Digitization, Strategies & Issues of Digital Preservation: An Insight View to Visva-Bharati Library

Tapas Kumar Das **Ajay Kumar Sharma** **Pritam Gurey**

Abstract

The digital preservation is of utmost importance in order to maintain the mission of preserving a record for future use. Digital preservation is the management of digital information over time. This paper explores an initiation towards digitization in order to preserve the rich collection, mostly rare documents at Visva-Bharati Library. The authors tried to highlight the need of digital preservation, strategies and issues of its related as well as digital preservation activities carried out by Visva-Bharati Library with intention to preserve the rare documents of the library.

Keywords: Digitization, Digital Preservation, Preservation, Visva-Bharati Library

1. Introduction

Library as information service provider has come to rely increasingly on digital information both as supplements to and parallels of print materials. Libraries are also procuring new resources, which are "born digital" that have no print or analogue equivalent - they exist only in digital form. Digitization has become a practical necessity and reality with technology interventions to provide improved access to information resources, preservation and dissemination as required, at any time; anywhere and any place as it were.

Digital preservation includes the preservation of print and non-print material in digitized form for effective, efficient and purposeful use. The purpose of preservation is to ensure protection of information of enduring value for access by present and future generations. Digital preservation is indeed a very challenging task for Library and Information Centre. The future of Library and Information services is closely associated to the preservation and the new technologies will create, collect, store, process and retrieve the information and deliver across the globe. Several issues of digital preservation including digital storage during the digitization process, migration of digital material, storage media are being faced to preserve the rare documents of the library.

2. Concept of Digital Preservation

A process by which data is preserved in digital form in order to ensure usability, durability and intellectual integrity of the information contained therein is called digital preservation. Digital preservation comprises of planning; resource allocation and application of preservation methods and technologies necessary to ensure that digital information of continuing value remain accessible and usable. The main purpose is to ensure protection of information of enduring value for access by present and future generations.

The term "digital preservation" refers to both preservation of materials that are created originally in digital form and never exist in print or analog form

(also called "born digital" and "electronic records") and the use of imaging and recording technologies to create digital surrogates of analog materials for access and preservation purposes. This means taking steps to ensure the longevity of electronic documents. It applies to documents that are either "born digital" or stored online (or on CDROM, diskettes, DVD, or other physical carriers) or to the products of analog to digital conversion, if long-term access is intended.

- **Short Definition:** Digital preservation combines policies, strategies and actions that ensure access to digital content over time.
- **Medium Definition:** Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.
- **Long Definition:** Digital preservation combines policies, strategies and actions to ensure the accurate rendering of authenticated content over time, regardless of the challenges of media failure and technological change. Digital preservation applies to both born digital and reformatted content.

3. Visva-Bharati

Visva-Bharati was founded by the great poet, prophet and philosopher, Gurudev Rabindranath Tagore. It aims at conferring Indian knowledge resource to rest of the world through various academic disciplines and on the other hand enriching Indian resources with derived philosophy from the world. Visva-Bharati was established on 23rd December 1921, located in the district of Birbhum, Santiniketan, West Bengal. By the Act of Parliament it became central University in 1951.

'Yatra Visvam Bhavatyekanidam' or 'where the world makes a home in a single nest' with this moto, Visva-Bharati offers the education from Kindergarten (KG) to Doctoral studies.

4. Visva-Bharati Library

The genesis of the library in Visva-Bharati goes back to 1901, with the founding of the Brahmacharya Asrama at Santiniketan. Rabindranath Tagore emphasized the use of books in the educational development of students. He personally supervised the selection of books, remaining alert to the needs of Santiniketan students and teachers and keeping himself aware of what was being published. The library of Visva-Bharati grew under his care with help coming from great minds all over the world. With the evolving of Visva-Bharati, Rabindranath toured Europe and America with the intention of collecting funds. He was often given large donations of books from Universities, individuals and groups of well-wishers.

Presently, the Visva-Bharati Library (Central & 12 sectional libraries) has more than 8 Lakhs volumes of books and bound volumes of Journals. The library subscribes around 300 Current Journals, some of these journals are available on-line and more than 5000 e-journals available from the INFLIBNET Centre. The Central Library has also purchased more than 800 e-books on various subjects from Elsevier, Cambridge and SAGE Publications, besides these more than 850 e-books are available in our home page that is free of cost. User strength is nearly 5818 and 550-user uses Central Library per day. Total

transaction per day in the circulation section is near about 300 books. Visva-Bharati has the intuitional membership of DELNET (Developing Library Network). The Central Library has introduced Inter Library Loan (ILL) and Photocopy of journal Article requisition services to its library members through DELNET. The library beside regular text holds enriched and important collections of old and rare documents. Special collections like: Prabodh Chandra Bagchi, Pramatha Choudhuri, Humayun Kabir, Sati Kumar Chattopadhyay, Lila Ray, Ashok Rudra, Abanindranath Tagore, and Panchanan Mondal.

