Digitisation of Archival Records at National Archives of India and Department of Delhi Archives: A Detailed Study of Methods, Standards and Protocols

Akhlak Ahmad[1] and Swadesh Sharma[2]

[1,2]G.D. Goenka Univeristy, Gurgaon - 122 103, India
Email: [1]akhlaqa@yahoo.com

Abstract

The National Archives of India (NAI) and the Department of Delhi Archives (DA) have undertaken an enormous initiative of digitization of records to preserve and enhance the access of archival collection. Storage and maintenance of records in both forms physical and digital pose an immense challenge to archival science. The purpose of this paper is to illustrate how these institutions have employed methodology to improve the preservation and access to archival resources. The study analyses the methods, standards and protocols adopted in the process of digitisation to highlight how it can be sustainably preserved and retrieved digital information resources for generation to come. The data is collected through a detailed questionnaires and findings are based on the comparisons of different digitisation procedures, metadata preservation and retrieval of records. The efficacy of these massive projects has been studied in detail to put forward critical recommendations in a bid to provide best practices which can be used to achieve digital sustainability and interoperability among the digital information systems to enhance and provide long-term access to these historical digital resources.

Keywords: Archives, Digitization, Digital Preservation, Standards, Protocols, India.

Introduction

An archive can be defined as "a place where public records and other important historical records are kept and preserved".1 Further, the archive is the major source of primary information, it in-houses the non-current documents and preserves them for future generations. The history of the archival organization was traced to the 4th and 5th centuries B.C. when Athenians kept their valuable documents in the temple of Athens. Although the development of archives declined in ancient civilization and in the middle age but the early modern period, the European countries and the United States strengthen these organizations for the archival resources of national importance2.

India, being one of the oldest civilizations has the largest collection of the rare and valuable collection of manuscripts recorded in various forms along with the distinct languages and different scripts available across the country. These heritage resources are collected, stored, and preserved in libraries, archives, museums, and temples, etc. Records of Indian history belonging to pre-European period are in the custory of Indian states archvies, individual collections, and religious houses etc.3

Archival System Of India

The Indian archival system was started with collection of records of "Dutch and French Administration" in the year 1765 just after the battle of Buxar. The first office called "General Record Office" was established in Fort Williams in Calcutta collecting East India Company records but the breakthrough came in a system of records keeping after H.L. Smith Committee's recommendations and proposals and this record-keeping system is even being followed after the post-colonial period4.

In 1820, the British India government thought of establishing the General Record Office for managing judicial and revenue records. "The Imperial Record Department" was established 1918 in Calcutta and just after a, the Capital of India was transferred from Calcutta to Delhi which inevitably transferred the all

records of the Government of India to the new Capital of India. Since 1947, the department is known as the National Archives of India which is a repository of public records, oriental records, cartographic records, including private papers of eminent personalities including the Arabian and Persian manuscripts. The manuscripts and other resource collections makes India a rich documentary heritage and it provides invaluable source of information for the public administrators, research scholars, and other users5.

Objective Of The Study

The National Archives of India has a quantum of records generated by more than 500 government agencies including rare manuscripts personal collections of eminent personalities and the number is increasing rapidly. While Delhi Archives in-houses records of Delhi Residents, and the Commissioner's Office including documents of the history of Delhi City and special collection of a trial of Bahadur Shah Zafar trial. These two organizations are not merely the store house of the material related to India and Delhi but also the custodian of collective memory, heritage, culture, and historical information resources. Both the organizations have started the digitization of their resources at a large scale for preservation and outreach their collection to a larger audience across the globe.

The study seeks to explore and examine the practices and methods adopted by these organizations while preserving, digitizing, storing, and providing access to these resources. Further, it reviews the development of digitization and archival standards and practices that are followed for digital preservation and retrieval of resources by both the archives. This study seeks answers in the following areas:

•   The status of digitization and retrieval of digital resources;

•   To compare standards and protocols for metadata records and tools for preservation and retrieval used in the digital archive;

•   To understand the challenges of developing a digital archive and adopting standards.

Literature Review

Digitization has become the necessity for the libraries, archives, and information centres to preserve the analog information resources and make them retrievable to a larger audience. The digitization process ensures the preservation of original records and documents in their original form that can be consulted and viewed by future generations6.

