Status of Archives' Digitization at the National Archives of Pakistan

Manan Hafeez[1] and Muhammad Arshad[2]

Abstract

Purpose: The main purpose of this study was to explore the current prospects of archives digitization at the National Archives of Pakistan.

Design/methodology/approach: Qualitative research approach was used to collect the data. The fifteen employees of the NAP were interviewed. The recorded data was meticulously transcribed, after which the responses were qualitatively assessed using thematic analysis.

Research limitation(s): This study is limited to the National Archives of Pakistan only.

Key finding(s): This study found that National Archives of Pakistan hold various types of materials such as government records, old newspapers, manuscripts, Quaid-e-Azam papers, Fatima Jinnah papers, rare books, Muslim league records, and catalogs, is extremely valuable for digitization. However, this study highlighted various issues in the digitization of these documents included lack of technical manpower, lack of digitization tools, and lack of software.

Practical implication(s): This study provides an alternate medium of preservation. Lack of human resources, funds, awareness and high salaries of professionals and high cost of technology/equipment are the major hurdles in initiating an archives digitization project. The Federal Government is the main source of funding. IT, managerial & technical skills for archives handling and expertise to build a digital library are the basic needs to run such projects. Digital cameras or overhead scanners i.e. Book-Eye", the open-source software i.e. Greenstone, D-Space and E-print, and Dublin Core or MARC metadata standards are the most suitable options for archives digitization.

Contribution to knowledge: The major contribution of this study is to inform about the benefits of archives digitization which includes easily accessible, multiple copies that can be made, dissemination of knowledge, and online access to multiple users, transferability, and its long term preservation.

Paper type: Research.

Keyword(s): Digitalization; Archives; National Archives of Pakistan (NAP).

Introduction

All civilized nations maintained records of human actions and preserved them. Archives exist as proof of human activity and are important sources of information for historical study. Achieves are examined, obtained, accessed, processed, and reviewed to gather information, and they are of huge continuous worth, with the constant requirement to maintain the knowledge in case it is needed again. Manuscripts contain centuries of accumulated knowledge in a variety of subjects, including philosophy, science, literature, and the arts. According to Guar and Chakraborty, manuscripts are more than just historical records. They are a synthesis of earlier tinkers' and thinkers' combined knowledge and experience (Guar & Chakraborty, 2009).

The digitization of archives is important for a variety of reasons in today's world. It is needless to describe all of its advantages; it is useful for data storage, preservation, and accessibility. It now has a large following. This study will look at the National Archive of Pakistan's present digitalization trends and future opportunities. The benefits of digitization programs are being recognized in industrialized countries, and these countries have conducted extensive research on the digitization of manuscripts and archives. However, this initiative is still in its early stages in Pakistan. In Pakistan, there have been very few attempts to digitize archives. Many information professionals undervalue the importance of archives digitalization

---

[1] National Electric Power Regulatory Authority, Islamabad, Pakistan. Email: manan_hafeez@yahoo.com

[2] Department of Libraries, Islamabad, Pakistan. Email: marshadnlrc@gmail.com

activities. In the field of archive digitization, there is a need to develop professional knowledge. According to the aforementioned literature; no systematic analysis of Pakistan National Archives' digitization projects at the national and local levels has been done too far. With this in mind, it's critical to consider the perspectives of seasoned information professionals in this field.

A South Asian archiving paradigm has been created by NAP. It started with manuscripts from the Moghul dynasties and continued through the first year of government archives kept at the Pakistan National Archives. The information was separated into three groups: public, newspapers and magazines, and private collections. A private manuscript commission has been established by NAP. After a two-year hiatus, the Pakistan Archives, which was published by the NAP from 1984 to 1988, was re-launched in 1989. NAP is a member of the ICA (International Council on Archives) (ICA).

Research Objectives

The major objectives of this study are to explore the current digitization process/practice, technologies utilized, funding, access, and cooperation in the National Archives of Pakistan (NAP).

Review of the Related Literature

Infrastructure Required for Digitization Projects

The process of converting information or input data from any physical form, such as a print, picture, photograph, videotape, or other physical media, into a machine-readable digital form for computer processing is known as digitization. Digitization is the most prevalent way of transmitting data. The infrastructure listed below is required for digitizing the archives.

