DESIDOC Journal of Library & Information Technology, Vol. 40, No. 5, Sept 2020, pp. 321-325, DOI : 10.14429/djlit.40.5.15822

Sustainable Digital Preservation and Access of Heritage Knowledge in India: A Review

Akhlak Ahmad and Swadesh Sharma

School of Humanities and Social Sciences, G.D. Goenka Univeristy, Gurgaon - 122 103, India
Email: akhlaqa@yahoo.com

ABSTRACT

India is an ancient country and its literature in the form of manuscripts is as rich as its culture. These heritage resources need to be documented and preserved by digitizing and securing them for future generations and also to make them accessible to the larger audience. This paper reviews scholarly literature and analyses it to understand the application of standards for digitisation, access, and preservation in the digital process by Indian institutions. Further, it explores the feasibility of establishing of national standards for sustainable digital preservation and retrieval of heritage resources across the country.

Keywords: Digital preservation of heritage resources; Digitisation standards; Metadata standards.

1. INTRODUCTION

Heritage knowledge is the legacy that is inherited from the past; it is represented through oral history and written communication in the form of songs, poems, paintings, stories, philosophical ideas, etc.[1] These heritage resources were at times documented and recorded on stones, clay tablets, woods, palm leaves, papers, and also on various metals, etc. It is the responsibility of the cultural society to preserve its history and culture and make these resources accessible rests with society. Galleries, libraries, archives, and museums (GLAMs) are collecting, selecting, and managing heritage resources[2]. The current study focuses on heritage resources such as manuscripts, rare books, and other material housed by selected Indian libraries and archives.

India, being one of the oldest civilisations, has a vast reservoir of ancient heritage knowledge resources written in various languages and scripts, which have been passed down for generations. Serious and effective initiatives have been taken to collect, manage, and archive these heritage knowledge resources which are scattered across the country. India also has one of the largest and oldest collections of manuscripts in the world. These heritage resources are available across the country in heritage institutions and personal collections[3]. It is estimated that the country has a collection of over 5 million which have been catalogued so far, and over 60,000 are available in European countries. Additionally, over 1,50,000 manuscripts are available in South Asia and other Asian countries[4].

These valuable heritage resources are spread across the country and most of them are not in a good condition and face a threat of extinction in the absence of proper archiving. For that, they need to be identified, documented, preserved, and made accessible to the research community. But to make these heritage resources accessible has become the biggest challenge for organisations while undertaking collection, preservation, and dissemination of these information resources. Every single usage or manual handing makes these rare and vulnerable resources more fragile[5]. Therefore preservation and making them accessible are the main challenges faced by the archival institutions. To achieve these objectives, they have to preserve not only the analog information resources but also digitise them and preserve them, along with the digitally born documents. Digitisation and digital preservation are different processes that require diverse methods to be followed for the preservation[6].

Received : 18 May 2020, Revised : 20 September 2017
Accepted : 28 September 2017, Online published : 04 November 2020

1.2 Background and Objectives

This objective of this study is to understand the current practices of the digitisation of heritage resources in India with existing available knowledge. It analyses quality literature that provide a foundation to the research question i.e. sustainable digital preservation of Indian heritage resources. A number of online databases were queried that provided the access of a wide range of information resources but not limited to journal articles, conference proceedings, theses, monographs and books etc.

The review starts with the identification and introduction of heritage resources and then looks at initiatives taken for their digital preservation using established standards and protocols. It reviews the digitisation projects carried out by the Indian heritage organisation to highlight the success and also challenges faced by the information professionals while adopting new information communication technologies (ICTs) as they change rapidly.

Further, it attempts to answer questions why standardisation and uniformity in digitisation process are necessary and what are the existing standards and protocols available for digital

preservation and adopting best practices for the sustainability of resources. The last segment of this endeavor includes determining whether any important gaps in the literature still exist and if so, to highlight those areas so that appropriate future work can be undertaken

2. DIGITAL PRESERVATION

Digital preservation can be defined as the conversion of analog information resources through the digitisation process and also the inclusion of the born-digital resources available in the various digital files formats containing the various types of information. According to Digital Preservation Coalition, 2008, digital preservation is "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary"[7].

