Digitization of Monographs, Rare Books at Indian Institute of Advanced Study Shimla: A Case Study

Lambodara Parabhoi   **Rupinder Singh**

Abstract

This paper highlights and discuss about digitization process, digital preservation which is an ongoing process at Indian Institute of Advanced Study Library, Shimla. The IIAS Library has been digitizing IIAS Monographs, Rare books and Images and also it is discussed the methods and equipment'sare being used in digitization process. The details volumes of woks and the tools and techniques has used. The paper also discussed objectives of digitization, technical issues, faults have found in digitization process.

Keywords: Digitization, Indian Institute of Advanced Study, Preservation, Digital Library

1. Introduction

Information plays a vital role for user community and Library plays a major role to dissemination of information. In recent past few years in the invent of ICT the dissemination of information has become very easy. Now days every Library has a digital Library to provide boundary less service. Libraries are doing digitizationof those documents which are very rare and valuable for future reference. In India there are lots of digitization project are carried out like National Digital Library Project (NDL) Digital Library of India (DLI) project,Digitize India Platform (DIP) is an initiative by Government of India, National Mission for Manuscripts launched by Government of India in 2003, and Indira Gandhi National Centre for Arts (IGNCA), Traditional Knowledge Digital Library and Panjab Digital Library etc. In abroad there are huge numbers of project carried out like Project Gutenberg, EBBA (English Broadside Ballad Archive). Many Universities and IITs have digitized their library collections and have come forward to contribute their digitized collection to a national platform.

2. About India Institute of Advanced Study

Indian Institute of Advanced Study, Shimla is one of the prominent research organization situated in Shimla, India. It was built in the year in 1964 by the Ministry of Education, Government of India and it started functioning as an institute for higher studies from October 20, 1965. The building that houses the Institute was originally built as a home for Lord Dufferin, Viceroy of India from 1884 - 1888 and it was called as the Viceregal Lodge prior to independence. It housed all the subsequent viceroys and governors general of India. After independence it was rename as RashtrapatiNivas.It occupied the Observatory Hill, one of the seven hills that Shimla is built upon (Chand, Prem;Rahman, Ajazur; Ahmed, 2010). The Institute has an International Centre for Human Development (IC4HD), Tagore Centre and Inter-University Centre. The institute offers Fellowship like National Fellows,

Tagore Fellows and Fellows and Associates for doing research in the field of Humanities and Social Science. The Institute has their own publication Studies in Humanities and Social Sciences (SHSS) is a bi-annual, Summerhill is an academic review journal published twice a year and Himanjali.

3. About IIAS Library

The IIAS Library, Shimla has a rich and robust collection of resources and other documents. The library was build up in the year 1965. It is one of the well-known special library in the field of Humanities and Social Science. The collection comprises of books manuscripts, Audio/Visual CDs, DVDs, Microfilmsand other resources in the area of History, Religion, Philosophy, Socio-linguistics, Psycho-linguistics, Social and Cultural Anthropology etc. The library has currently 1.5 lakh books, 40,000 bound volumes, and 213 current subscriptions to journals, magazines and newspapers. The Library subscribed wide range of e-resources though UGC-Infonet, NLIST and directly from the publishers as well. Library provides remote access services through EZ-proxy.

4. What is Digitization?

In recent few years the concept of digitization is heard often and has now become a topic of discussion among all library thinkers and librarians. Digitization has bestowed upon librarians and archivists of the late 20ᵗʰ and early 21ˢᵗ centuries the opportunity to reexamine how they access their collection (Brenner, Larsen, & Weston, 2006).Digitization is the conversion of materials or analog media into electronic form for creating digital collection (Singh, 2006). According to Wikipedia "Digitizing or digitization is the representation of an object, image, sound, document or signal (usually an analog signal) by generating a series of numbers that describe a discrete set of its points or samples".ODLIS stated that "The process of converting data to digital format for processing by a computer. In information systems, digitization usually refers to the conversion of printed text or images (photographs, illustrations, maps, etc.) into binary signals using some kind of scanning device that enables the result to be displayed on a computer screen". Digitization is a process where physical media and documents e.g. books, journal articles, monographs, old manuscripts which are rare in nature and convert them in to digital form i.e. a computer readable file such as PDF, Rich Text Format, HTML or EPub and can be accessible by computer, electronic device such as tablet and e-book readers.

