PRESERVATION OF DIGITAL RESOURCES IN THE 21ST CENTURY: STRATEGIES, CHALLENGES AND THE LIBRARY PROFESSIONALS

NIMAI CHAND SAHA   **SUCHISMITA SARAKAR**

TAPAS MONDAL

Abstract

This papers deal with the various aspects related with the digital preservation. Preservation problem is complicated by the rapid obsolescence of the hardware and software required for it. Further, the technological ignorance of library personnel and users is one of the major obstacles on the way of development of Digital Libraries and its effective preservation. In this context, the paper highlights the strategies, issues & challenges, suggestions, skill requirement of the library personnel, infrastructure, system development and UNESCO's guidelines in short.

Keywords: Digital Preservation/ Medium Preservation/ Intellectual/ Preservation/ Digital Archaeology/ Copyright /IPR.

1.   Introduction

Due to the information explosion and the advancement of new technologies the concept of libraries shifted from traditional to digital. The nature of collection development has also been changes from printed document to electronic / digital document. As a result, the library professionals and users have gradually up-date themselves to offer and have the proper service with the help of the modern technologies in all respect e.g. Library software (SOUL, LibSys, CDS/ISIS, KOHA etc.), Digital Library Software (Dspace, Fedora, E-prints, Greenstone etc.), digital preservation project (DIAS, PANDAS, NDIIP, CAMiLEON, PADI etc.), hardware, E-Mail, Internet etc. In India most of the library professional as well as user's communities are lacks from the required skill to handle the new systems. But digital world challenges our notion of presentation. For several decades, specialists have voiced concern about the preservation of the portion of our cultural heritage in E-form. Widespread transition of knowledge from P-format to E-format has also given rise to the problem of its preservation in digital form. The major challenge rapid obsolescence of the hardware and software required to interpret and present digital documents has been widely discussed. Therefore, it is crystal clear that unless otherwise the equipment of the library professionals and users we can't get the actual benefit as offered by the new technologies.

2.   Concept and definition of DP

Digital Preservation (DP) refers to the various methods of keeping digital material alive into the future, typically centers on the choice of interim storage media, the life expectancy of digital imaging systems and the life expectation to migrate the digital files to future systems while maintaining both the future systems, full functionally and the integrity of the original digital system.

It includes everything from electronic publication on CD-ROM to online databases and the collection of experimental databases in digital format maintains the ability to display, retrieve and use digital collections in the face of rapidly changing technological and organizational infrastructure and elements. [3]

There are so many definitions for DP. Some important of them are as below;

The Wikipedia (2006) defines "DP as long term, error free storage of digital information, with means for retrieval and interpretation, for all the time span that the information is required for", where "retrieval" means obtaining required digital files from the long-term, error-free digital-storage, without corrupting the error-free stored digital files and "interpretation" means that retrieved digital files, which may be texts, charts, images or sounds, are decoded and transformed into usable presentations for access to human.[1]

"DP is the ability to keep digital documents and files available for time periods that can transcend technological advances without concern for alteration or loss of readability" (the Association for Information and Image Management).

Therefore, DP is defined as the managed activities necessary;

- For a long term maintenance of a bit stream (including metadata) sufficient to reproduce a suitable facsimile of the original document.

- For the continued accessibility of the document contents through time and changing technology. [10]

3.   Objectives

- Keep a place for new and emerging technology in the context of DP programs,

- To provide World Wide Accessibility,

- To reduce the effect of deteriorating factors such as temperatures, light, humidity, flood, fire, fungus, bacteria, insects, pollution, dust and the important factor called human,

- To maintain the historical value of information,

- To make it easy to use and handle,

- To make it information survives longer,

- Illustrate the combination of developments, events and decisions that led us to where we are today, in regard to technology that pertains to DP,[3]

- To protect the original documents,

- To present originals,

- To transcend originals etc.[2]

4. Utility

It helps to preserve rare and fragile objects without denying access to those who wish to study them.

Once books are converted to digital, users can retrieve them in seconds by searching for words, combination of words, phrases or ideas, readers can choose whether to view. Or store on a computer or take prints, several people can simultaneously read the same books or view the same picture.