Hardware

- Computers with a Pentium IV, Dual 2 Core, and PCI Bus for data flow, Ethernet for data transmission, and RAM for loading, reloading, or creating a digital picture of various sizes and colors.
- Storage devices such as hard disc drives, removable hard disc drives, optical drives, and digital audio tapes for backing up and storing digital objects, as well as archiving and retrieving data.
- Sharpness and color clarity monitors are essential for producing professional-looking digital pictures.
- Digitization Scanners are used to convert photos, artwork, and slides into digital files. Images must also be captured with digital cameras before being downloaded to the computer.
- Printers, modems, CD writers, and other output devices.

Software

For digitalization, Chhatwal et al. (2009, p. 149) recommend installing software with the following capabilities.

- That which can modify pictures.
- Anything that uses page layout programming to combine text and images.
- That which allows you to exchange files across computers with ease.
- That which has files translation software that can convert files from graphics to text and vice versa.
- That which can compress files.

Digitization Efforts in Pakistan

According to Ahsan (2009), Pakistani libraries are still in the process of digitizing. Launching digitization efforts in Pakistan is difficult due to a lack of funding, computer literacy, and digitization expertise. Some private organizations and institutions are also digitizing. Social science institutes are digitizing manuscripts and other old texts to preserve Pakistan's cultural heritage. There are about 0.15 million Arabic, Persian, Urdu, Pashto, Sindhi, and Sanskrit manuscripts in Pakistan. A comprehensive approach for digitizing manuscripts is required. Khan (2010) claimed that when Pakistan was established in 1947, the subcontinent left no major manuscript collections behind.

Two record rooms, one in Lahore and the other in Karachi, contain a significant collection of government papers. About half of them are handwritten. After 1947, many new institutions arose, each acquiring a significant amount of manuscripts. There are 2 million manuscripts in Pakistan. These writings span a wide range of subjects, ideas, languages, and styles. Punjab University Library (2010) stated that the manuscript division was founded on July 31, 1920 and it now houses Pakistan's largest manuscript collection. This collection contains about 13167 manuscripts, including theses in Urdu, Arabic, Persian, and indigenous languages. As of September 2002, there were 22574 manuscripts in Hindi and Sanskrit, with more being added daily. Sanskrit and Hindi literature is written on palm leaves. Warraich and Tahira (2010) say the library has Pakistan's largest manuscript and archival collection. An old and rare manuscript in Arabic, Persian, Urdu, Turkish, Punjabi, Sanskrit, Hindi and other regional languages is said to be at the Punjab Library. According to Khan (2010), the digitization of manuscripts and archives is still in its infancy in Pakistan.

Few organizations have recently started digitizing manuscripts and documents. Pakistan could not make substantial progress in this field due to a lack of funds, facilities, and expertise. Sindh Archives, Pakistan National Library, International Islamic University, University of Punjab, Pashto Academy, and Pakistan's Higher Education Commission have all begun digitizing their collections in recent years. According to Khan (2010), the Sindh Archives in Karachi is a pioneer in digitalization. In 48 personal collections, there are almost 124,000 books. There are also 6134 letters from the Urdu language and literary greats dated 1933. GCU received 4970 letters and scripts of well-known short stories, poetry, and essays recently. The Higher Education Commission (HEC), according to Warraich and Ameen (2008), is critical in the development of Pakistan's knowledge-based economy. The Digital Library Program was launched in 2004 as part of the UI-based International Network for the Availability of Scientific Publications. Over 30 databases, 24,000 e-journals, and 45,000 e-books are available through the HEC National Digital Library. According to Qutab, Ullah, and Saqib (2017), the "Nuqoosh Research Center" (NRC) and museum collection is a one-of-a-kind experience for Urdu scholars worldwide to enhance evidence-based research possibilities in Pakistan.