But one of the main challenges with the digital preservation is that the digital resources may not be readable and usable in the future as digital resources are more complex as compared to the analog. The retrieval of these digital documents depends on storage devices and hardware and software technologies. But there is a greater element of risk involved in a situation where these technologies become obsolete or incompatible as they become inaccessible or unreadable given the available hardware and software. The fragility of digital media can lead to the corruption of data; also poses a challenge to preserve their original and integrity as digital documents and the related metadata may easily be modified or tempered[2].

2.1 Digital Preservation Methods

Uwe M. Borghoff has suggested three methods to preserve digital data. Firstly, a migration which is the most common method is used for digital preservation; this can be simply done by transferring digital data from one to another new storage medium or by translating digital data from one format to another. Secondly, emulation preserves the digital data as it is. In this process, a computer environment is created in which different computer systems act in similar pattern and digital information resources are originated by emulation. Thirdly, the markup languages method is used to describe suitable metadata of digital document. The markup languages use tags to define elements within digital documents such as defining text, graphics, audios, and videos[8].

2.2 Sustainable Digital Preservation and Retrieval of Heritage Knowledge

Sustainability in digitisation broadly deals with the digital preservation of digital images and electronic resources constantly, until it continue accessibility and ensure the conservation of metadata[9]. While data is searched across computer systems, interoperability facilitates search and harvesting of metadata from the repository system. These are very important elements in creating digital library systems to make resources accessible across the globe[10].

Sustainable digitisation, long-term preservation, and retrieval largely depend on the application of standards and protocols in the digital process. When preserving and exchanging knowledge resources, these standards play an important role to make metadata interoperable and help in assisting unwanted variations in the metadata[11]. Also, when an Internet user retrieves these records in a standard format it allows to link to these records and harvest these resources in their database and offering various indexing and abstracting library services. Therefore sustainable digitisation and interoperability for data preservation and exchange should be achieved only by applying established standards and protocols[12].

3. DIGITISATION OF CULTURAL HERITAGE KNOWLEDGE IN INDIA

Various attempts were taken by the heritage institutions in India such as Indira Gandhi National Centre for the Arts (IGNCA), the National Archives of India, and the National Mission for Manuscripts to digitise its cultural heritage resources and integrated them in digital collection to start digital information services to cater to larger groups of users. The IGNCA is a premier institution in the area of the arts and culture has a large collection of cultural heritage resources in a wide range of formats, including manuscripts, microfilms, micro-fiches, negatives, photographs, audio-visuals, etc. IGNCA has adopted various methods of preservation, tools, and techniques for the digitisation of its heritage resources. It is estimated that IGNCA has digitised over 46 per cent of its heritage collection which has over 5,50,000 in the form of books, periodicals, manuscripts, microfilms, images, audio, and video recordings[13-14]. The UNESCO, IFLA, UNC guidelines for digitisation projects, standards, and protocols related to digitisation, metadata preservation, multilingual open standards for fonts such as Unicode have been followed while digitizing heritage resources[15].

The 'National Mission for Manuscripts (NAMAMI),' established by the Ministry of Culture, Government of India, aims to discover the manuscripts on arts and culture scattered across the country and digitise and make them accessible by creating a resource base of Indian manuscripts. It has "Manuscript Resource Centre" (MRCs) and manuscript collections across four regions of India. The digitisation of manuscripts has been undertaken to maintain several documents and total pages along with a zone-wise list of Manuscript Resource Centres (MRCs)[16].

The author recommends GLAMs need to explore the crowdsourcing potential in the process of digitisation especially converting contents from one form to another such as handwritten documents, transcribing text, or notation of audio and artifacts description[17].

Khuda Bakhsh Oriental Public Library has a collection of about 21,000 oriental manuscripts and 2.5 lakh printed books. It is perhaps well on its way to becoming the country's first library to computerise its handwritten collection for universal dissemination through the Internet. A pilot project of digitisation of 10,00,000 pages of manuscripts has been started and 1,214 manuscripts covering 3,57,915 folios have been completed for use[18].

National Library of India, Kolkata, National Archives of India, Nehru Memorial Museum and Library (NMML), Rampur Raza Library are prominent heritage national importance institutions, are the stakeholder of India's heritage collection of manuscripts. These institutions have also started the

digitisation process of manuscripts and other rare collections to preserve and provide access to their users or clients[13]. Some regional institutions started digitisation to preserve records from damage, larceny, and decay of manuscript such as Panjab University, Chandigarh has started a digitisation project to digitise manuscripts and rare literature at the Punjab level[19].

