DIGITIZATION OF MANUSCRIPTS AND RARE LITERATURE: INITIATIVE OF ARCHIVAL CELL, PANJAB UNIVERSITY, CHANDIGARH (INDIA)

MRITUNJAY KUMAR **NEELAM SHARMA**

Abstract

Digitization in Indian setup is a bigger challenge than it appears. This is true in the light of several issues governing the process. Some of these includes, monetary factor, bureaucratic setup, absence of the policy-frame, unavailability of the skilled labour, and above all the absence of the environment in which these needs to be carried out. Still, to create a definite space, in this age of Information Communication and Technology (ICTs), Government of India in their own way, through the support of its ancillary bodies is trying their best to create a niche for their users, so that they could stay in the race in this fast changing and growing world of information technology. In the area of Manuscripts, Department of Culture, GOI made an ambitious plan in 2003 to preserve, conserve and possibly upgrade the same to the level of electronic excess through its digitization and digital preservation. To achieve the above objective, National Manuscript Mission was constituted in 2003 to accomplish the desired objective in 5 years time. In this paper, effort has been made to capture the NMM Guidelines and how Panjab University, Chandigarh is utilizing the same to digitize its multilingual holdings. The paper starts with the standard definition of digitization and based on the constraints, particular to the Indian scenario, it has focused upon the policy frame which came out with the findings of the National Manuscript Mission.

Keywords: Digitization/ Manuscripts/ Rare Literature/ Archives

1. Introduction

Digitization means acquiring, converting, storing and providing information in a computer format that is standardized, organized and available from demand from common system.1 But for many others, it is a means to convert an image or signal into digital code by scanning, tracing on a graphics tablet or using an analog to digital conversion device. 3D objects can be digitized by a device with a mechanical arm that is moved onto all the corners. The language could change but certain universal will remain the same for undertaking digitization around the globe. The digitization, as I have maintained in my abstract in the developing world is a challenging task owing to the

specific problems of these nations in particular. Some of these include the financial problems, bureaucratic hassles, no specific guidelines, unavailability of the desired work-force (though many from these parts of the world work for the leading nations) and above all the unavailability of the environment in which these have to be operated. Despite these problems in offing, Panjab University had started the process of digitizing its Manuscripts and Rare Literature which has both historical and cultural significance. The paper is an attempt to understand the recent initiative of National Manuscript Mission (a body constituted by Department of Culture, government of India) and in their light how Panjab University is approaching this problem.

2. Manuscript and Rare Literature at Panjab University

Panjab University Chandigarh has the rare distinction of representing a plural culture of different nationalities existing in the region. Looking back into the history, the University originally had been established in Lahore (present Pakistan) was the 4th University of India and was the hub of education, during this period, in the entire northern territory of the colonial India. With the partition of Colonial India into two independent nations as that of India and Pakistan, the University in India got its own campus in 1963 as Panjab University, Chandigarh. Panjab University, owing to its unique history, developed a sizeable collection of literature (Rare Collection), which represents the aspect of changes and continuity in this period. The crucial question of digitization of the university heritage came up in 2004. This was essential largely because of the importance of these holdings having both cultural and historical significance. This also explains in short the need and significance of the colonial government of India to open such University in its northwest province, and also the character of the holdings to a great extent. This partly explains the rationale of digitizing the collection. The paper is also important as it highlights the process of digital techniques being followed in the developing world, especially India. The technique in India became operational, largely due to the efforts of National Manuscript Mission (Hereafter NMM).2

3. University Archives and its Importance

As early in 1982, the University Grants Commission has requested the Panjab University to open an archive having its office in main library. In response to the UGC circular and the importance of its holdings the University had taken the decision to open an archive for the upkeeping of its numerous collections. In its holding, we have a total of 1492 manuscript. These are in Hindi, Urdu, Persian, Punjabi, Sanskrit and also in Sharda Script. It covers a range of subjects like Persian Court Etiquette, Poetry, Writings of the Sikh Gurus and various translations. Alongside this, we also have the government reports and other items as part of general archival trends.