5. An Initiation towards Digitization at Visva-Bharati library

The library has a huge treasure of knowledge, as it consists of old and rare documents which includes multi-lingual and multi-discipline books, reports, and manuscripts. Among them some are more than hundred years old, which are rare but now brittle in nature. Since the library holds mostly book documents, which are made up of paper and due to ages it makes them brittle. So preservation of these documents is a necessary. The primary objective of the library is to disseminate knowledge to its users from own resources. With the "Mega Digital Library Project of India" the digitization work had started with the help of Department of Information Technology, Ministry of Communication & Information Technology, and Government of India in 2005. The Ministry empowered CDAC-Kolkata for implementation of this project in Visva-Bharati. As the MoU has been signed between Visva-Bharati and CDAC-Kolkata on 12th March 2005 with the intention to digitize the following rare documents, which had published upto 1923:

a. Tagore's writings
b. Old books and journals
c. Rare collections
d. Special collections
e. Visva-Bharati publications

5.1. Phase I

After identifying the resources to be digitized and scanning, the following three processes were being taken into account by the CDAC-Kolkata:

- Noise cleaning
- Preparing metadata
- Stored the processed document(s) into DVD

Noise cleaning: The process of removing unwanted pixels from the scanned image is known as noise cleaning as it arises due to so many reasons like old document, low paper quality, dust particles, and use of low quality Ink as well as low quality printing machines. The process involves clear and unambiguous recognition, analysis and understanding of the document content. Here the processing of document includes text recognition, text presentation, formatting, skew correction and omni font handling etc. The first and foremost step after noise cleaning is skew detection and correction.

"ABBYY FineReader" software had been used for Optical Character Recognition "OCR" to clean and translate images (usually captured by a scanner) into machine-editable text only for English language documents. In the case of other languages documents, it is still awaited for OCR due to unavailability of proper OCR Software.

Preparing metadata: Description of a digital object it is "data about data" called metadata. Metadata is an important part of any digital preservation strategy. Within a digital repository, metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information. Metadata structure the data that

makes useful the other data. It provides contextual information.

The Dublin Core Metadata Standard having the "ISO-8859-1" is maintained for preparing metadata, a standard character encoding of the Latin alphabet.

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description like:

1. Title
2. Creator
3. Subject or Keywords
4. Description
5. Publisher
6. Contributor
7. Date of publication
8. Document type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

The following elements/tags are also used:

1. Copyright data
2. Scanning center
3. Scanning number
4. Digital republishes and
5. Digital publication date

Storing the processed document(s) into DVD: Transferred digital information is handled by portable storage devices such as recordable Tapes, Floppies, CDs, etc. and more recently by DVDs. The capacity of the DVD is greater than other devices. Processed documents were stored into single layer-single sided, 12cm in diameter with a capacity of 4.7GB DVD-R (mostly in Samsung-Pleomax and Sony). Each DVD contains the more than one document and the details of the stored documents are provided into the DVD by the CDAC. For example say a DVD contains 45 documents (Books), an Excel Sheet provided in the DVD with the details about the contents of document stored in it:

Table 1: Excel Sheet given into the DVD

Sl.	Date Digitization	Book Name	Author Name	Language	Pages
1	16.06.05	The Philosophy of Civilization vol.1	Towner, .R.H	ENGLISH	304
2	18.06.05	Africa In Trans formation	Maclean, DO Norman		306

Each stored document (in form of folder) contains five (5) sub-folders namely HTML, OTIFF, PTIFF, RTF, TXT, and Metadata Tag.

HTML: It contains the link among the pages

OTIFF: It contains original scanned image

PTIFF: It contains processed image (i.e. after OCR)

RTF: Rich Text File

TXT: Text file

So it is a work for the library professionals to compile a whole database out of all the documents that had been digitized rather than individual DVD. E-document Accession No. is also provided in the note area of each catalogue (manual & automated) so that when the user goes through the catalogue can have a proper retrieval facility of the digitized document.

List of Digitized Documents: Received DVD from CDAC-Kolkata is thoroughly checked by the library personnel whether it functions properly or not and preparing a list of the digitized documents in the Excel sheet followed by given format.

Table 2: Excel Sheet for Digitized Documents at Visva-Bharati Library

Sl. No	TITLE	AUTHOR (S) / EDITOR (S)	LANGUAGE	PAGES	CALL NO.	ACC. NO.	DVD I.D.	E-DOCUMENT ACC. NO.	REMARKS
1	1897-1904 Zweibund Englisch-Deutscher Gegensatz	Kohler, Wilhelm	German	144	940.3 K82	29143	DLI/CDAC(K)/0133	ED029001	
2	A Brief History Of Civilization	Hoyland, John S.	English	292	902 H85	51147	DLI/CDAC(K)/0033	ED001009	

There are two types of Accession Numbers given in the sheet, one is for print document and another is for e-document that helps in identifying the actual documents on the shelf and scanned documents in DVD respectively. We are using the remarks column for mentioning the original information about the document if there was any mismatch happened in between the details assigned by the CDAC and original ones. The DVD ID, a unique number that is used to identify a particular DVD from the DVD rack.