The initiatives to preserve and conserve heritage knowledge through digitisation may differ in the selection of material, adaptation of technologies, applying standards and tools all over the country. Possibly, heritage organisations working in India are applying different methods, techniques, standards of digitisation to preserve the heritage resources from deterioration. It will be more challenging while the technology for digital preservation is changing rapidly unless comprehensive digital policies at national and institutional levels are formulated. Digital technological platforms should be integrated for digital preservation and physical conservation of manuscripts and other rare materials. A well equipped infrastructure, regular funding, new digital technology, well-trained manpower, and applying uniform digitisation standards are the fundamental requirements to achieve sustainable digital preservation of our heritage knowledge.

Therefore uniformity in standards and protocols are essential qualifications to meet the above objectives. There are standards and protocols related to digitisation and file formats, preservation of metadata, and information retrieval. These standards deal with the technical specifications for image quality parameters of master and access files requisite standard metadata enhances appropriate access, and facilitate discovery and interoperability by the digital information system; and overall quality control of digital resources and associated data[20].

3.1 Standards for Digital Heritage Resources

A number of digitisation of cultural heritage projects started funded in India and other parts of the world, in the past few decades various initiatives have been taken to develop standards and unify them to optimise the outcome, and achieve high quality of digital resources within and outside the country. In India, National Mission for Manuscripts (NAMAMI) has designed guidelines for digitisation of archival material and metadata scheme in 2003, the Centre for Development of Advanced Computing (C-DAC) has developed EGov-PID preservation metadata standard in 2013 for government e-records.

While, European Union (EU) has developed Europeana Data Model (EDM) to manage large-scale metadata repository and aggregation service for all kinds of cultural heritage information from Europe.

3.2 National Mission of Manuscripts

NAMAMI aims to preserve and provide access to the manuscripts wealth of India in a big way. To deal with the born-digital and digitisation of heritage resources, Namami has prepared a guideline and standard of digitisation for creating digital archives of heritage resources including manuscripts dealing with digitisation, file formats, metadata, and retrieval of digital resources.

The metadata standard called Manus e-Granthavali is based on the Dublin Core standard having 24 descriptive and 16 technical fields. The guideline deals with creating a digital image of manuscripts, printed books, maps, photographs, slides, negatives, microfilms, etc. specifying image quality, file formats, storage, and access with the objective of long-term usage and reduce the need to rescan material. Further, the guideline deals with the file naming of the digital resources, organisation, and storage of images[21].

3.3 The Standard for Preservation, Information, Documentation (Egov-PID) of Electronic Records

"The Centre for Development of Advanced Computing (C-DAC) has been pursuing pioneering research in language technology and heritage computing. It is also involved in standardisation and representation of heritage scripts such as Grantha, Vedic, Samavedic, modi, etc". It has designed and developed Egov-PID preservation metadata standard in 2013 having basic elements defined by Dublin Core Metadata Initiative (DCMI) with 22 descriptive fields to describe government records.

To make government digital records readable which may be inaccessible due to obsolescence of storage media, software, format, etc. Digital preservation standard and guidelines are developed to assist country digitisation projects and based on open information archival system (OAIS) Reference Model which will preserve digital records as specified in the ISO 14721[22].

3.4 Europeana Data Model

The European Union (EU) has funded various digital culture research and development projects. The Europeana was started in 2005 to make European information resources easier to use in an online environment and it was hosted in 2008 at Europeana.eu. The main function of Europeana standard is to facilitate of access of resources of various heritage institutions in Europe. These heritage institutions index their contents following different metadata schemas. To make them searchable from a single window, these different metadata schemes were mapped into a single standard[23].

Europeana used elements from a combination of Dublin Core, IFLA Bibliographical Records Standard FRBROO, and Conceptual Reference Module (CIDOC-CRM). Europeana Data Model (EDM) is based on Dublin Core (DC) and Metadata Encoding and Transmission Standard (METS) elements was developed a metadata standard to provide its users more and better information retrieval. The main advantage of the EDM is that it maps the contents of various organisations such as archives, libraries, museums, etc. engaging in collecting and managing heritage resources, it makes possible interoperability between these organisations to meet the common information requirement of users[2].