4. Digitization Trends and the Government of India

Digitization in a developing country is a challenging job owing to the unavailability of the skilled labour force, efficient technology, finances and above all the working

environment suited to the job. Still, the requirements have prompted the authorities to take few measures in this area to match the global challenge. In India, The National Manuscript Mission for Manuscripts was launched in February 2003 by the Department of Culture, Ministry of Tourism and Culture, Government of India, to save the valuable but less visible, of our cultural inheritances as that of Manuscripts3. An ambitious five-year project, the mission seeks not merely to locate, catalogue and preserve India's manuscripts but also to enhance access, spread awareness and encourage their use for educational and research purposes.

5. NMM Guidelines

The National Manuscript Mission (NMM) has the primary objective of using digital technology to preserve the manuscripts for posterity. The working group of the mission in its first meeting observed that there are no digitization standards thus far available those the mission in its massive digitization initiative can adopt. In this regard NMM in association with National Informatics Centre, a premier Govt. of India IT organization, jointly studied the several digitization projects4 being undertaken by world libraries, museums etc. In depth study of the digitization processes of these organizations has helped in developing a viable standard. This study, however, addresses the standards for creating archival quality digital still images of manuscripts. These guidelines are meant not only for the archives, but all those organizations planning to digitize their collection. The guidelines affecting image quality file formats, storage and access standards for images. Guidelines5 are prescribed to maintain consistency, high quality scan and meeting the global standards. It is advisable to use these guidelines to create images of long-term usage and reduce the need to rescan material. Some of the specific guidelines of the same as follow:

- Digitization standards for audio and video recording
- Born digital material
- Conservation process for preparation of material for digitization
- Scholarly work required deciphering manuscripts like folio numbers etc.

These standards are aimed at decision makers, library managers, and curatorial and technical staff members. The reasons for implementing a digitization project, or more precisely for digital conversion of non-digital source material, are varied. The preservation of our cultural heritage is one, but it has many more purposes which are associated with the day to day work, administration and based upon the user requirements. Some of the immediate benefits which one can have from the same as follows;

- It increases the access: This is the most obvious and primary reason, where there is thought to be high demand from users and the library or the source has the desire to improve the access to a specific collection.
- It also improves the services to an expanding user's group by providing enhanced access to the institution's resources with respect to education, life long learning etc.

- It also reduces the handling and use of fragile or heavily used original material, for instance, manuscripts and create a "back up" copy for endangered material.

6. Process of digitization

The process of digitization requires a number of components; which needs to assessed at the beginning of any such planning. Some of the important component, which the NMM has issued in its proceeding, is selection policy, conversion, quality control programme and the issue of collections management.

6.1 Selection policy

In selecting the source material for digitization, three basic questions are to be addressed. These include; whether the source material needs to be converted? Should be converted? and can be converted. The selection therefore has to be converted in such a way that it will assure that only those issues like the value of the selected material and in its content are considered but also demands concerning technical feasibility and institutional requirements. However the issues involved in selection of material will be examined from two perspectives:

- Principal reasons for digitization
- Criteria for selection

In response to the above question; we get an answer that the principal reasons for the digitization could be, for enhanced access, to facilitate new forms of access and use or for the preservation of these invaluable materials. In the first case, it is basically for the research purposes that we enhance its access, that too in the light of growing demands. In the second case, the main purpose is to enable the use of original manuscript or rare literature that cannot be consulted in its original form other than by visiting its specific repository, and for manuscripts that has been damaged and where technology is needed to reveal its content and shape. As regards the third case, i.e. for preservation, the purpose is, in the first place, to create accurate reproductions of the original manuscripts on a long lasting medium. These reproductions, however, needs to satisfy both users of today and future potential users, and must therefore be of high quality and should also posses a physical stability that can be maintained over time.