5. Need of Digitization

Books, manuscripts and other rare documents which are very precious material for our human civilization and cultural heritage. These documents comprise paper which is made up of organic material such as cellulose pulp derived from wood or grasses which decay with time, slowly damaging them and making them unreadable. Loss of such valuable information is devastating. The main needs of digitization are to conserve and preserve the original documents and use those documents in electronic format infuture. On the other hand, it is a solution to solve the problem of housing and accommodating the library resources and it overcomes the space problem in every library.

6. Objectives of Digitization

Digital documents come in vary handy and fulfills following objectives.

❖ To conserve and preserve the knowledge

❖ To reduce space problem

❖ To easily disseminate the information to the users

❖ To provide easy mobility to the user

❖ To facilitated remote access

❖ To save the many as well as maintenance

❖ To saving the time of the user

❖ To increase the accessibility

❖ To facilitate the reproduction of the rare documents

7. Important aspect to Consider in awarding Work of Digitization.

It is a crucial decision for the librarians whether to send the documents out to some agency for the scanning, or whether to engage a team for digitization and execute the whole process in house. But it all depends upon the policies of library, and the nature of the documents. If the documents selected for the scanning are extremely rare and classified in nature, it is advised to engage a team or agency to do that work for you within your library premises, leaving no chance of theft and leaking of vital information. The IIAS library called for open tender for digitization of IIAS, monographs, photographs and rare books from different agencies and has received from no. of agencies. The tender has two section Financial and technical bids after deciding proper qualifying criteria.

8. Selection of Documents

The documents to be scanned are selected from the library collection, old publications of the institute which are not available in the digital format and other rare books and archival material available in the library is selected for the scanning and digitization.

The number of the IIAS publications goes to 300 books with approx. 450 Pages each, and 250 rare books and manuscripts of various page numbers and sizes. However, a different procedure is followed for Institutional publications and rare book. Books which are in the IIAS library collection and are older than 100 years from today were also included in this process.

9. Hardware and Software use for Digitization

9.1. Computer Configuration

The IIAS Library has engaged 6 pcs with appropriate configuration for the purpose of scanning and storage work. The configurations of the PCs are Core i3 Processor, 4 GB RAM and 250 GB HDD on each PC. All the computers are connected on an isolated LAN, to insure the data security. For data backup and redundancy 2 external hard drives are used and each of hard drive has the capacity of 2 Tera Byte. Data backups from all computers are taken manually on weekly basis.

