

---

---

## **DIGITIZATION AT GANDHI SMRITI LIBRARY OF LBSNAA: AN INITIATIVE TO PRESERVE THE RARE BOOKS**

***Rajender Singh Bist***

### **Abstract**

*Manuscripts provide rich, authentic information, facts and evidence for historical, legal and research references. Physical condition of the manuscripts may be decayed due to factors such as climate, insects, acidified paper and old age. A core function of libraries is to maintain and preserve the decaying/ damaged manuscripts for the present as well as use of the future generation. The most significant development in the recent times is to digitize the fragile and rare documents for better access, storage, preservation and dissemination. Digitization is an important aspect of developing digital libraries as it opens up new avenues of access, use, research and preservation of valued information resources. This paper deals with the key issues of digitization at the Gandhi Smriti Library of LBSNAA, which have started digitizing its rare collection with the aim of providing the cost-effective technical solutions for efficient development and delivery of digital services to provide enhanced access to the Library's collections and other documentary resources, and also to provide shared access to digital collections in cooperation with other institutions. Further, it strives to provide the infrastructure for long-term management of digital material in the Library's collection through provision of hardware and software systems supporting integrated collection management in a digital environment.*

**Keywords:** Digitization; Digital library

### **1. Introduction**

Information being an intellectual resource has the capacity and power to transform the image of society and of changing the very direction of human life by way of preserving and sharing the essence of knowledge passed down by innumerable wise men to the future generations for posterity and use. The libraries of the new millennium find themselves in the difficult position of having to maintain and extend traditional library services, and develop, implement, and upgrade electronic and networked based information services. Thus, there is an increased demand on library services and provision of electronic resources in the libraries. With the emergence of Information and Communication Technology (ICT), the role of Library has expanded and it is necessary to incorporate technological gadgets to modernize the library services with utmost care and diligence in order to attract the society and justify its existence. During the present Internet and e-resources era, efforts are needed to adopt the newer technology to meet the needs and aspirations of the society. Gordan's opinion is that, the recent developments are creating new pressures, however, and the expectations on informational professionals are changing [1]. Similarly Alex Byrne said that, the development in ICT give us good opportunities to enhance research and scholarship by returning to some of our

---

roots in value adding, to revisit some of the activities which we have declined to undertake in the pursuit of efficiency [2]. Digitization is an important aspect delivered by these ICT developments. In the words of Sudha S. Murthy "digitization has become a practical necessity and reality with technological interventions to provide improved access to information resources, preservation and dissemination as required, at any time; any where and at any place as it were. The tremendous impact of digitization especially on scholarly communications across the globe not only supports research but also to large extent prevents duplication of research". [3]

## **2. Definition of Digitization**

Digitization is defined as, "conversion of analog items into digital format for the purpose of extending access and, where appropriate, to assist with preservation. Digitization is not an activity which can be seen in isolation. It is linked to all aspects of services provided by the library." [4]. In simple words digitization means acquiring, converting, storing and retaining information in standardized and organized manner with technological support. With specialized scanners documents are converted into digital formats and stored for further reference via CDs or web based application. The digital content may be locally held or accessed remotely via computer networks. The fundamental requirements for the digitization are process of information selection; selecting documents, and selecting data, digitization gadgets, software, etc.

## **3. Why Digitization?**

Following are some of the reasons why the library needs to digitize.

- To make available the reading specially rare and fragile materials.
- This helps to search automatically and swiftly and also they are more quickly and efficiently accessible. It also furthers e-learning opportunities.
- Preservation of material is made possible here by making the digital copy available to the readers, which saves the original document.
- Digitization also helps in promoting and marketing of library resources worldwide. It also attracts the reader to the library to see the original material.
- Further, it increases the revenue of library.
- Information has got the economic value and also for the sake of society, democracy, education, advancement of science and technology universally.
- Maximum utilization of the documents leads to, good life consisting of freedom, health, ethical conduct, wisdom and well-being.