In deciding the criteria for selection, issues like content, demand, and the condition plays pivotal role. Regardless of the purpose for implementing a digitization project, the selection of the source material will always be more or less content driven. The level of demand is of course of great interest when selecting the source material for digitization. Involving scholars and other researcher in the original decision is therefore a traditional selection methodology. As regards condition, selection of material for digitization will be affected both by its physical condition and by existing quality. Material, which is fragile, damaged and in poor condition may present too many risks of further damage being caused by handling to allow it to be scanned without special care, or some basic conservation techniques. Similarly, if the material being considered as a candidate for

digitization lacks detailed cataloguing or descriptive data, it is essential for future access to such material to create such data, and it will therefore need to be considered whether the necessary costs of doing this can be included in the overall budget of the digitization project.

6.2 Conversion

A digital image is an "electronic photograph" mapped as a set of picture elements (pixels) and arrange according to a predefined ratio of columns and rows. The number of pixel in a given array defines the resolution of the image. Each pixel has a tonal value depending on the level of light reflecting from the source document to a charged-coupled device (CCD) with light-sensitive diodes. When exposed to light they create a proportional electric charge, which through an analog/digital conversion generates a series of digital signals represented in binary code. The smallest unit of data stored in a computer is called a bit (binary digit). The number of bits used to represent each pixel in an image determines the number of colours or shades of grey that can be presented in a digital image. This is called bit-depth.

Digital images are also known as raster images to separate them from other type of electronic files such as vector files in which graphic information is encoded as mathematics formulas representing lines and curves.

Source documents are transformed to bit-mapped images by scanner or digital camera. During image capture these documents are "read" or scanned at a predefined resolution and bit-depth. The resulting digital files, containing the binary digits (bits) for each pixel, are then formatted and tagged in a way that makes it easy for a computer to store and retrieve them. From these files the computer can produce analog representations for on-screen display or printing. Because files with high resolution images are very large it may be necessary to reduce the file size (compression) to make them more manageable both for computer and the user.

When a source document has been scanned, all data is converted to a particular file format for storage. There are a number of widely used image formats in the market. Some of them are meant both for storage and compression. Image files also include technical information stored in an area of the file called the image "header".

The goal of any digitization programme should be to capture and present in digital formats the significant informational content contained in a single source document or in a collection of such documents. To capture the significant parts, the quality assessment of the digital images have to be based on a comparison between those digital images and original source documents that are to be converted, not on some vaguely defined concept of what is good enough to serve the immediate needs.

6.2.1 Input specifications

- The input documents are manuscripts/rare books of generally A2 –A4 size.

- These manuscripts/rare books are available at various repositories/libraries located at different parts of the country and need to be digitized at the Digitization Centre to be located at site. In the case of Panjab University, it is the Central library.
- Manuscripts of our collection are primarily available on paper (various types), Palm Leaves, Cloth, Clay Tablets, Parchments, Tamra Patra, Wooden Covers, Ivory Covers/Sheets, Wooden beads, Scrolls, dear Skin, Micro film etc. However, the Rare Books/Literature is mostly in the Paper form.
- In the case of manuscripts, they are generally very old and brittle and need special and sophisticated handling techniques, primarily archival in nature.
- Again, some manuscripts are having illustrations/charts created using ancient inks, vegetable dyes, metals such as silver, gold etc. they are very likely to get oxidized with the effect of bright light and heat.
- All the pages of the manuscripts shall be numbered before scanning, if not already numbered.

6.2.2 Handling and Preservation of Manuscripts/Rare Literature There are a number of archival specifications necessary for these materials before getting them digitized. Some of them as follow:

- Manuscripts/Rare Literature taken up for digitization should have undergone conservation process as necessary.
- Standard conservation techniques are to be used for cleaning and to increase the legibility of the documents as and where necessary.
- These are handled in the best way as prescribed by scholars. Therefore, while digitizing the same, the effective guidelines from the experts and the working scholar in the field may be solicited.
- In general, binding is not allowed to be taken out as it any damage the literary text, however, in some cases where it is absolutely necessary, due care shall be taken to remove them and rebind them using prescribed archival guidelines in this regard.
- Scanner/ camera operators should bear surgical gloves so as not to damage any of these vital documents.
- Soft bristled paint brushes to be used to wipe away the years of accumulated dust and dirt as necessary.
- Long, horizontal format requires special handling considerations
- To maintain the sequence of loose leaf local archives and archival helps are to solicited to remove the thread, enumerate the pages and record the missing folios and rethread them after digitization.