9.2. Scanner

Since natures of the documents are different, the library is using two different types of scanner for digitization of manographs and rare books. Two ADF scanners iball A305 are used for scanning of the Institutional publications and monographs. And Fijistu Truecolor overhead scanner which is used for scanning of rare books and other archival materials which is to be handled with extreme care.
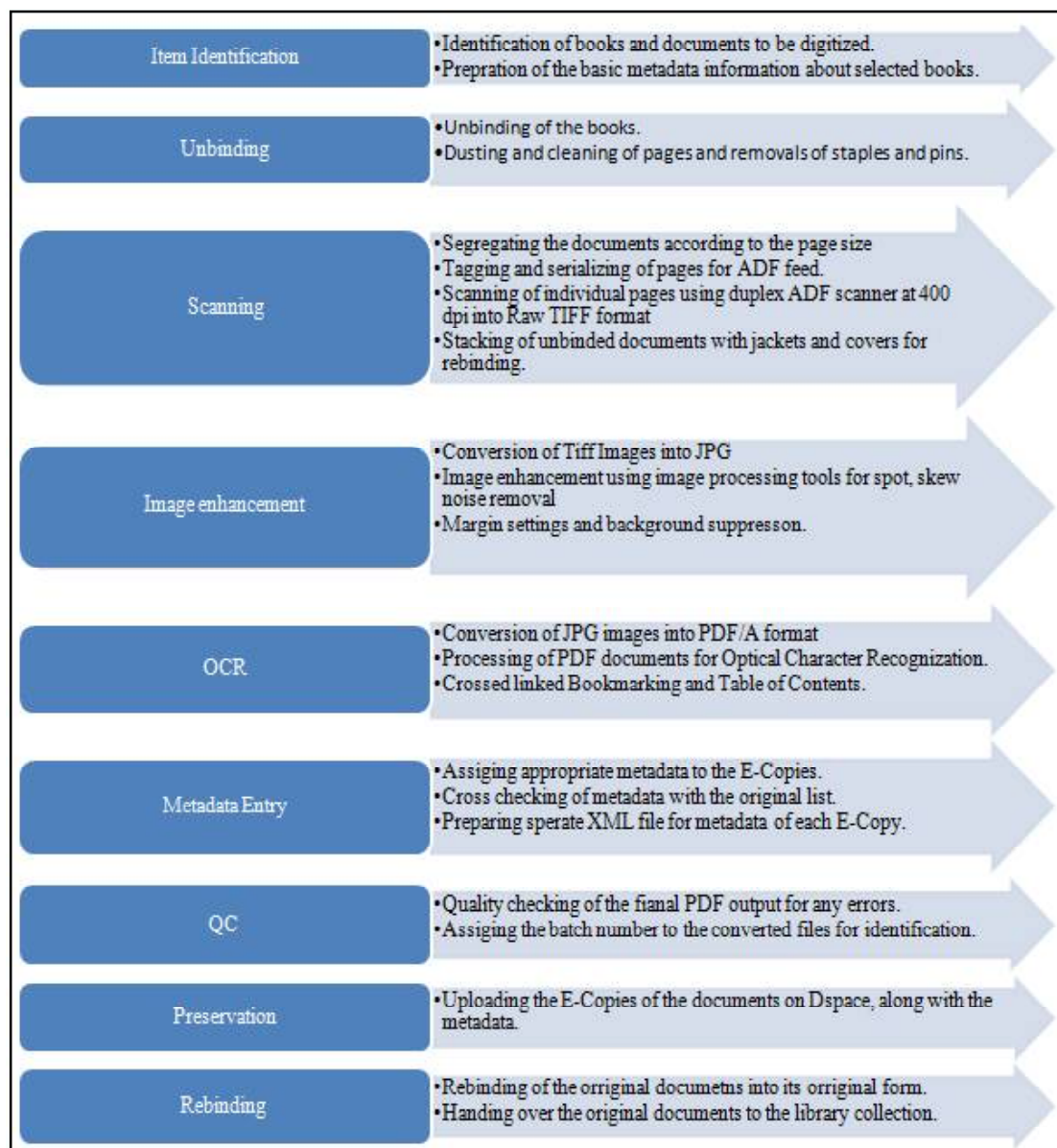
9.3. Image Editor

Scan tailor version is a free ware interactive post-processing toll and it is used for image enhancement and background suppression. Equal margins are added all around the text. Adobe Photoshop CS3 is used for removal of patches, skew, distortion and noise.

9.4. OCR and Conversion

Adobe Extender Pro which is a proprietary PDF editing tool used to perform optical character reorganization and conversion of the JPG images into PDF.

Work flow for Digitization

9.5. Manpower

There are 6 persons are engaged with the digitization process and they have assigned different tasks to them.

The following process and ware involved in the digitization process and the flowchart shows the work flow.

Item Identification	• Identification of books and documents to be digitized. • Prepration of the basic metadata information about selected books.
Unbinding	• Unbinding of the books. • Dusting and cleaning of pages and removals of staples and pins.
Scanning	• Segregating the documents according to the page size • Tagging and serializing of pages for ADF feed. • Scanning of individual pages using duplex ADF scanner at 400 dpi into Raw TIFF format • Stacking of unbinded documents with jackets and covers for rebinding.
Image enhancement	• Conversion of Tiff Images into JPG • Image enhancement using image processing tools for spot, skew noise removal • Margin settings and background suppresson.
OCR	• Conversion of JPG images into PDF/A format • Processing of PDF documents for Optical Character Recognition. • Crossed linked Bookmarking and Table of Contents.
Metadata Entry	• Assiging appropriate metadata to the E-Copies. • Cross checking of metadata with the original list. • Preparing sperate XML file for metadata of each E-Copy.
QC	• Quality checking of the fianal PDF output for any errors. • Assiging the batch number to the converted files for identification.
Preservation	• Uploading the E-Copies of the documents on Dspace, along with the metadata.
Rebinding	• Rebinding of the orriginal documetns into its orriginal form. • Handing over the original documents to the library collection.