Electronic copies occupy millimeters of space rather than meters on shelf and the problem of space vanishes when libraries opt to digital medium. [2]

5. Infrastructure

The computer system having the multimedia facility integrates all the media components to a single platform and provides interactivity to the system. The required IT components are as below;

- Central processing unit with high-speed processor,

- Large memory (RAM),

- Hard disc with high volume,

- Floppy drive,

- CD drive (Read / Write),

- Modem,

- Sound blaster card,

- Key board having multimedia keys and mouse scroll,

- Color monitor,

- Color / B & W laser printer,

- Better digitization tools (Scanner),

- Digital camera (still / move),

- Video camera for capturing analog video and convert into digital and

- Graphic cards etc.[3]

6.   Skill Requirement

The electronic environment of 21st century will encompass a wide range of technologies including computer, communication storage, recognition and other technologies. As such it is easy to say that future LIS professionals require knowledge and operating skills in all these areas. Mainly LIS professionals should have to acquire the following three kinds of skill to impart the service to its target users;   1) Technical skill, 2) IT skill,   3) managerial skill. Before getting into these three broad groups of skills, we may note that skills are not generally acquired by self-study or listening to lectures rather it should be acquired by 'practical sessions'.

6.1   Technical (cognitive and professional) skills

As part of professional skills the LIS professionals need to develop extraordinary access skills. The other technical skills, which are likely to be emphasized in future, are skills required to sift analyze, synthesis, assimilate, interpret and reformulate the information accessed and retrieved. This process not only involves providing quality information provision with censorship and quality control, but also calls for improving exploitation of new tools and resources.

6.2   IT skills

Like basic literacy (skills), information literacy and computer literacy have become a necessary of every walk of life. A computer literate is a person who has acquired the skills needed to use computer effectively. Such a person has to be 'comfortable' in his work in this computer age. This 'comfort' is the outcome of his familiarity, experience and understanding with computer. As a part of IT skills each and every library personnel should have acquire knowledge about operating system, IT products, software, physical handling of Gadgets, telecommunication products, DBMS, data and file management, DTP word processing, generation of reports etc.

6.3   Managerial skills

The managerial skill expected of new LIS professionals is quite vast. What is attempted to explain here is the managerial skills required for working in the E-world environment. To start with technological awareness coupled with skills for technology and product is fundamental. LIS professionals need to have managerial skills required for information management in team. A sort of leadership as well as fellowship skills is required to work in teams. [9]

7.   System Development

Various initiatives have been taken and various systems have been developed all over the world of digital preservation and archiving. With the help of the following table some of the initiatives cited: [8]

Sl No.	Name of the initiatives / project	Project Organizer
1	DIAS	Koninklijke Bibliothek (KB), the National Library of Netherlands a joined project with IBM. Extended through a new project called KOPAL with the German National Library, 2004.
2	CAMiLEON (Creative Archiving at Michigan and Leeds: Emulating the old in the new)	Universities of Michigan (USA) and Leeds(UK)
3	PADI (preserving Access to Digital Information)	National library of Australia
4	OCLC Digital Archive: Web Archiving: item by item Batch archiving: for collection	?
5	PANDAS (PANDORA Digital Archiving System)	? 2001, 2002. 2003.
6	NDIIPP (National Digital Information Infrastructure and Preservation Program)	Library of Congress, 2006
7	FEDORA (Flexible Extensible Digital Object Repository Architecture)	University of Virginia Library and Cornell University's Digital Library Group, 2003.
8	LOCKSS (Lots Of Copies Keep Stuff Safe)	Standford University
9	PMC (Portable PubMed Central)	U.S. National Library of Medicine (NLM).
10	GREENSTONE	University of Waikato, New Zealand Digital Library Project.
11	The E-Prints	University of Southampton
12	DSpace	MIT Libraries and Hewlett-Packard Labs.

