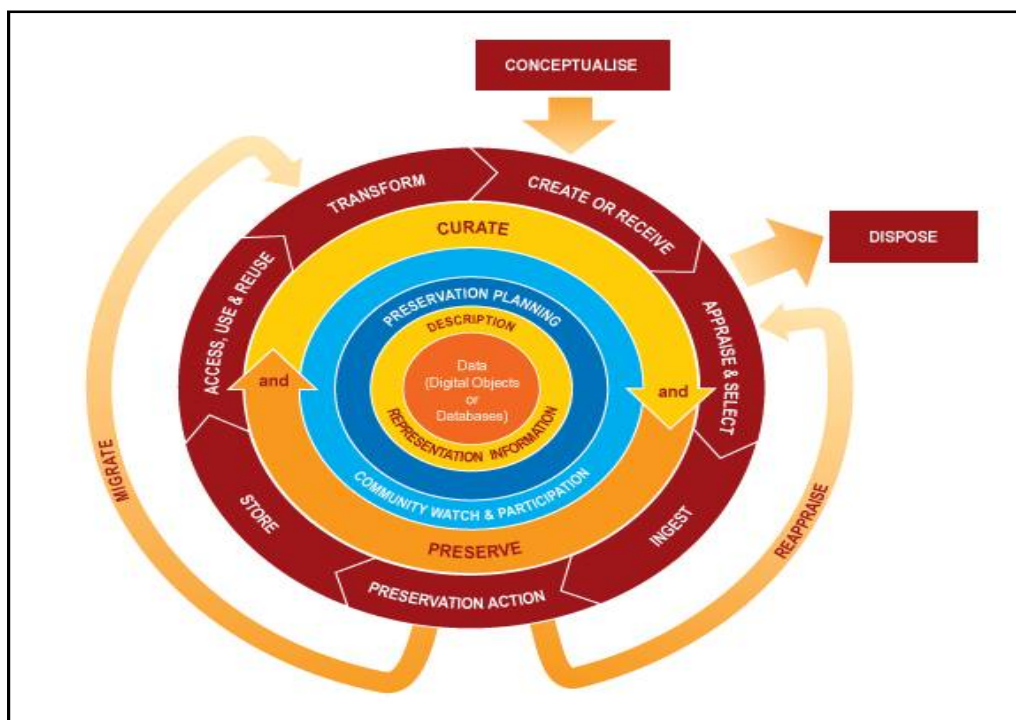

Figure: 1:DCC Digital Curation Lifecycle Model

The model (Fig. 1) acts as an effective guideline that assists in planning activities in an organisations or group of organisations collaborating together, that ensures a set of sequence of steps that best preserves the digital information. It assists in defining roles and responsibilities; and sets a framework of standards and technologies that needs to be adopted, suggests steps that may be required

from texts, images, audio, and their related metadata, to structured collection of records or data stored in a computer system or server. The model delineates a series of actions that are as follows:

Table 1: DCC Digital Curation Lifecycle Actions

Sequential Actions	Description of Actions
Conceptualise	Conceive and plan the creation of data, including capture method and storage options.
Create and Receive	Create data (administrative, descriptive, structural and technical metadata.) - Preservation metadata may also be added at the time of creation.- Receive data, in accordance with documented collecting policies, from data creators, other archives, repositories or data centres, and if required assign appropriate metadata.
Appraise and Select	Evaluate data and select for long-term curation and preservation. Adhere to documented guidance, policies or legal requirements.
Ingest	Transfer data to an archive, repository, data centre or other custodian. Adhere to documented guidance, policies or legal requirements.
Preservation Action	Undertake actions to ensure long-term preservation and retention of the authoritative nature of data. Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its integrity. Actions include data cleaning, validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures or file formats.
Store	Store the data in a secure manner adhering to relevant standards.
Access, Use and Reuse	Ensure that data is accessible to both designated users and re-users, on a day-to-day basis. This may be in the form of publicly available published information. Robust access controls and authentication procedures may be applicable.
Transform	Create new data from the original, for example ❖ 'By migration into a different format. ❖ By creating a subset, by selection or query, to create newly derived results, perhaps for publication.
Occasional Actions	
Dispose	Dispose of data, which has not been selected for long-term curation and preservation in accordance with documented policies, guidance or legal requirements. Typically data may be transferred to another archive, repository, data centre or other custodian. In some instances data are destroyed. The data's nature may, for legal reasons, necessitate secure destruction.

Reappraise	Return data which fails validation procedures for further appraisal and reselection.
Migrate	Migrate data to a new format. This may be done to accord with the storage environment or to ensure the data's immunity from hardware or software obsolescence.

Source: (Higgins, Sarah, 2008).

The lifecycle model is an attempt at identifying all the stages of life for a digital material while it stays in the collection of a repository, archive, library. This identification is very important in order to set standards of actions that are required at each stage in order to ensure the digital materials last longer and be available for reuse in future.