## **4. Designing the Project: Policy and planning**

For any digitization project we are bound to consider various issues relating to its design and prepare a list of high level requirements. This normally includes information that the library contain, how that information will be generated and to what audience the information is indented for and how this data will be accessed through a well designed policy. Digitization policy for the libraries should clear in its stand towards sharing the resources at local, regional, national and International level and clear about standards and specifications for digitization and make provision for staffing,

---

for conservation, digitization and dissemination. Digitization project also requires thoughtful planning, designing, implementation and evaluation. Inadequate and unplanned project without sound management and control are likely to fail, therefore an essential element of project planning is to ensure that the purpose or vision of the project is clear and well communicated. Tanner states that to develop a vision for a project and its planning it is necessary to be able to perceive how all the elements of the proposed project fit together. A holistic overview of the whole lifecycle of the project is important to ensure that the aims, objectives, available resources and the deliverables are complementary. He has listed some thoughtful elements for planning the project. They are [5]:

- Assessment of the need for digitization.
- Selection of material.
- Deciding the goals from the information content of the originals.
- Deciding how to reach the final product and to maintain/ preserve the resource over time.
- Finding the funds for the project.
- Planning the project and assigning resources.
- Digitization feasibility study to determine procedures and benchmarks.
- Preparing the originals for digitization.
- Conversion and scanning.
- Adding metadata (technical, administrative and subject based).
- Quality assured checks to ensure the output conforms to specification.
- Make provision for archiving and preserving the data.
- Return originals to their place in the collection.
- Rights management, mounting data, resource support and final evaluation of the project.

## **5. Core Components of Digitization**

There is variety of tools like information organization tools, information retrieval tools and collection management tools and techniques required to build good information retrieval systems and user interfaces which leads to building a digital library. The library, which desires to develop a digital library by way of digitization, should have to consider the following:

### **5.1 Technological issues**

like accessibility; search engines; outsourcing DL resources for digitization; fiber optic connectivity; manpower; training; DL management skills; ICT skills; information skills; management skills; research and project management skills

### **5.2 Tools and Techniques**

Such as scanning and indexing; security; delivery; interoperability; knowledge management; standards etc.

---

---

### 5.3 Hardware requirements

High-end servers, CD/DVD servers, CD/DVD writer; computer PIV with high capacity hard disk for server and clients in the LAN, web servers and FTP servers; workstation PCs; capture devices-scanners, cameras; high power UPS ( 10 – 20 KV); printers, consoles and test computers; output devices CD-ROM, DVD, OCR; high speed LAN, WAN, internet connectivity etc.

### 5.4 Software: Operating software:

OCR Software; scanning software; Acrobat reader; CD-read / writer software; DL (OS) software e.g. Dspace, Greenstone, Fedora, E-Prints etc.; windows NT networking software, SQL server software, database management software; web designing software like java, front-page, XML etc; RDMBS (Relational Database) supporting the variety of digital databases like Oracle, Posgre SQL, MYSQL etc.; full text search engine to index and provide access to digital resources.

## 6. The Institute and Library Background

The Lal Bahadur Shastri National Academy of Administration (LBSNAA), Mussoorie (Uttanchal), set up in 1959, is a premier training institution for the higher civil services in India. Common foundation courses are held for entrants to All India Services and group 'A' service of the Union. The professional training to regular recruits of the Indian Administrative Service (IAS) and members of the Royal Bhutan Service is conducted after the foundation course. The Academy also conducts in-service training courses for members of IAS and officers promoted to the IAS from the state civil service, as well as workshops and seminars on policy issues.

Gandhi Smriti Library (GSL) of the Academy is one of the most prestigious, automated and well-equipped Libraries in the country. The Library is situated on the first and mezzanine floor of the "Karmshilla" building, from where one can feast their eyes on the serene snows and majestic deodars. The Gandhi Smriti library strives to provide material and resources, including the emerging technologies that are diverse, accessible and up-to-date to serve the needs of the user community at the LBSNAA. The Library caters to the reading, research and reference needs of the faculty members, officers trainees, staff, research associates and participants of the various courses, workshops, seminars etc. run/ organized at the main as well as Indira Bhawan and NIAR campus. GSL has a collection of over 1.55 lakh books, which has been rebuilt from scratch after a devastating fire accident that occurred in may 1984. The library also has a separate audio-visual unit with a collection of about 4,000, audio, videocassettes, CDs and DVDs and subscribes to about 275 journals, periodicals, and magazines. Hindi section comprises of 16,000 well stocked books and a separate state specific section houses about 25,000 books, reports etc. The GSL is fully automated and computerized library, providing online services through the web-enabled libsys-LS Premia and web OPAC accessible from the Academy web site: [<http://www.civilservices.gov.in>]. One of the recent goals and objectives of the GSL has been to develop the GSL into a technologically advanced library and provide technologically advanced services for most economical cost possible. Technologies like digitization (in progress), electronic databases, maintaining up-to-date web presence etc. are the recent priorities.