8.   Strategy

Many DP strategies have been proposed, but no single strategy is appropriate for all data types, situations, or institutions. Tristram describes the following options available for DP: [1]
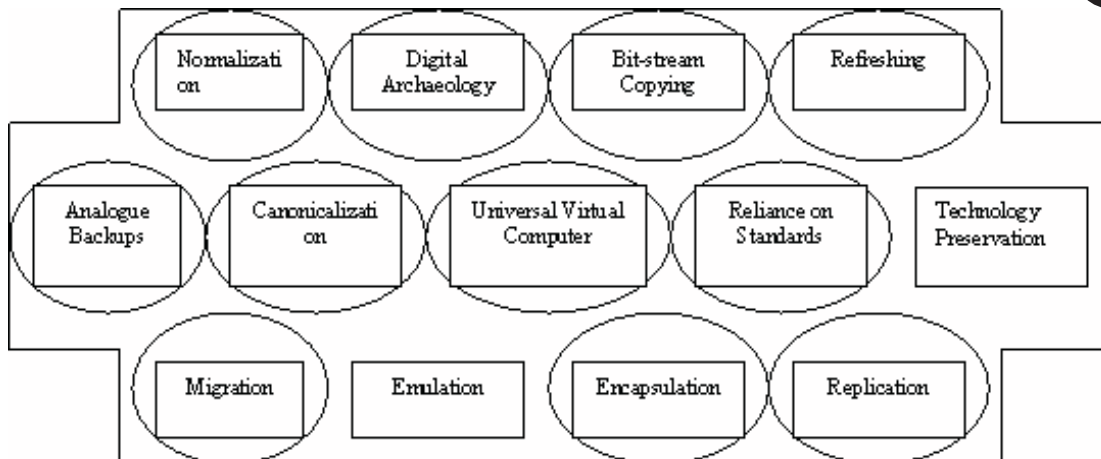
8.1 Normalization

Normalization is a formalized implementation of reliance on standards. The advantages and disadvantages of reliance on standards also apply to normalization.

8.2 Digital Archaeology

It includes methods and procedures to rescue content from damaged media or from obsolete or damaged hardware and software environments. Digital Archaeology is explicitly an emergency recover strategy. [5]

8.3 Bit-stream Copying

Bit-stream Copying commonly known as "backing up data" refers to the process of making an exact duplicate of a digital object .it should be considered the minimum maintenance strategy for even the most lightly valued, ephemeral data.

8.4 Refreshing

Refreshing essentially means copying digital information from one long term storage medium to another of the same type, with no change whatsoever in the Bit-stream (e.g. from an older CD-RW to a new CD-RW). "Modified refreshing" is the copying to another medium of a similar type with no change in the Bit- pattern (from a 100 MB Zip Disk to a 750 MB Zip Disk). It is a necessary component of any successful DP project. [5]

8.5 Analogue Backups

Analogue backups combine the conversion of digital objects in to analogue form with the use of durable analogue media. The digital object of an analogue copy can preserve its content and protect from obsolesce, without sacrificing any digital qualities including sharability and lossless transferability.

8.6 Canonicalization

It is a technique designed to allow determination of whether the characteristics of a document have remained intact through a conversion from one format to another.

8.7 Universal virtual computer

It is a form of emulation. It requires the development of a computer program independent of any existing hardware or software that could simulate the basic architecture of every computer since the beginning.

8.8 Reliance on standards

Reliance to standards is to mean "haredn" the encoding and formatting of digital objects by adhering to well-recognized standards and discarding proprietary or less supported standards. It is no more a permanent preservation solution than the use of gold CDs or stone tablets.

8.9 Technology preservation

Technological preservation is based on preserving the technical environment that runs the system; including operating system, original application software, and media drives etc. it is some times called the "computer museum" solution.

8.10 Migration

Migration is broader and richer concept than "refreshing" for identifying the range of options for digital preservation. It is a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/ software configuration to another or from one generation of computer technology to a subsequent generation.

8.11 Emulation

Emulation uses a special type of software, called an emulator, to translate instructions from original software to execute on new platforms. The old software is said to run "in emulation" on newer platforms. This method attempts to simplify DP by eliminating the need to keep old hardware working. This concept has been tested in several projects.

8.12 Encapsulation

Encapsulation may be seen as a technique of grouping together a digital object and metadata necessary to provide access to that object. Appropriate types of metadata to encapsulate with a digital object include reference, representation, province, and fixity and context information.

8.13 Replication

Replication is used to represent multiple digital preservation strategies. Bit-stream copying is form of replication. LOCKSS is a consortial form of replication, while peer-to-peer data trading is an open, free-market form of replication.