4. THE MAPPING BETWEEN METADATA STANDARDS

Digital archival systems use different kinds of metadata standards to meet their information management requirements.

Table 1. Metadata standard crosswalk - This (Getty) crosswalk is given as example of selective fields to map the metadata for few metadata standard schemas.

CDWA	MARC	MODS	Dublin Core	EAD
Title Text (core)	24Xa Title and Title - Related Information	<title>	Title	<titleproper> (in <eadheader>) <unittitle>(in<archdesc>)
Creator Description (core)	1xx Main Entry 7xx Added Entry	<name>	Creator	<author> (in <eadheader>) <name>,<origination>, <persname>, <origination>, <corpname>, <origination> < famname> (in <archdesc>)
Edition Description	250 Edition Statement	<edition>		<revisiondesc> (in <eadheader>)
Subject Display	520 Summary, etc.		Description.Abstract	<abstract> <scopecontent> (in <archdesc>)
Copyright Statement	540a Terms Governing Use and Reproduction	Rights	4.1 Conditions Governing Access	

https://www.getty.edu/

The future consequences of these variations in describing and managing digital heritage resources followed by the heritage institutions may face a challenge for consistency, migration, preservation of metadata, and retrieval of digital resources. Presently, there is no mechanism is available which support to fix a large digitisation projects carried by heritage institutions.

Heritage institutions within the country may follow different metadata formats/ schemas. To enhance the interoperability between metadata applications, mappings between metadata standards can be done at the scheme level as well as record level. Each institution can use the existing crosswalk to map and transform data directly to an intermediary schema or one of the standard metadata schemas[24].

Crosswalk table enables mapping of a relationship between elements of different metadata schemes. It is used when multiple digital systems following different metadata schemes are interfaced to get certain schema based on national and international standards. For example, an institution maintains a particular metadata standards crosswalk mapping the connection between several commonly used standards such as Dublin Core, the Categories for the Description of Works of Art (CDWA), The Encoded Archival Description (EAD), the Metadata Object Description Schema (MODS), and MARC, among others[25].

Table 1 illustrates the relationship of classes between different metadata standards. For example, a book title classes (24Xa Title) of MARC standard from Library X can be correspond to the digitized document in Dublin Core (Title). Further, the description of the documents can be mapped to numerous instances of the class of a digital object.

By following the crosswalk table, heritage institutions in India could create an appropriate digital archival system for common information system after studying the digitization mechanism followed by major Indian heritage institutions. This type of information system can solve the problem of interoperability and can achieve some elements of the digital sustainability.

5. CONCLUSIONS

Every heritage organization has its own goals to achieve through digitization and have adopted different methods and practices for digital preservation. Therefore it is the urgent requirement to develop a digitization policy at the national level and also draw up standards for digitization and metadata preservation to improve the digital transformation of its heritage and cultural resources. Using standards for digitization, access and preservation make it possible for organizations to migrate data from one generation to another generation, together with compliance with new hardware and software, to preserve resources for future generations.

Based on the above exercise, it is argued that most of the case studies conducted related to digitization of heritage resources are institution-specific. No comparative study of the digitization process and practices followed among the heritage organizations have not been found so far. Therefore, it is felt that applications of standards for digitization, metadata preservation, and discovery of heritage resources need to be undertaken institutionally and collectively. Further, it needs to understand the challenges and benefits of applications of established standards and tools in the practice among Indian organizations.

Further, a sample crosswalk table was presented for various standards to understand the relationship between elements of different metadata schemas. It is suggested that a comprehensive study of heritage organizations digitization projects should be conducted and through a detailed crosswalk table and implemented to ensure the sustainable digital preservation and accessibility of heritage resources for the coming years. Also, we expect that this could be an experiment for the development of Indian national standards for the sustainable preservation of digital heritage assets such as “Europeana for European Union”.