## 10.	Things to Consider

Physical condition: It refers to the documents that will be digitized. The condition of documents must be suitable for digitization. It should not be completely damaged

10.1 Purpose

The library must have a clear idea about the whole objective for this work. Why digitization is required and what purpose is it going to serve. Who would be benefitted from the outcome of this project. of course primary purpose of digitization is to preserver the originals and make them available to the audience without causing any damage to the original books which are in physical form.

10.2 Cost

In-house scanning and outsourcing of the scanning work both involve large financial implications. It is to be considered that the library has sufficient funds for digitization of documents. However, for a small collection, outsourcing of work would be advisable whereas, if the library has a large collection of rare books and the policy or condition of documents doesn't allow them to be taken out from the library premises, In-house scanning setup can be established but it would cost on the higher side as compared to the out-sourcing.

10.3 Copyright & Permission

While digitizing the rare and old documents if you are planning to publish the e-copies online, one of the important thing to take care is to check if the document is available in public domain if it is not, you must have permission from the copyright owner to digitize the material.

10.4 Audience

The true beneficiary of these resources would obviously be the members of the library. But if the documents are of sensitive nature, the library policy must be framed to restrict its access to a particular category of users.

10.5 Discovery & Access: The scanned copies of the documents must be stored on a centralized server with appropriate software for search and retrieval of the documents. The software must be chose considering its cost, flexibility, data security and access control system. Being open source and very robust, DSpace is a good choice. The same is in use at IIAS, Shimla Library.

## 11.	Minimum Scanning Guidelines

The library of Congress has described some standards for the scanning of the textual documents and images (both monochrome and color). But it depends on your library and nature of the documents to increase and decrease the optimum standards. At IIAS Library we have followed different criteria for different documents, which is as:

11.1 Rare Books

Scanning resolution 400 dpi full color, TIFF document. With No background suppression or text re-construction. NO OCR is done for these files. No rare document is unbounded or changed from its original state. The scanning of these is done only using high quality overhead scanner.

11.2 IIAS Monographs

Scanning resolution 300 dpi grayscale for text. 400 dpi full color for images in uncompressed RAW TIFF files. The RAW images are converted into Grayscale JPEG images after spot removal, margin adjustment

and text reconstruction. OCR and bookmarking is done for these file after converting them into PDF/A format.

It should also be kept in mind that the entire process should be on the track all the time and the work should be executed within the proposed time, so that the documents are not engaged for long, in the scanning phase. Since the library users keep coming to the library for using these documents, Library services should not be hindered just because the books remain unavailable for the users for so long times. The books should return to their designated place, immediately after their scanning part is complete. Other thing to consider is to keep the work space clean and moisture and dust free. If the project is executed within the library, access to the scanning workplace should also be restricted and should be monitored all the times.

12. Collection Management

The IIAS Library use DSpace software to manage to digitize documents. After digitization process, the digital documents go through quality control process. The following procedure has follow by IIAS Library:

❖ Organization the digitize documents

❖ Naming the documents

❖ Proper checking of metadata

❖ Cataloguing

❖ Indexing the documents

❖ Give subject keywords

❖ Give access right to the documents

13. Challenges of Digitization

❖ The digitization project of IIAS Library has faced a number of challenges.

❖ Library staff should have technical qualification so that they can handle the digital asset very easily.

❖ The technology has been changed day by every few years ICT devices has dramatically changed in both software and hardware

❖ The library should have proper design search and retrieval tools for easy cataloguing of digital material and facilitates easy retrieval of information.

❖ The most important thing to consider when you are uploading a huge data into a server. It should have the ability to migrate over the other softwares in case the need occurs.