The main objective of the digitization of archival material is to extend public access of its historical records to maximize efficient and effective use of resources achieving cost-saving benefits whenever possible. The success of a digital archive largely depends on connecting the clienteles both internal and global7.

The archives are dealing with the complexities of management of physical archival resources and it has been troublesome for decades. To deal with it, various standards for e-record management have been developed which is possible by using new information technologies, further, it helped in managing not only the record themselves but also the deal with their complex interrelationships. Now, an archive can deal with these challenges not only to maintain the e-records keeping in an efficiently and effectively for archival resources but also can also preserve them digitally and enhanced the access via Internet technologies for future generations without human handling of existing resources8.

The digitization of archival resources using standard practice opens up new opportunities for both users and archival professionals. It enhances the retrieval of collection and handling of records becomes simpler as compare to physical records. The powerful search engines enable to locate of resources with the combination of geographical and chronological dimensions from the user's interface and also in the manner the collection is managed and makes retrievable. Further, digitization also implies appraisal of archival

records and has a significant emphasis on the need for developing a flexible appraisal mode9.

On the other side, the archival community also faces challenges to cope with new guidelines and standards are being envisaged for metadata preservation, digitization, retrieval, and interchange of digital archival records among the archival community. To enhance the retrieval and to get the attention of impatrient user, the web developer have to adopt various new techniques. The ever-advancing Internet technologies force professionals groups and archival organizations to find an innovative and flexible solution to reach the end-users10.

Methodology

The study has opted a qualitative research approach to explore the digitization practices followed by the two archives based in Delhi. It is based on primary and secondary information resources gathered from books, journal articles, official records, and website of archives on the aspects of digital preservation, application of archival standards and protocols, and retrieval of digital resources. Also, it will explore multiple forms of data collection such as questionnaires, interviews, and literature including the website of these institutions With regards to this study, the two hypotheses have been framed i.e. Institutes do not differ significantly in strategies and policy adopted for digitization; and Institutes do not differ significantly in the usages standards and tools for metadata, file formats, image resolution, and retrieval/harvesting formats.

National Archives Of India (Nai)

National Archives of India (NAI) is a Central Government agency responsible for identifying records of various ministries, departments, offices including PSUs, Corporations, Commissions, Committees, etc. under the Public Record Act, 1993 and Public Records Rules, 1997. Currently, it has an archival holding of more than 47.8 lakhs of valuable and unique documents in the forms of files, volumes, maps, bills, treatises, and manuscripts, and 7500 microfilms rolls. The collection is broadly categorized into four categories such as public records, oriental records, manuscripts, and private papers.

National Committee of Archivists constituted in 1955 under the Chairmanship of the Director, National Archives of India (NAI) to address technical issues faced by the states archives and to achieve uniformity in archival practices through exchanging ideas and expertise. As a result states archives have evolved over the period on the pattern of the National Archives of India. NAI provides them all professional, technical support in the form of Archival training, sharing of resources and expertise besides extending financial assistance under two Grant-in-Aid Schemes. The NAI also works with the vision to preserve national heritage documentary for prosperity and working closely with national and international archivists and archives to encourage greater access of archival holders to make proper care use of documentary heritage11.

Besides the above records, the NAI has a special Archival Library supplementing the information available in public records. The library holdings comprise a collection of Government reports and publications, gazetteers, native newspaper reports, and books on various disciplines. NAI's Regional Office at Bhopal and Records Centres at Bhubaneswar, Jaipur, and Puducherry have also a huge collection of records with research facilities for the scholars visiting these offices.

Electronic Records Management

The advancement of the recent information and communication technologies (ICTs) has revolutionized archival record management as they offer more effective tools for the preservation and retrieval of archival records. The NAI has integrated new ICTs technologies to keep with the new phenomena of electronic

records management of archives. Because of meeting the requirement of quick retrieval of archival information records, it also started the computerization of its records holding in 1998. A customized software called "Archival Information Management System (AIMS) was developed using Visual Basic at the front-end and MS Access at the backend in 1998 for a standalone system but later in 1999. They upgraded the software in 2006 with the support of the National Informatics Centre (NIC) using ASP and MySQL RDBMS to operate and retrieve queries on NAI Interanet through a Web browser. The database has over 23 lakhs records for retrieval by staff, scholars, and users of Archives. Shortly, it will be hosted online on the NAI website to make it searchable globally. Also, the NAI has computerized its library catalog using LibSys Software but now planning to move to open source KOHA library management software shortly.