9.   Issues & Challenges

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

■   Medium preservation-the preservation of physical media on which the bits and bytes of electronic information reside,

■   Technology preservation-refreshing technologies from old to new as they become available,

■   Intellectual preservation-addressing the integrity and authenticity of the information is originally recorded. [4]
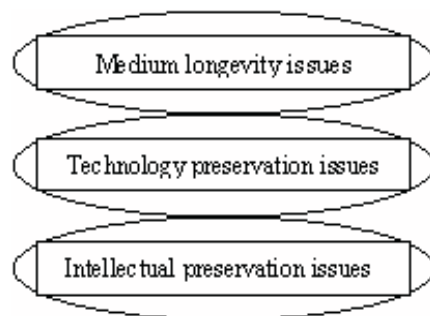

9.2   Challenges : Although, the digital technology offers several advantages over their print counter part, it along with other associated Internet and Web technologies are in a continuous flux of change. New standards and protocols are being defined on a regular basis for file formats, compression techniques, hardware components, network interfaces, storage media and devices etc. the digital contents face the constant threats of "techno-obsolescence" and transitory standards. Some of the important challenges for preserving digital contents are as follows: [1]

9.2.1 Machine dependency Digital contents are machine-dependent. It may not be possible to access the information unless there is appropriate hardware and associated software, which will make it intelligible. Access to digital contents may require specific hardware and software that were used for creating them. Digital contents stored on 5.25-inch floppy disk, for example, cannot be accessed since it has been suppressed by 3.5-inch floppy disk along with the drives to access data from it.

9.2.2 Technological obsolescence Unlike the situation that applies the books, digital archiving requires relatively frequent investments to overcome rapid obsolescence

introduces by galloping technological changes (Feeney, 1999). Technological obsolescence can effect hardware, software and file formats. Not only computers are continuingly superseded with their faster and powerful versions, the media used to store digital contents also become obsolete in two or three years before they are replaced by newer and denser version of that medium. As a result, even if the storage media is retained in the best condition, it may still not be possible to access the information it contains. [6]
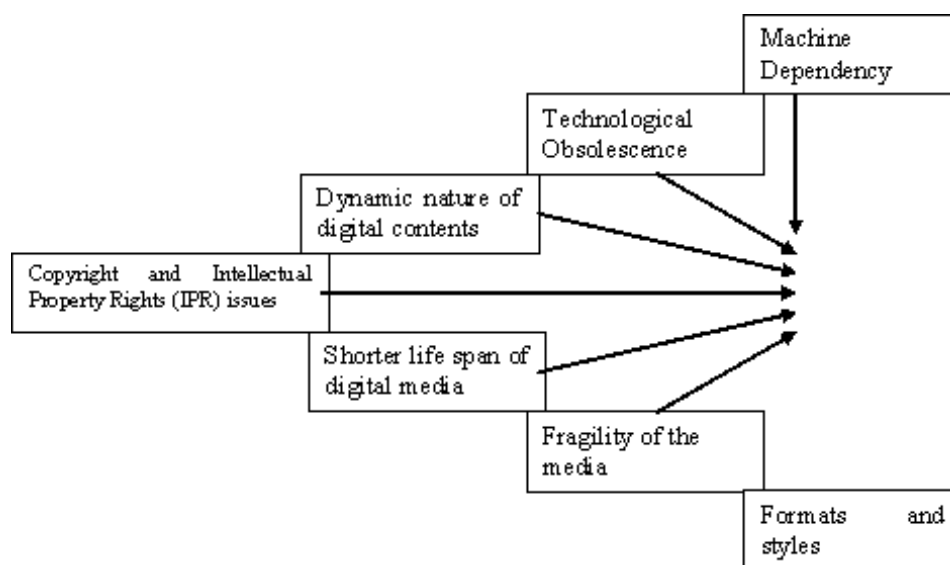
9.2.3 Dynamic nature of digital contents The initial problem with DP is the content itself (Chen, 2001). Preservation in analogue world involves static objects like printed documents, manuscripts and other artifacts, collecting and storing these items in some form is simple and straight word process. As the object grows and changes over time, new question emerge about what it means to preserve a digital object. Internet users are well familiar with the link failure syndromes that plague the web.

9.2.4 Copyright and Intellectual Property Rights (IPR) issues Copyright and IPR have a substantial impact on DP. The IPR issues on Digital contents are much more complex than for printed material. IPR issues in digital environment have implications not only on digital contents but also to any related software. Simply refreshing digital materials onto another medium, encapsulating content and software for emulation, or migrating content to new hardware and software, may lead to infringement of IPR unless statutory exemptions exist or specific permissions have been obtained from the right holders.