❖ It is one of the major issue when the digitize the documents It addresses legal issues of physical and digital materials like, copying, access, and dissemination of the resources.

❖ Library should have the proper plan to facilitate digitize documents and Integratethem both digital and physical materials

❖ Library should have proper plan to provide and make use of digitize documents for the use of user.

❖ Provide more efficient and more flexible tools for transforming digital content to suit the needs of end-users sustaining the resource

❖ Library should have proper plan and budget to maintain the digitized materials hence it is very expensive to maintain the digital materials.

15. Conclusion

It is suffice to say that the digitization is just going one step forward in improving the user services at the library. The IIAS Library has initiated the work of digitization of the rare books, IIAS monographs and photographs. Which we believe, will be very useful for the library members and for Institutes own academic and administrative purposes. It will improve the accessibility and preserve the rare documents for future. This also ease the process of reproduction of the vulnerable documents. However, it is worth mentioning that the digitation process not end of story, the digitized documents needs to be properly maintained and access to those documents must be easy and in controlled manner for use of the library patrons and scholars across the globe.

References

1. Ahmed, F. (2009). Digitization as a Means of Preservation of Manuscripts: Case study of Osmania. In 7th International CALIBER-2009 (pp. 93–97). Pondicherry: INFLIBNET.

2. Brenner, M., Larsen, T., & Weston, C. (2006). Digital collection management through the library catalog. Information Technology and Libraries, 25(2), 65–77. http://doi.org/Article

3. Chand, Prem;Rahman, Ajazur; Ahmed, J. (2010). Planning for Digital Library of Special Collections in Indian Institute of Advanced Study (IIAS) Using Dspace/ : A Case Study. In 7th Conventional Planner 2010 (pp. 480–491). Tezpur: INFLIBNET.

4. Dadzie, P. S., & Walt, T. Van Der. (2015). Planning for Digitization of University Libraries in Ghana/ : Challenges and Prospects. Library Philosophy and Practice.

5. Davis-Perkins, V., Butterworth, R., Curzon, P., & Fields, B. (2005). A study into the effect of digitisation projects on the management and stability of historic photograph collections. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3652 LNCS, 278–289. http://doi.org/10.1007/11551362_25

6. Fabunmi, B., Paris, M., &Fabunmi, M. (2009). Digitization of library resources: challenges and implications for policy and planning. International Journal of African & . Retrieved from https://ojcs.siue.edu/ojs/index.php/ijaaas/article/view/80

7. Hughes, L. (2006). Why digitize? The costs and benefits of digitization. Public Information Technology and E-Governance: Managing the Virtual State. Retrieved from http://www.facetpublishing.co.uk/downloads/file/sample_chapters/digitizing collections chapter 1.pdf

8. Jagboro, K. O. (2012). Digitization of Library Collection in Developing Countries/ : The Hezekiah Oluwasanmi Library Experience. Library Philosophy and Practice, 1–11.

9. Kanndppanavar, B. U., Rajanikanta, S. T., &Tandur, S. K. (2010). Importance of Digitization of Library Materials. SRELS Journal of Information Management, 47(1), 61–70. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=51145455&site=ehost-live

10. Seifi, L. (2011). Digitization and digital preservation of manuscripts and access in organization of Libraries, Museums and Documents Centre of Astan Quds Razavi Library in IRAN: A, (280), 280–284. Retrieved from http://ir.inflibnet.ac.in/handle/1944/1618

11. Singh, S. N. (2006). Major Digitization Initiatives in India. In 4th Conventional Panner 2006 (pp. 9–10). Aizawl.

12. Vijayakumar, J. K., & Jeevan, V. K. J. (2001). Digital library development/ : major issues of externally published contents, 174–184. Retrieved from http://eprints.rclis.org/7228/

About Authors

Mr. Lambodara Parabhoi, Professional Assistant, Indian Institute of Advanced Study, Rashtrapati Niwas, Shimla.
Email: suresh19871987@gmail.com

Mr. Rupinder Singh, Indian Institute of Advanced Study, Rashtrapati Niwas, Shimla.
Email: rupinder@outlook.in