Microfilming

The NAI also started microfilming on a large scale to preserve its national heritage collection for the future and to protect its resources against the loss of fire, flood, war, and sabotage. For this, a Micrographic Laboratory was set up in Bhopal Regional office where negative copies of microfilms are produced and a master copy of the microfilming collection is kept in its Regional Office at Bhopal in a controlled environment while a backup copy is in the Main office in Delhi. There are more than 35,000 microfilms roles in the microfilm repository. Further, these microfilms were also converted into a digital format using new microfilm digitization technologies12.

Digitization

In the year 2000, the NAI started a pilot project to digitize its rare collection of manuscripts, Bhagwatgita, Ramayan, Mahabharata, etc. using a digital camera. Subsequently, several other digitization projects were also started to convert its files into digital format. To enhance its capacity in long-term digital preservation, an MOU was signed with "Centre for Development of Advanced Computer (C-DAC), Pune" to the fulfillment of the mandate of the Public Record Act of 1993. To make its digitized documents and reference media online to scholars worldwide, the NAI launched an online portal, Abhilekh Patal https://www.abhilekh-patal.in/

Department Of Delhi Archives

State and local archives are instrumental in the development and preservation of the history and their documentary heritage. The Department of the Delhi Archives was established in the year 1972 to preserve and make available the rich documentary heritage of all the events witnessed by the historical city of Delhi since 1785 to support and catering its information services to the academic community and general public. It in-houses archival records of various departments, PSUs, autonomous bodies under the Government of NCT including the records of Delhi Residents, and posts the Commissioner's Office records post 1857. These records are managed and preserved following the Public Records Act, 1994. Highlights of the collection, having a 600-page document on the trial of Bahadur Shah Zafar, documents related to the demolition of Chandni Chowk, and records on the setting up Imperial Delhi13.

In general, the Delhi Archive has lagged and less visible than the National Archives of India (NAI) possibly because of lack of funds and staff. However, the cataloguing of scattered material of Urdu and Persian languages have been started and digitization of its entire collection has already begun as the Delhi Government has provided funds.

Records Management

The Department of Delhi Archives has acquired and preserved the rich history of Delhi making. The primary function of the Department is to acquire and preserve the records related to Delhi City and to expand and

to enrich their holdings. Also facilitating the supply of the records to creating agency and the scholar to support their research and issuing certified copies of the Registration Records of properties, Gazettes, etc. to the general public. These facilities are offered only by implementing a systematic archives records management program to cater to a vast amount of information services to its diverse clientele.

The acquired records exist in any formats including paper, audios, videos, digital-born or maybe in email, etc. are selected based on its historical, administrative, legal, fiscal value and public interest, for permanent preservation under the supervision of the head of the Delhi Archives as per the collection development policy. The records received after appraisal, the shelf life of the records decided as the Record Retention Scheduled to ensure that any record should not be destroyed prematurely or retained for a longer period than required. Finally, the records are transferred for the permanent collection. The physical records that are preserved are made available at the premises following the access policy of the Department.

Digitization and Micro-filming

The Department has started converting its analog and print records into digital and microfilming formats aiming to preserve the original archival records for a longer period without harming them further while handling them manually; to make its collection accessible instantly to the larger community by using customized software for managing its digital resources. Under this project, over 4 cores of pages will be converted in digital microfilm formats. It is one of the archival preservation project of its kind in Asia costing Rs.25.40 crores. Subsequently, these digitized records will be available online through e-Abhilekh portal.

Digital Preservation Practices At National Archives Of India (Nai) And Delhi Archives (Da)

The prime objective of digital preservation of the archival records is to enable access digitally and preserving rare and fragile original records for the future. National Archives of India and Department of Delhi Archives are well-established archival organizations that have adopted best practices for the acquisition and preservation of records and follow a scientific approach using new technology for cleaning, repairing, treatment of records, etc. Additionally, adopted photocopying, reprinting, microfilming and digitization practices to conserve its analog or printed collection through a preventive, curative and restorative process to prevent the loss and to prolong life of the archival records.