9.2.5 Shorter life span of digital media The greatest concern of DP is relatively short life span of digital media and higher rate of obsolescence of the hardware and software used for accessing the digital records. Rapid change in the IT industry and the move

from science-based developments to commercial development of software and hardware systems has resulted into media becoming inaccessible in to faster pace.

9.2.6 Fragility of the media The storage media used for storing digital contents are inherently unstable and highly fragile because of problems inherent to magnetic and optical media that deteriorate rapidly and can fail suddenly because of exposure to heat, humidity, airborne contaminants, or faulty reading and writing devices (Hedstrom and Montgomery, 1998). Deterioration of storage media may lead to corrupt digital files in such a fashion that it may not be easy to identify the corrupted portions of digital contents. Besides unintentional corruptions, digital contents are amenable to intentional corruption and abuse.

9.2.7 Formats and styles Information contents that were earlier confined to traditional formats like books, maps, photographs, and sound recording are getting increasingly available in diversity of digital formats. New formats have emerged, such as hypertext, multimedia, dynamic pages, geographic information system and interactive video. Each formats or style poses distinct challenges relating to its encoding and compression for digital preservation.

10. Suggestions

- Preservation of digital content is a continuous process that requires commitment and involvement, not only from heritage institutions, but also from governments, policy makers, producers of information and the software industry.

- Existing legislation should be adopted to support national heritage institutions in the preservation of digital materials. Deposit legislation should also be extending to all published materials.

- The information and communication technology (ICT) industry should be made aware of the need to take preservation requirements in to account. The value of standard and open source software should be promoted among software developers.

- Copyright legislation should not act as an impediment for preservation of digital heritage.

- Extensive training programmes are needed to disseminate the expertise and experience gathered so far widely among management and staff of heritage institutions. Programmes will have to focus, not only on technical aspects, but also on training staffs to deal with a changing environment and new directions.

- A vast amount of world's wide information is now produced digitally, and most of this exists in digital form only. Much of these digital materials is potentially of cultural value and new pro-active strategies need to be developed to ensure it is saved for posterity.

- Further research to develop promising models and technology should be widely supported in order to come to fully operational systems for preservation of digital content as quickly as possible.

■ The leadership role in digital preservation of a number of heritage institutions worldwide should be acknowledged.

■ A clear division of tasks and responsibilities, based on existing roles and expertise, needs to be established. It should be established how tasks can be shared between national heritage institutions and discipline-oriented organizations working (internationally) for specific communities.

■ Awareness of preservation issues should be raised with producers of digital information. They should realize the importance of use of standards and open source software and of adequate description and documentation. [2]

11. UNESCO Guideline

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

■ Not all materials need to be kept, only those that are judged to have ongoing value: these forms of the digital heritage.

■ Materials cannot be said to be preserved if access is lost. The purpose of preservation is to maintain the ability to present the essential elements of authentic digital materials.

■ Digital presentation must address threats to all layers of the digital objects: physical, logical, conceptual and essential.

■ Authenticity is best protected by measures that integrity of data is not compromised and by documentation that maintains the clear identity of the material.[10]

12. Indian Scenario

Million Book Universal Digital Library Project (Million Book Project): The Million Book Project initiated by Carnegie Mellon University, in some countries like China, India, Egypt etc. As on November 2005, around 170000 Books in India have been scanned. In India 22 scanning centers are running at different places with 4 nodal centers at different places like Allahabad, Goa, Pune, New Delhi, Hyderabad etc.

The Digital Library of ERNET India: the digital library hosted by education and Research Network (ERNET) is a collaborative effort between various Indian and US institutions. More than one million books are already digitized and available to access everyone over the Internet.

Center for Development of Advance Computing (C-DAC) digital Library projects: C-DAC is known as specialized agency to digitize the rare manuscripts available in India and have been working on various projects. [7]

Ministry of Human resource Development (MHRD), Govt. of India has advised all the consortium members of INDEST to set up E-prints archives using appropriate OAI

complaint E-print Software. MHRD also recommended that a central server may be deployed to harvest metadata from all such E-print archives. Again, INFLIBNET, the inter university center of UGC under MHRD has initiated Institutional Repository and archive its publications, proceedings etc using Dspace (dspace @ INFLIBNET). INFLIBNET's Institutional Repository and Archive-INDIA is an online electronic repository especially created for Indian Academia by INFLIBNET Center (UGC).