---

---

## 7. Digitization at GSL

Rare collection is the most invaluable collection of GSL. There are approximately 8,000 rare and fragile documents that need special care and preservation, which is maintained in a separate section, called Gyanlok and is organized systematically in wooden shelves and godrej compact shelves. The GSL is digitizing the archival collections housed in the Gyanlok section, basically for preserving them and as well as to facilitate their access to the LBSNAA user community. The main goal is to enable the users to have easy access to these archival materials without physically handling the documents and thus avoiding further damage and the ultimate intention is to make a repository of old publications, the literature produced within the institution to cater to a larger group of administrators and researchers.

It was felt that rare and fragile documents housed in Gyanlok are being irretrievably lost in the sense that their proper utilization is being neglected on the user's part, though they are valuable for its aesthetic and historic value. In addition, numerous in-house publications are not well organized up to the extent which facilitates their use, though they are the most demanded by the officer trainees. Lecture notes, government circulars, short reports, policy documents etc. having futuristic value are scattered in various sections and their retrievals on demand is almost nil or requires a lot of manual searching through files and folders. Hence, to overcome this lacuna of inadequate documentation and a proper integration of all institutional documents the concept of digitization gained consensus within the authorities. A comprehensive project proposal for digitizing the rare collection was initiated in 2005. It was necessary to clarify the digitization needs, investigating our resources i.e. manpower, equipment, money, time etc., free use, and deciding the audience i.e. target users. The document categories to be digitized were clearly segregated as institutionally produced (Assignment reports, syndicate reports, socio-economic surveys, village study reports, term papers and back volumes of the journal 'Administrator' published by the institute) and the old publications published prior to 1950. The faculty gave due consideration to the proposal and recommended the digitization of the proposed documents. The rare documents intended for open access are to be uploaded on the web whereas the documents intended for internal use will be placed on the intranet.

The main reasons advocated to the digitization project are:

- To improve services to our esteemed users through the provision of enhanced access to the institution's resources with respect to education, training and long life learning.
- To increase access and decrease handling of fragile documents.
- To create digitized content of term papers, syndicate reports, assignments and other documents submitted by officer trainees/faculty/participant of various courses, which will facilitate their multiple use from a single original.
- To create a highly interactive electronic network at GSL.
- Whole knowledge cannot be incorporated in a single library; therefore, resource sharing in the digital era receives special attention. In particular some form of exchange of digital data or some means of common access these data is the prime concern of GSL.
- Library is a growing organism, therefore space problem are likely to arise in future. After digitizing the materials, the original source documents can be shifted to the storeroom situated

---

at the ground floor of the "Karmshilla" building, thus providing space for accommodating the current documents.

- Nearly 75% of our rare collection is unknown, inaccessible and fragmented leading to an unutilized cost of handling and shelving them. Digitization will surely do away with the cost of housing and shelving it and simultaneously provide better facilities of data collection coupled with quality control and faster access.

### **7.1 Selecting material**

Selection regarding the right content for the right user at the right time is must and it includes defining value of information to be digitized, recognizing it, and then deciding to address its preservation needs in most appropriate way. The selection of source material is more or less content driven as we are dealing with preserving intellectual content of the document and not just physical preservation. Choosing the material for digitization require a deep study regarding the relevance as well as the condition of the material. For the purpose involvement of our faculty/experts was sought, whose managerial skills ensured that the relevant and documents of immense use are selected. Careful assessment of how the original is used by the library patrons was made (i.e. collection's past history of use and potential for use in future, document's aesthetic, historical and bibliographical values. Frequently used documents (Assignments written by OT's, Syndicate reports, Term papers, Village study reports, Surveys, Case studies, Core readings of different courses and many other in house publications.) were selected for its multiple usages from a single original to avoid wear and tear and easy retrieval. Rare and fragile and oversized documents were identified to prevent their further deterioration and make rare objects like; manuscripts, books with important annotations, maps, and photographs etc, more accessible and preserved.