11. Digital Curation – Best Pract

Best practices are by no means laws or governing principles. They simply imply that best and most desirable results were found while engaging in certain actions when they were done in a certain way. Taking this view of the best practices certain observations are mentioned here that could be applied while framing the organisations digital curation policies.

12. Conclusion

Our society today is data driven, whichever walks of life whether it is education, science and technology, conduct of business, communication, etc., it is digital information which lies at the heart of it all. The rising significance of digital material has also brought the realisation that it needs to be carefully managed in order to ensure its survival and continued access for future reuse. The initial efforts towards digital preservation were more focused to preserve the material so that it survives technological obsolescence and organisationsal mismanagement. Having said that Ambacher B.I. et. Al (Ambacher, B.I. et al., 2013) discusses the technical, societal, and conceptual challenges faced by digital curation in public, private and not-for-profit sectors at both national and international level. It is to be noted that the digital curation is a relatively new field and the implications of its requirement are still being realized however, the rapid growth of electronic information and the need to actively manage this information is recognized in diverse communities (Ambacher, B.I. et al., 2013).

The Article presents the international standard of digital curation brought forward by the Digitial Curation Centre and their Digital Curation Lifecycle model which acts as a directional guideline for institutions to enumerate, follow, and adapt to their own requirements and challenges.

References

1. American Council of Learned Societies. (2006). Our Cultural Commonwealth: The Report of the ACLS Commission on Cyberinfrastructure for the Humanities and Social Sciences pp. 1– 51). USA. Retrieved from https://www.acls.org/ c y b e r i n f r a s t r u c t u r e / OurCulturalCommonwealth.pdf

2. Beagrie, Neil. (2006). Digital Curation for Science, Digital Libraries, and Individuals. The International Journal of Digital Curation, 1(1), 3–16. Retrieved from www.ijdc.net/index.php/ ijdc/article/download/6/2

3. Beagrie, Neil, Semple, N., Williams, N. P., & Wright, R. (2008). Digital reservationPolicies Study Part 1, final report. Retrieved from www.jisc.ac.uk/media/documents/programmes/

preservation/jiscpolicy_p1finalreport.pdf

4. Bearman, D., & Sochats, K. (1996). Metadata requirements for evidence. Retrieved January 31,2016, from http://www.archimuse.com/papers/nhprc/BACartic.html

5. Brown, A. (2007). Developing practical approaches to active preservation. The Internationa Journal of Digital Curation, 2(1), 3–11. http://doi.org/http://dx.doi.org/10.2218/ijdc.v2i1.10

6. Chowdhury, Gobinda. (2010). From digital libraries to digital preservation research: the importance of users and context. Journal of Documentation, 55(2), 207–223. http://doi.org/http://dx.doi.org/10.1108/00220411011023625

7. Dappert, Angela, & Enders, Markus. (2010). Digital Preservation Metadata Standards. Information Standards Quarterly, 22(2). Retrieved from https://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf

8. Graham, Peter S. (1995). Requirements for the Digital Research Library. College and University Research Libraries Group, 56(4), 331–339.

9. Harvey, Ross. (2010). Digital Curation: A How-to-Do-It Manual. London: Neal-Schuman Publishers, Inc.

10. Hedstrom, Margaret. (1998). Digital Preservation: A Time Bomb for Digital Libraries. Computers and the Humanities, 31(3), 189–202. Retrieved from http://www.jstor.org/stable/30200423

11. Higgins, Sarah. (2008). The DCC Curation Lifecycle Model. The International Journal of Digital Curation, 3(1). Retrieved from http://www.ijdc.net/index.php/ijdc/article/view/69/48

12. Innocenti, P, McHugh, A., & Ross, S. (2008). Tackling the risk challenge: DRAMBORA (Digital Repository Audit Method Based on Risk Assessment). Retrieved January 31, 2016, from http://eprints.gla.ac.uk/45946/

13. In Wikipedia, The Free Encylopedia. (2016a, January 26). ASCII. Retrieved January 31, 2016, from https://en.wikipedia.org/w/index.php?title=ASCII&oldid=701696841

14. In Wikipedia, The Free Encylopedia. (2016b, January 27). UTF-8. Retrieved January 31, 2016, from https://en.wikipedia.org/w/index.php?title=UTF-8&oldid=701857706.

15. Kondayen, Kate. (2015). Saving the digital record. Retrieved June 2, 2010, from http://news.harvard.edu/gazette/story/2015/05/saving-the-digital-record/

16. Library of Congress. (2008). Digital preservation: national digital information infrastructure & preservation program: a collaboration initiative of the Library of Congress. Retrieved from www.digitalpreservation.gov/.

About Authors

Dr. J. Shivarama, TISS, Mumbai
Email: shivaramatoo32@gmail.com

Ms. Sheela.V, KSOU,
Email: she579@gmail.com

Mr. A . B Deepak. TISS, Mumbai

Dr. K B Agadi, Assistant Librarian, CUG, Gandhinagar, Gujarat