Research Methodology

Personal, face to face interviews with fifteen National Archives of Pakistan technical experts involved in the digitization project were undertaken (NAP). The professional interviews were recorded using a digital sound recorder. Their answers were subsequently analyzed qualitatively. The interview performance covered all bases. Thematic analysis of semi-structured interview data was performed. The taped interviews in English were transcribed objectively. To avoid using qualitative tools, interview transcripts were hand summarized. Themes from the interview guide were utilized to organize textual material and to create tables of content analysis of interview transcripts. These were discussed utilizing transcripts, interpretations and comments. The research protected the respondents' identity by using quotations from transcripts without identifying the person.

Analysis and Interpretation of Data

Table 1 shows the persons interviewed, all were male and the place was Islamabad.

Table 1. Persons Interviewed

Sr#	Designation	Frequency
1	Deputy director	1
2	Microfilming officer	2
3	Junior microfilming officer	1
4	Microfilming assistant	1
5	Photo stating assistant	1
6	Audio visual assistant	1
7	Dark room assistant	7
8	Laboratory assistant	1
Total		**15**

Process/Practice of Digitization

Interviewees mentioned factors for choosing content for digitalization, such as national heritage. Preservation of material/document academic significance by minimizing harm to original age of material/document historical-cultural worth of resources. Some interviewees said it depended on the papers' significance. Most respondents said they had no notion or that the number of papers was uncountable. Some said 50,000 papers were digitized last year. However, few respondents responded that 200,000, 90,000, or 100,000 documents were digitized in the previous year. They have digitized the final 20-year-old newspaper, however, the quantity of papers digitized is unknown. The main subjects for digitalization are ancient newspapers, according to interviewees Quaid-i-Azam draughts Fatima Jinnah papers rare book most respondents said that digitization is done to preserve historical material for future generations. Most respondents said that the National Archive of Pakistan's microfilming and picture archiving division does digitalization. Several respondents said that the IT/Reprographic department handles the digitizing process. Most respondents said that the personnel engaged in digitization had adequate knowledge and don't need any training. Some respondents said that the National Archives of Pakistan does not have technological export for digitalization (NAP). All workers engaged in the digitization process have broad knowledge and have self-taught the task of digitalization.

Status of Digitization Technologies Used

The majority of respondents said that image-based digitization resolutions range from 300 to 600 dpi. The resolution, according to several interviewers, is 1200 dpi. In answer to this question, every respondent said that they always prefer the document's original color. The majority of respondents said that they favor JPEG file format for archive material digitization, with a few mentioning that they also utilize PDF file format. The majority of respondents said that they digitize Latin characters using image format, but a minority stated that they utilize both techniques (image and OCR). The majority of respondents said that they digitize Latin characters using image format, but a minority stated that they utilize both techniques (image and OCR). The respondents said they had digitized papers in whatever scripts or languages they were working with. The National Archives of Pakistan do not have a digital library. The majority of respondents said they don't utilize specialist software to digitize their materials/documents, instead opting for microfilming. The majority of respondents said they store digitized information on microfilm, CD-ROM/DVD, and external hard drives; nevertheless, some interviewees said they have to keep digitized content on the organization's server.

Funding and Access

The Federal Government is the primary source of financing for digitization initiatives at the National Archives of Pakistan (NAP), according to all respondents. The majority of respondents said that the government's funding for digitization initiatives are insufficient and that additional money is needed. The majority of respondents said they do not provide online access to digitized material to the public. They have not yet scanned all of the materials and are still digitizing them. The majority of respondents said that they provide photocopies of papers to academics and researchers' on-demand, as well as scanning and distributing them on CDs. They have a reference room on our grounds where researchers and students may go to examine our catalogue and index, and they can also get their chosen papers on demand. They also have a microfilm reader, and if a copy is required, they may print it using a microfilm printer.

Cooperation/Collaboration

The majority of respondents said that the material held by the National Archives of Pakistan (NAP) would not be shared with any other library or institution. Scholars and students, on the other hand, may get their requested documents by visiting the Pakistan National Archives in person. The majority of respondents said that any group that provides training for the digitization process would be beneficial to them. Some respondents said that they need technical knowledge or equipment to digitize their data. Future Development Professionals at NAP are being questioned about any national or international digitization initiatives they may be aware of. The majority of those interviewed said they were unaware of any national or worldwide digitization initiatives. According to several respondents, other provincial archives departments are also conducting a microfilming initiative to some degree, and Sindh archives are doing an excellent job throughout the province. The majority of respondents agreed that the research addressed all important elements of the digitalization process. According to several respondents, the current research has covered everything related to digitization and microfilming at the National Archives. The primary issue at the National Archives is a shortage of qualified personnel in the area of digitalization. The crew is entirely self-trained in microfilming and digitization, with no formal training from any Pakistani institution. They believe that, apart from the National Archive, Pakistan has no authority for digitalization. According to a few respondents, the research addressed all areas of digitization at the National Archives, and they did their best to respond to the questions properly.