### **7.2 Digitization Options**

The choice about whether to undertake the digitization of the texts in-house or to outsource the scanning of the texts to an external agency was the key. Thus a number of critical factors and options were weighed up carefully before the final decision was made. An in-house digitization operation was deemed more cost effective. No matter, in the initial phase we can outsource the documents but the ultimate intention is to train the staff in the digitization technology so that the digitization can be maintained on the continual basis for the institutionally produced literature, with the library having all the required hardware/software. The institute have mostly all the required components and gadgets required for digitization. Only one book scanner along with the required accessories and the digital library software need to be procured. For in-house digitization there were two options; using a Minolta scanner, or purchasing a flatbed scanner with sheet feeder. However, there would be a single capital outlay of approximately Rs. 17 lakh and staff would need to be trained in its operation in quality assuring output. For initial cost identification we contacted various agencies involved in digitization work. A complete market survey was carried out and a visit to the libraries where the process was in progress was organized in July/August 2005, after the constitution of the project team.

We had to make a selection of the digital software as it is essential to have a robust and flexible digital library management and preservation software for creating as well as delivering the digitized

---

contents. The best available choice for us was the open source software, basically Dspace and Greenstone. But after a prolonged discussion about the software it was asserted that the retrieval should be interlinked with the bibliographic database, therefore it was decided to digitize our resources, with LS Digital software-DRMS (Digital Resource Management System), which is linked with Libsys. The unique feature of integration of LS Digital with our Libsys software offers users a single interface through OPAC for accessing bibliographic records and digital contents. This system supports both window and web interface.

## **8. Mega Digital Library Project and GSL**

Soon after we finalized the digitization project at GSL and were moving around to select a vendor/ or outsource, CDAC (Centre for development of advanced computing) Noida contacted us relating to the digitization in August 2005. After successive discussions between the GSL and CDAC, it was finally agreed that the rare and fragile as well as the special collection published before 1950 will be digitized by CDAC, whereas the institutionally produced documents will be digitized in-house. The library professionals were directed to get acquainted with the digitization technology while working with the CDAC team for gaining practical insight and necessary skills for the purpose. The necessary specifications for the equipments to be procured by the GSL will be provided by the CDAC.

Finally signed a MOU between LBSNAA and CDAC, for rights to digitize books, journals from GSL for digital library mega centre project was signed on 8th November, 2005. GSL was recognized as one of the sources of multilingual data for digital library creation. The ultimate aim of the project is to digitize identified books, magazines and journals and make use of multi-lingual content for research purposes in OCR test bed and create other digital library tools. Only those books and resources, which are free from copyright will be uploaded on the web. The documents, which are useful for the project are identified by CDAC and LBSNAA experts. A copy of the digitized CDs DVDs will be provided to LBSNAA free of charge.

The Mega Digital Library Project was envisaged by Carnegie Mellon University USA as "Million book Universal Digital Library (UDL) Programme". CDAC is the agency for scanning documents for this project. This project is coordinated by Indian Institute of Science, Bangalore and is supported by Ministry of Communication and Information Technology, GOI. This initiative covers scholarly material, like books, journals, manuscripts, conference papers and reports. The digitization of materials for this initiative is carried out in a number of organizations spread over India. CDAC started the scanning at GSL in November, 2005 and since then about 3, 626 documents covering more than 3.1 lakh pages have been digitized.

## **9. Digitization Process**

As a process towards making a digital library the CDAC have set up a digitization station at the GSL with the following H/W:

- Zeutschell Omnican 5000 OS Scanner
- 3 P- IV Computer Systems with Hard disc capacity of 256 GB
- 3 KW Battery Backup
- DVD Writer

---

The source documents are transformed to bit-map images by scanning and during image capture these documents are read at a predefined resolution and bit-depth. Scanning resolution is chosen to give acceptable readability. The digitization process uses the following techniques for documents digitizing:

### **9.1 Scanning and making Image files**

The documents are scanned at 300 dpi resolution or above, depending on the physical condition of the document. At the time of scanning the resolution, proper centering of documents etc. are very important aspects to keep in mind. There are various file formats, like Tiff, Gif, bmp, etc. for storing the scanned images. The scanned images are stored in their original uncompressed O TIFF (Original Tagged Image File Format) format. However, the charts and illustrations are scanned separately and then merged with the text at the appropriate location. These uncompressed images are further subject to compression in order to reduce the size of the file for processing, storing and transmission of digital images. The use of Tiff format results in high quality image saving as it does not make any loss to the original image and moreover conversion from Tiff to other formats is quite easy and straight forward. For displaying the digitized contents via the web proves to be ideal with Jpg and Gig formats.