- As regard language issue, the experts having the knowledge of grammar and script may be invited to read and verify the contents of the script.

6.2.3 Image Capture There are two important issues involved with the capturing of image in the process of digitization. They are:

1. Selection of Imaging Equipment
2. Maintaining the requisite image quality

- Selection of the imaging equipment has an important impact on the quality of the image. Equipments from different manufacturers can perform differently, even if it offers the same technical capability. Face up scanners/digital camera or any other non-touch device shall be used as they might harm the original state of the documents. Flat bed and other touch devices shall not be used as they might harm the original state of these precious documents. While selecting the imaging equipments the cost plays a crucial role. The total cost of imaging not only involves the cost of the capture device but also associated peripheral devices, lighting equipments, labour cost, processing equipments, storage cost etc.

  The workflow of image capturing, processing, and storing should be automated to reduce the cost. Face up scanners with low processing time can be better choice. While in digital photography flexible lighting arrangements can make the object better lit.

- Image Quality at capture can be defined as the cumulative result of the scanning resolution, the bit depth of scanned image, the enhancement processes and the compression applied the scanning device or technique used, and the skill of the scanning operator.

6.3 Quality control

It is an important component in every stage of a digital imaging project. Without this activity it will not be possible to guarantee the integrity and consistency of files.

6.3.1 Methods : The automated image evaluation tools that are available today are normally not sufficient for materials that are required for cultural and scientific purposes. Therefore, visual quality evalution has to be done. The same could be done the following way;

- On-screen
- Print-outs

However, the technical limitations that can effect the evaluation must be considered, beginning with the possibilities of getting good quality printed hard copies of grey scale and colour images. Some of the highlights of the On-Screen Evaluation as follow;

- View scanned images at 1:1 (100% enlargement).

- Use of target to evaluate grey scale and colour reproduction.
- Use resolution targets and histograms to evaluate spatial resolution and tonal reproduction.
- Specified image quality control as prescribed by NMM for different types of the document is must.
- Moreover, the specifications of the Print-Out Evalution consider;
- Examining hard copies created from the images to see if they fit the quality requirements.

Also, compares the printouts with the source documents.

6.3.2 Scanner Quality Control : Before a scanner is bought, vendors should be required to deliver measurable digital results from relevant digital image quality evaluation tests. When a digital image project is running, scanning quality control measures must be set to enable operators to ensure that the scanning device is operating within anticipated tolerances. Issues of main concern in performance are; spatial resolution, tonal reproduction, colour reproduction and noise.

6.4 Collection Management

The possibility of being able to use a collection of digital images in the way it was intended depends not only on conversion standards and quality control, but also on how efficiently the collection is managed. The purpose is not only to meet the short term needs but also to provide accessibility over time. Therefore, the steps have to be taken to satisfy both current use and the expectations of the future users. Hence, the plan must be made, for instance to;

- Make scanned images appropriate to the ultimate intended use.
- Upgrade distribution of images and user interface functionality.
- Transfer images to new technical platforms to meet increasing capacity for processing and handling of digital information.
- Migrate digital images to new file formats or physical media to ensure long-term accessibility.

To make scanned images usable, further steps are require to store them properly. All image files that are produced by a digital image project must be organized, named and described in a way that fits the purpose of the project.

6.4.1 Organization of Images : Before a name and a description of an image file is considered, decision is required as to how it should be stored. Normally, the source documents being scanned are physically organized according to the principles of library management. Collections of manuscripts generally have numbers given by the library

or repository where they are stored. So the organization should be in such a manner that just by looking at the name one should tell about the manuscript digitized.