Some other initiatives in this context are: Institutional Repositories for research output and open access to full text Electronic Theses Databases (ETD); Indian Institute of Science (NCSI); Digital Library of Library and Information science (DLL) at DRTC Bangalore; Nalanda Digital Library project, NIT, Calicut; IIT Kharagpur, IIM kozhikode. [8]

13. Conclusion

Preservation is the oldest and most fundamental function of the libraries and archives. Archiving and related issues of digital preservation are becoming ever more significant within the scientific and scholarly communication chain. Digital resources, undoubtedly, have several advantages over its analogue counter part; however, preservation is definitely not one of them. The fact that the risk of loss of data in digital form is much greater than any other physical form is well understood and addressed to. Long-term preservation of digital information is plagued by short media life, obsolete hardware and software, slow read times of old media, and defunct website. For the successful completion, conduction, offering services and maintenance of the DP initiatives the concerned Library personnel should be active in nature and also conscious / aware for the required technology (both software and hardware). Mission (Training, Workshop, Refresher course, Orientation, initiation program etc.) be taken for the up gradation of the both, related Library personnel and target users in the periodically instead of occasionally. Because, human beings are the service providers and seekers of the digitized and non-digitized document. Therefore, it is very much essential to equip the concerned human factor in a planned way, so that they can able to cope with the frequent change of the technology.

References

1. Arora, Jagdish. (2006). Digital preservation and management. In. PLANNER-2006. Digital preservation, Management and access to information in the twenty first century [proceedings]. Ahmedabad: INFLIBNET center, 01-20p.

2. Chattopadhyay, Samir. (2006). Digital Preservation in the twenty first century: concepts, needs, problems and solutions. In. PLANNER-2006. Digital preservation, Management and access to information in the twenty first century [proceedings]. Ahmedabad: INFLIBNET center, 166-172p.

3. Garg, Babita & Chand, Salek. (2006). Digital Preservation. In. PLANNER-2006. Digital preservation, Management and access to information in the twenty first century [proceedings]. Ahmedabad: INFLIBNET center, 286-292p.

4. Graham, Peter S. (1998). Long-term intellectual preservation: Collection Management. Vol. 22, No. 3/4: 81-98p.

5. Hussain, Akhtar & Kumar, Krishna. (2006). Digital preservation system: A new approach. In NCIMDiL-2006. Information Management in Digital Libraries. Kharagpur: Central Library, IIT kharagpur, 563-568p.

6. Mulla, K.R. & Chandrasethara, M. (2006). Preservation of Digital information system. In. NCIMDiL-2006. Information Management in Digital Libraries. Kharagpur: Central Library, IIT kharagpur, 513-518p.

6. Muqueen, Shaista. (2006). Preservation of digital collection in India. In. PLANNER-2006. Digital preservation, Management and access to information in the twenty first century [proceedings]. Ahmedabad: INFLIBNET center, 247-250p.

7. Saha, Biswajit. (2006). Digital archiving and preservation: Present scenario. In. PLANNER-2006. Digital preservation, Management and access to information in the twenty first century [proceedings]. Ahmedabad: INFLIBNET center, 265-273p.

8. Sridhar, M.S. (2004). Skill requirements of library professionals in the electronic environment of 21st Century. In. Mahapatra, M. & Ramesh, D.B. (eds.) Information Technology applications in Libraries: A text book for beginners. Orissa: Reprorint (P) Ltd. 597-604p.

9. Sur, Subrata. (2006). Preservation of digitized information: new challenges in the 21st century. In. PLANNER-2006. Digital preservation, Management and access to information in the twenty first century [proceedings]. Ahmedabad: INFLIBNET center, 258-264p.

10. Sur, Subrata. (2006). Preservation of digitized information: new challenges in the 21st century. In. PLANNER-2006. Digital preservation, Management and access to information in the twenty first century [proceedings]. Ahmedabad: INFLIBNET center, 258-264p

BIOGRAPHY OF AUTHORS

Mr. Nimai Chand Saha presently working as Deputy Librarian in Visva-Bharati, Santiniketan, West Bengal.
Email: nimaichandsaha@yahoo.com

Mrs. Suchismita Sarakar is working as Assistant Librarian, Visva-Bharati, Santiniketan, West Bengal.

Mr. Tapas Mondal is working as Librarian at Lab Pur Jadavlal Institution, Lab Pur, Birbhum, West Bengal.