### **9.2 Image Processing and Batch Processing**

Image processing is the process after the images are scanned. Professional image cropper software is used for processing which helps in the elimination of blurs and dark spots or unwanted noise level within the text areas. It also helps in reducing the size of the image thus saving the disk space.

After the images are processed, the whole batch is processed with the help of scanfix 4.21 software, to enhance the quality and clarity of the scanned images; margins and text clarity is maintained. The noise removal option of the software is applied for the image that still has the noise in the text area. The intelligent crop option of the software is used to maintain all the margins i.e. top, bottom, right and left equally and giving the original shape to the image.

### **9.3 OCR**

The next step which is important task is to rectify the errors and make text files error free with the help of optical Character Recognition (OCR) process. This process also involves proof reading and checking quality of text files. When a document is scanned it is an image file and it is not possible to search an image file, therefore it needs to be converted to a text file to enable searching and indexing possible. For OCR we are using the Abbey fine reader 6.0 v. as it is considered the best OCR software for its accuracy of recognizing the characters. After applying the OCR the complete document is saved in Text, Rich text, etc. The bibliographic citation is developed in Dublin Core Metadata format for the document for subsequent retrieval by OPAC

### **9.4 Preparing for access and retrieval**

As PDF image files are not searchable, therefore HTML format was chosen as the best format to provide web operability and readability as it is easy for conversion, has small file size, readability by all browsers, and most important search ability. The files saved in TXT, RTF etc are converted to



HTML files; along with complete tagging at least up to chapter/ section level as applicable, which will enable the database to be integrated with any web application. In such types of documents, a full text search is feasible and advanced search like search within the search and Boolean operators etc can be built or incorporated easily. Hyper linking of HTML files with each document structure is a prerequisite for retrievals, as for easy and step to step access to digitized information of the whole document. As each heading and subheading appear in different pages and different chapters, these are linked to the content page of the digitized document (if the user wants to refer to a subheading in say chapter 9 of the book he can just go to the content page and click on the heading/ sub heading thus leading directly to the required page). Hyper linking, therefore, provides quick accessibility to the required data/ information with CD-ROM based or web based access.

## 9.5 Metadata

Metadata is "data about data," or information about an image. It's critical to capture as much information about the image in addition to the image itself. Metadata has to be preserved along with the image. Metadata is one of the key aspects for information access and is provided by the GSL professionals. A sample metadata encoding is shown below.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <dublincore>
<language>English</language>
<title>Soonderbai Power</title>
<title1 />
<title2 />
<creator>Storrie Kate</creator>
<creator1 />
<creator2 />
<creator3 />
<creator4 />
<keywords />
<keywords1 />
<keywords2 />
<keywords3 />
<description />
<subject />
<subject1 />
<subject2 />
<publisher>Pickering And Inglis London</publisher>
<contributor />
<contributor1 />
<date>1886/00/00</date>
<documenttype>Print - Paper</documenttype>
<documenttype>Book</documenttype>
<format>Tagged Image File Format</format>
<identifier />
<source>LBS National Academy Of Administration</source>
<relation />
```

---