Conclusion

Digitization is an excellent method for archival long-term preservation. Archives should be maintained in their original formats. The main barriers to launching such initiatives are a lack of money, a lack of qualified personnel, a lack of knowledge about such projects, and a lack of appropriate infrastructure. The main advantages of archival digitization include ease of access, the capacity to make multiple copies, knowledge diffusion, and online access to many users, transferability, and long-term preservation. The main sources of funding for such initiatives are international organizations, the Higher Education Commission (HEC), and the Federal or Provincial governments. Archives digitization initiatives are less expensive because of the use of open-source software and low-cost technologies. There are no copyright issues with internet access to digital records, and they should be made available. IT abilities, management skills, experience in technical handling of archives, and the ability to create a digital library are all needed to operate such initiatives. The technology required for archive digitization initiatives is accessible locally and may also be imported. Over-the-shoulder scanners, such as "Book Eye," open-source software, such as Greenstone, D-Space, and E-Print, and metadata standards such as Dublin Core and MARC are the best options for digitizing manuscripts. Only a few digitization initiatives have been started in Pakistan, i.e. the National Archives of Pakistan, the National Library of Pakistan, Government College Lahore, and Punjab University.

Recommendations

Based on the study's findings, the following recommendations are made.

1. Archives should be maintained through digitalization, with the original copy of the archives being preserved as well.
2. The government and relevant authorities should encourage archives digitization initiatives and give enough funding.
3. Library and information science experts should look for funding from international organizations such as UNESCO and USAID.
4. It is necessary to raise awareness about the significance of archival digitalization to preserve our valuable knowledge heritage.
5. A nationwide survey should be undertaken to determine the state of archive digitization initiatives.
6. Scholars and the general public should have online access to digital archives.
7. Training/workshops for archive professionals to improve their professional skills should be arranged.
8. Digitization techniques for archives should be taught in library and information science (LIS) courses.
9. Professional groups and institutions should share their digital experiences.
10. The government and relevant agencies should take the lead in preserving and disseminating our invaluable intellectual heritage. They should generously give the necessary money and resources in this respect.

References

Ahsan, M. (2009). Digital divide and digitization initiatives in Pakistan: A bird's eye view. *Nuova Cultura,* 515-520.

Chhatwal, A., Kanwal, P., & Lal, P. (2009). Digital heritage archiving in India: A case study of Punjab university library, Chandigarh. *Paper presented at the Conference INFuture: Digital Resources and Knowledge Sharing.*

Gaur, R. C. & Chakraborty, M. (2009). Preservation and access to Indian manuscripts: A knowledge base of Indian cultural heritage resources for academic libraries. *Vision and Roles of the Future Academic Libraries,* 91-99.

Khan, I. N. (2010). *Islamic manuscripts in Pakistan.* Retrieved October 7, 2010, from http://www.sindharchives.gov.pk/articles.aspx

Punjab University Library. (2010). *Manuscript Section.* Retrieved September 03, 2010, from Punjab University website.

Qutab, S., Ullah, F. S., & Saqib, F. (2017). Antiquity into Digital: A case study of Urdu literary magazine 'Nuqoosh Research Center (NRC) and Digitization Project. *Trends in Information Management*, *11*(1).

Waraich, N. F. & Ameen, K. (2008). Perceptions of the library and information science professionals about a National Digital Library programme. *Library Hi Tech News*, *8,* 15-19.

Waraich, N. F. & Tahira, M. (2010). Manuscript collection in the Punjab University Library: Assessing management issues. *Chinese Librarianship: An International Electronic Journal*, *29.*