5.2. Phase II

After completion of first phase work, Central Library had received 106 DVD's from CDAC-Kolkata, which contains 5084 numbers of documents having 2036001 pages. Further work relating to the digitization of the rare documents at Visva-Bharati Library is under process. CDAC-Kolkata is waiting for Government sanction for second phase. The "List of Digitized documents of Visva-Bharati" is available on campus LAN (http://172.16.2.2).

6. Conclusion

A significant advantage of digitizing older documents is to make them accessible and searchable for future use. A well-thought-out plan to ensure preservation and accessibility of digital information must be developed and implemented. The Visva-Bharati Library, during the digitization process has learnt lots of major concerns and got practical experiences to resolve that. The future of research and scholarship depends on the ability to preserve information in digital form for the future use. Preservation keeps materials alive, intact and available for use so that they can be authoritatively used as long as possible to document our heritage and our society and to guide to others who will come in the future. As Philip Ward poetically states-

"Our heritage is all that we know of ourselves; what we preserve of it, our only record. The record is our beacon in the darkness of time; the light that guides our steps."

Digital information forms an increasingly large part of our cultural and intellectual heritage and offers significant benefits to users. At the same time preservation and access to this information is dependent on impermanent media and technologies; retaining metadata on the provenance and context; and retaining the authenticity and content of the resource. This paper explores about the process of digitization and the factors, issues related to its preservation of our cultural and intellectual heritage.

Acknowledgement: The authors are very much grateful to Dr. Subodh Gopal Nandi, Librarian; Sri Nimai Chand Saha, Deputy Librarian; Sri Achintya Kumar Mandal, Information Scientist and other staff members of the Visva-Bharati Library for their full co-operation, continual support and encouragement for preparing this paper.

References

1. **Alhaji, Ibrahim Usman.** Digitization of past question papers, dissertations and theses: A case study of 30 Nigerian university libraries. International Information & Library Review, 2007, 39, pp. 228-246.
2. **Bansal, Alka** [et. al.]. Securing the Future of Information: Digitization and Preservation of Documents in e-Format. DESIDOC Bulletin of Information Technology, January 2005, 25(1), pp. 19-26.
3. **Bullock, A.** Preservation of Digital Information: Issues and Current Status, Information Technology Services, Ottawa: National Library of Canada, 1999.
4. **Central Library,** Visva-Bharati. Annual Report of the Central Library, Visva-Bharati for the period from 1st April 2007 to 31st March 2008. 2008.
5. **Central Library,** Visva-Bharati home page: http://172.16.2.2 (available only in campus LAN) (Accessed on 05/12/2008).
6. **Chalravarty, R.** Digital Libraries: Issues and Future. In New Challenges in Information Management and E-learning in the age of Globalisation: Issues and Opportunities. SIS2003 21st Annual Convention and Conference, Roorkee, India, 9-11 April, Organized by Society for Information Science, 2003.
7. **Conway, Paul.** Digitizing preservation. Library Journal, February; 1990, pp. 206.
8. **Gladney, H. M.** Trustworthy 100-year digital objects: Evidence after every witness is dead. ACM Transactions on Information Systems, 2004, 22 (3), pp. 406–436.
9. **ICPSR.** Digital Preservation Strategies. http://www.icpsr.umich.edu/dpm/dpm-eng/terminology/strategies.html (Accessed on 28/11/2008).
10. **Kaur, A.** Archiving / Preserving Digital Information. In Road Map to New Generation of Libraries using Emerging Technology. 2nd International Convention on Automation of Libraries in Education and Research (CALIBER), New Delhi. India, Feb 11-13. Organized by INFLIBNET, edited by Murthy, TAV, et. al. 2004. Pp. 282-288.
11. **Lynch, Clifford.** Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information. http://www.dlib.org/dlib/september99/09lynch.html (Accessed on 21/11/2008).
12. **Witten, Ian H. and BRIDGE, Brian Bridge.** How to Build a Digital Library. San Francisco: Morgan Kaufman. 2003.

About Authors

Mr. Tapas Kumar Das, Assistant Librarian, Central Library, Visva-Bharati, Santiniketan.
E-mail: tapaskumar.das@visva-bharati.ac.in

Mr. Ajay Kumar Sharma, Assistant Librarian, Central Library, Visva-Bharati, Santiniketan.
E-mail: ajaysharma.lib@visva-bharati.ac.in

Mr. Pritam Gurey, Professional Assistant, Central Library, Visva-Bharati, Santiniketan.
E-mail: pritam.vb@gmail.com