```
<coverage />
<rights />
<copyrightdate>0000/00/00</copyrightdate>
<scanningcentre>Cdac, Noida</scanningcentre>
<scanningnumber>OS 5000</scanningnumber>
<digitalrepublisher>Digital Library Of India</digitalrepublisher>
<digitalpublicationdate>2006/08/04</digitalpublicationdate>
</dublincore>
```

## 10. Staffing and Training

Digitization needs the dedicated staff and highly motivated; and each of their role must be determined. Staff is also essential to promote the digitization collection. Regular training to the staff involved with digitization activities are required to impart periodic training and the newer technologies of digitization. It is necessary that the professionals acquire and understand digitization technologies and skills.

## 11. Integrated Access Management

With the help of a technically sound IT infrastructure at the LBSNAA the GSL is fully equipped to build a digital library to serve the users effectively and efficiently. The prerequisites for retrieval at the GSL are that; it should be fast and deliver the page image for reading and printing; it should be able to perform full text searching and it should be Browser and web based. Academy is moving for a paperless environment. Moreover, tight schedule of OT's/ Faculty hampers their frequent visit to the library. Therefore web based digitized data may be of immense help to them in academic pursuit. For accessing the digitized contents, a separate comprehensive and interactive library website will be created. Accordingly the digital contents will be linked to the bibliographic database, where the E-resources option can directly link to the full text of the digitized data. To meet this aspect the GSL have recently upgraded the libsys4 management software to the web-enabled LS Premia and for linking of full text of digital content to the bibliographic details LS digital have recently been procured.

Several issues are involved in access to digital libraries ranging from the design, information retrieval and user interface issues. Digital libraries should be cautious in the design of the user interface as this is the first thing that users notice when accessing a library's website. The user interface is a major area of concern. Unlike traditional libraries where users can walk through the library, see other people reading, browse through the shelves, or can talk to the library staff, in digital libraries the only communication link between the user and the collection is the user interface [6]. The GSL digital library will be judged on the basis of merits of the user interface. It will provide all necessary capabilities yet be easy to understand and easy to use. The choice between three kinds of search; direct search, simple search and fielded search will be maintained. Dillon [7] has listed few questions that the DL designers need to address:

- What will bring a user back to our resources again?
- How do I build an interface that supports a richer comprehension or appreciation of the contents?
- What makes material more learnable by users?
- Can novices learn for viewing an experts' construction of information space?

---

---

## 11. Conclusion

Digitization is costly affair and also it is an additional cost for the libraries to bear amidst budget cuts and decreased staffing. Digitizing does not replace the need to perform all the traditional tasks like acquiring, organizing, cataloging, and preserving materials, but it results in increased expectations to the user needs. The libraries of today have changed a lot and they become dynamic centers of information resources rather than mere storehouse of materials. Development of Digital Library technologies is enabling and enhancing the standards of information dissemination and also the quality of life as a whole and in future they should act as depositories of knowledge. Present day library services should cope up with the new operating environment and content. They should overcome the inhibitions and look ahead for the betterment of information services to the user community by using digital technology. The need of the hour is to keep pace with world. It seems that the days may not be far when the whole world would have digital libraries interconnecting all libraries to meet the academic and research needs within the short time. Safeguarding digital heritage is a major issue, especially for national libraries, because of their legal task of preserving the national heritage of a country.

## References

1. Reid, Gordon (2000). The Digitization of Heritage Material: Arguing for an interpretative approach based on the experience of the Pawys digital library project. Program Vol. 34 (2), 2000. pp.143-158.
2. Byrne?, Alex (2005). Promises and Challenges: Digital libraries in the global information society. Available at [<http://www.ifla.org/IV/ifla71/papers/176e-Byrne.pdf>]
3. Murthy, Sudha S (2005). Digitization: A practical Experience at the National Tuberculosis Institute, Bangalore. Information Studies, Vol. 11 (2), 2000. pp. 109-125.
4. Witten, I.H., Bainbridge, D and Boddie, S.J. (2001). Power of the People: End user building of digital library collections. In proceedings of the First ACM/IEEE-CS joint conference on digital libraries, New York, ACM Press, 2001. p.96
5. Tanner, Simon. (2001). Librarians in the Digital Age: Planning digitization projects. Program, Vol. 35 (4), October 2001. pp 327-337.
6. Chowdhury, G. G. and Chowdhury, Sudatta (2000). An Overview of the Information Retrieval Features of Twenty Digital Libraries. Program, Vol. 34 (4), October 2000, pp. 341-373.
7. Dillon, A. (2002). Technologies of Information: HCI and the Digital Library. In J.M. Carroll, (Ed). Human computer Interaction in the New Millennium. Boston: ACM press, 2002. pp. 457-474.