REFERENCES

1. Graham, B. Heritage as knowledge: capital or culture?. *Urban Studies*, 2002, **39**(5-6), 1003-1017.
2. Bountori, L. Archives in the digital age: Standards, policies and tools. Elsevier, Cambridge, 2017.
3. Tripathi, D.S. Guidelines for digitisation of archival material. National Mission for Manuscripts, New Delhi,

4. Kumar, S. & Shah, L. Digital preservation of manuscripts: A case study. *In* 2nd Convention Planner-2004, Imphal, 2004, 27- 28.
5. Arora, J. Digital preservation: An overview. Bridging the Digital Divide, 2009, 29-30.
6. Traczyk, T.; Ogryczak, W.; Pałka, P. & Śliwiński, T. Digital preservation: Putting it work, Springer, New York, 2017, (Studies in Computational Intelligence).
7. Digital Preservation Coalition 2008, www.dpconline.org/docs/digital-preservation-handbook, (accessed on 20 April, 2020)
8. Borghoff, U.M.; Rödig, P.; Schmitz, L. & Scheffczyk, J. Long-term preservation of digital documents: Principles and practices, Springer, New York, 2006
9. Calzolari, M.C. Digital sustainability. *Technology*. December 2016. https://www.infrastructure-channel.com/technology/digital-sustainability/ (accessed on 15 January, 2019)
10. Mukhopadhyay, P. Interoperability and retrieval. United Nations Educational, Scientific and Cultural Organisation, Paris, 2015.
11. Pfister, R. The role of metadata standards in EOSDIS search and retrieval applications.NASA/ Goddard Space Flight Center, Greenbelt, (n.d.)
12. Kong, Qinzheng & Gottschalk, John. Information retrieval standards and applications. CiTR, The university of Queensland, Brisbane, (n.d.)
13. Khan, Nida & P.M., Naushal Ali. Digital preservation of manuscripts: Initiative in India. *J. Indian Libr. Assoc.*, 2017, January-March, 25-30
14. Bakhshi , Samar I. Digitisation and digital preservation of cultural heritage in India with special reference to IGNCA, New Delhi. *Asian J. Inf. Sci. Technol.*, 2016, **6**(2), 1-7
15. Gaur, Ramesh & Chakraborty, Mrinmoy. Preservation and access to Indian manuscripts : A knowledge base of Indian cultural heritage resources for academic libraries. *In* ICAL 2009 – Vision and Roles of the Future Academic Libraries, 2009, 90-98.
16. Sahoo, J. & Mohanty, B. Digitisation of Indian manuscripts heritage: Role of the National Mission for Manuscripts. *Int. Fed. Libr. Assoc. Inst.*, 2015, **41**(3), 237–250. doi: 10.1177/0340035215601447.
17. Singh, A. Digital preservation of cultural heritage resources and manuscripts: An Indian Government initiative. *IInt. Fed. Libr. Assoc. Inst.*, 2012, **38**(4), 289–296. doi: 10.1177/0340035212463139.
18. Khuda Bakhsh Oriental Public Library http://kblibrary.bih.nic.in/ (accessed on 20 April, 2020).
19. Harpreet Kaur. Digital preservation of manuscripts: An Indian perspective with special reference to Punjab. *In* 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services *(ETTLIS),* 6-8 January 2015, Noida , 271-274.
20. Rieger, O.Y. Preservation in the age of large-scale digitisation: A white paper. https://www.clir.org/pubs/reports/pub141/ (accessed on 25 April, 2020)
21. National Mission for Manuscripts. www.namami.gov.in (accessed on 10 January, 2020).
22. C-DAC Brochure. www.cdac.in (accessed on 12 April, 2020).
23. Europeana Pro. https://pro.europeana.eu/ (accessed on 12 April, 2020).
24. Cambridge, Abby. Building a digital repository program with limited resources. Springer, New York, 2020
25. Getty Resources on Art and Research Collections. http://www.getty.edu/resources/ (accessed on 15 April, 2020).

CONTRIBUTORS

Mr Akhlak Ahmad holds the M.A., MLIS, M.Phil degree in Library and Information Science and currently is the Ph.D scholar. He is working with Centre for Women's Development Studies (CWDS) since 1993 and presently working Library Associate (Grade I).

In the current study, he collected, analysed the literature available on the topic and highlighted the research gap in the applications of standards tools for digitisation in Indian context. Also listed the challenges faced by the institutions while adopting new digitization technologies.

Dr Swadesh Sharma presently working as Dy. Librarian cum Asst. Prof. Head of Libraries, G. D. Goenka University, Gurugram Haryana. He is Ph.D. (Library and Information Science) from School of Studies in Library and Information Science, Vikram University, Ujjain. Having a rich experience of 17 years of working as a Library professional in Organisation, Management and services of University Libraries and College Library.

He contributed to this study by providing overall guidance in drafting and improving the contents of the paper.