6.4.2 Naming of the Images : There are two approaches to name a document: (1) to use a numbering scheme that reflects numbers already in use in an existing cataloguing system, or (2) to use meaningful names. In digitizing the collection, both the schemes are valid, and the criteria should be 'what fits a type of collection or source document" should be chosen or applied. For example, in the case of manuscripts, the images will be named according to Manus Id generated by the cataloguing software of the mission called Manus e-Granthavali. Meta Data information for each manuscript scanned will be stored in the database and is identified by its Manus Id. So the Manus Id and the Accession Number given by the institute where the digitization process is taking place forms the basis of naming the digitized images of each manuscript page.

6.4.3 Description of Images : To describe digital images there is a need for metadata that is structured data about data. Metadata can also be defined as data that facilitates the management and use of other data. Metadata information according to Dublin Core standards is stored through the specified software, for instance, manus granthavali software for each manuscript.

The highest quality file produces is referred as Digital Master. These files are created as a result of image capture. Master file represents the original manuscripts as close and correctly as possible. Derivative images are generated from the master image using photo editing software like Photoshop.

6.4.5 Output Specifications of Images : The detail specifications of the images discussed above may be as follows;

- Master Image (Original Uncleaned and Uncom-pressed)

File Format	:	Tiff latest version
Compression	:	Uncompressed
Spatial Resolution	:	300dpi
Subject Metadata	:	As per standards fixed by NMM
File Naming	:	As specified

- Clean Master (Cleaned Image)

File Format	:	Tiff latest version
Compression	:	Group 4 CCITT Compression
Spatial Resolution	:	8" X 10" at 300 dpi
Subject Metadata	:	As per standard fixed by NMM
File Naming	:	As Specified

- Access Image (Derivative Image)

File Format	:	JPEG latest version JBIG (in case of black and White)
Compression	:	Group 4 CCITT lossy compression
Spatial resolution	:	1024 x 786 pixels
Subject Metadata	:	As per standards fixed by NMM
File Naming	:	As Specified

- Thumbnail

File Format	:	JPEG latest version
Compression	:	Group 4 CCITT lossy compression
Spatial resolution	:	1" x 1"
Subject Metadata	:	Nil
File Naming	:	As Specified

7. Metadata creation

Metadata as we know is the data of the data and is equivalent to traditional cataloguing system. Thus, it will contain all the information related to a document in details. Thus, here for information we should know that there are two kinds of metadata which we are creating in the process of digitizing our collection. They are: Subject Metadata and the Technical Metadata.

7.1 Subject Metadata

Subject Metadata is created according to the Manus Data record using Manus Granthavali Software of National Informatics Centre (NIC). This software is already in use with in the NMM. Panjab University is also using this expertise for our collection. Some of the components of the Subject Metadata as follow:

1. Manuscript Number
2. Title
3. Other Title
4. Author
5. Organization
6. Commentary
7. Commentator
8. Scriber
9. Language
10. Script
11. Complete/Incomplete
12. Subject
13. Bundle Number
14. Folio Number
15. Pages

16. Material
17. Missing Portion
18. Illustrations
19. Condition
20. Catalogue sources
21. Remarks
22. Manuscript date
23. Manuscript length (in inches)
24. Manuscript width (in inches)

7.2 Technical metadata

Technical Metadata is that which describes the features of the digital file. Technical Metadata is automatically generated and assigned to the image file at the time of creation. Some of the important components of this metadata as follow;

File Name	:	Assigned at the time of scanning
Date Created		
Date Modified		
Image Format		
Width	:	Pixels
Height	:	Pixels
Colour Mode	:	RGB, grayscale
Resolution	:	Pixel per inch
File Size		
Colour Profile	:	ICC colour profile
Make	:	Scanner Software
X Resolution	:	Scanning resolution in the X axis
Y Resolution	:	Scanning Resolution in the Y axis
Resolution units		
Software	:	Imaging software

Note: one can view these technical metadata using Photoshop.

8. Digitization in the Panjab University Chandigarh

Panjab University Chandigarh (India) started the digitization process in the year 2004. The necessity of such an approach became true some because of few important factors. First and the foremost was the importance of the collection. The second was the process of physical preservation, as applied in the case of archival documents, though were in the practice, still this methodology of archival preservation failed to provide a solution to the growing demands of these documents for research and other ancillary purposes.

Third and the most important factor were now we had certain guidelines from the NMM for digitization and digital preservation.

Thus, based upon the guidelines, specified above, the university started digitizing its collection. At the outset, we had two goals, first, to harness the knowledge embedded in the manuscripts and the rare literature. Second was to preserve and conserve the cultural heritage of our nation through enhanced digital longevity. The outcome of the practice is visible on the ground. Research use of the manuscripts and the rare literature has increased manifold. These users are both national and international. A number of scholars from Europe, US and especially from the Arab countries visit our repository every year. Though we have not put our digitized collection on web, still, we hope that in the coming years research use of these collections may be more. The making of these resources more visible, we are planning to put them on the web using D-Space Software. However, this issue will come up only when we will complete the stage one.

9. Conclusions

Panjab University Chandigarh had started its digitization based on the guidelines issued by the NMM. This, in my opinion, is not only the case of India or Panjab University in particular, but of the developing nations at large. However, there are issues like, government apathy, attitude of the people and overall financial problems are coming in the way of achieving the target. Also, the issues like digital security, problems associated with the data migration, issues of intellectual property right are coming in the way, at the technical level to achieve the target. The factor is because of the absence of the proper digital environment. Despite all this, we are hopeful that in the years to come developing world may be able to overcome all this, with the regular interactions with the communities close to the digital environment. This is important in the light of the growing demands and the utility of these sources for research and other ancillary purposes.

References

1. National Mission for manuscripts: Guidelines for Digitization of Manuscripts, published by NMM Press, Department of Culture, Government of India , 2003.
2. Public Records Act 1997, Government of India.
3. Guides to the Records, National Archives of India, 2001.
4. Bernard, Richard C. The Arrangement and Description of manuscripts, American Archivist 23, 1960.
5. Massis, Bruce E, Practical Library manager, Ane Books, New Delhi, 2006.
6. Balram, Shivanand and Dragicevic, collaborative Geographic Information Systems, Idea Group Publishing, Hershey, 2006.
7. Dwivedi, Suresh Chandra, Digital Library (V.1), Shree Publishers, New Delhi, 2005.

8. Dwivedi, Suresh Chandra, Digital Library (V.2), Shree Publishers, New Delhi, 2005.

9. Sharma, Sarabjeet and Gopal, Krishan, Right to Information: Implementing Information Regime, Authors Press, Delhi, 2006.

10. Singh, R.K. Basic Facts of Digital Library, Shree Publishers, New Delhi, 2004.

11. Craft, Sir Alfred, Review of Education in India, GOI Publication, Calcutta, 1886.

BIOGRAPHY OF AUTHORS

Sh. Mritunjay Kumar is presently working as a Assitant Archivist at Punjab University, Chandigarh. He holds M.A, M.Phil (JNU), Associateship in Information Science (NISCAIR-CSIR), Diploma in Archives (National Archives of India), Certificate in Intellectual Property Rights (World Intellectual Property Organization, Switzerland. His Areas of Interest are Traditional Archives, Web Archiving, Content Management, Teaching, Application of Information tools in Libraries and Knowledge Centres.

Email : mailmritunjay@yahoo.com

Ms. Neelam Sharma is presently working as a Deputy Librarian in Panjab University, Chandigarh. She holds M.A, M.Lib. Her Areas of Interest are All aspects of library management, organization of conference, symposia and seminars, specialized in digital library, digitization and automation tools, represented university library at various professional forums.

Email : neelamkuki@yahoo.com

(Footnotes)

[1] National Manuscript Mission (hereafter NMM), "