Microfilming

The National Archives of India (NAI) established its first microfilm unit in 1948 to fill the gaps in its existing records. Presently under six series, it has created 12,656 microfilm rolls for preservation and access on nominal fee, details are provided at the NAI website. While Delhi Archives is equipped with the latest microfilming technology having SMA 51 Archive Writer and Kodak Prostars I microfilm processor. It has 1,553 microfilm rolls its printed catalogue in PDF format is available for download.

Digitization Process and Methods

These premium archives of India, NAI and DA represent a very unique and large collection of archival records but it is accepted that the entire collection of these archives cannot be digitized in-house with a lack of technological resources and manpower. As a result, it was necessary to outsource to scale up the digitization work to meet the target. For digitization of archival records, the material is selected based on the prescribed criteria of each archive including rare documents, those are frequently requested by the researcher for consultation, etc. and then set the priority for which collection will be hosted online

Outsourcing is a common practice for large-scale digitization projects to achieve the target in a stipulated period that institutions along with their routine responsibilities cannot often do, also it requires specialized

digitization equipment and expertise. NAI and DA have also hired external vendors for digitization work but the key archives' staff are engaged throughout the project to monitor the workflow, check quality control and adhere to the technical requirements prescribed by the institution. For instance, the complete inventory for the digitization of archival records is developed for the digitizing agency. After this exercise, digitization work through OCR and image conversion including image capture, processing, and cleaning, metadata creation etc. completed. Further, these digitized and physical print and analog records are rechecked with inventory prepared for digitization. Finally, these digital copies of records are hosted online on portal websites.

Digitization Formats

The digitization aims to increase the outreach of its collections through the web and minimize damage to the original material. It involves the conversion of print and analog material into a digital format which also becomes a challenge to manage a newly formed collection and its workflow. National Archives of India (NAI) follow well-established guidelines such as NARA and FADGI for scanning and digitization13. It applies multiple formats including Optical Character Recognition (OCRs) technologies to produce PDF/A format and image files stored JPEG format and also master digital images are stored in TIFF format. The text is usually captured in 300 ppi while images are captured in 300 ppi to 600 ppi depending on the type of material, file size in addition to the quality of the image and minimum required resolution for the display. The derivative files for user access are provided in JPEG and PDF/A format for images and text respectively. Post scanning, images are processed following the practice of Noise removal and de-skewed to achieve normal skew, sharpening, crop to outside the page edge, and split 2-page to produce 1 image per page unless information crosses both the pages14,15.

Delhi Archives (DA) devised its customized guidelines for the scanning process which follows standard digitization practices. The material files are in a mixed format of images and text. A text file is created through the OCRs process loading with ABBY Fine Reading XML software. A PDF/A file is created for each digital object while images are stored in JPEG format for access and the master file of both the format are stored in the PDF, XML, TIFF, and JPEG family format. The resolution for the master file is 300 ppi, 400 ppi, and 600 ppi but the resolution for derivative files is 300 dpi for users. All images are thoroughly viewed after scanning to check the satisfactory capture and rescanned if required and then post scanning processes such as noise removal, de-skewed, one image per page are performed.

Metadata

Metadata is an integral component of any digital management system and each digital object needs to be described. The quality of the retrieval output depends on metadata description and large-scale digitization projects throw many challenges while implementing a metadata component. NAI records hosted on Portal is primarily governed by the Dublin Core schema with descriptive metadata to describe digitize content online, structured metadata to demonstrate its authenticity and integrity, technical metadata - describes the hardware, software, and processes used for digitization, and administrative metadata which is customized to meet the requirement of the Archives and also keeping in view the needs of interoperability among the digital information systems. The file number is arranged by title which consists of allotted subject standards heads, the serial number of the file opened, the year of the opening of the file, and an abbreviated name/symbol identifying the section.

Figure 1: Part of a metadata record for a digital object from NAI – Abhilekh Patal (https://www.abhilekh-patal.in/jspui/)16

Representation from the Shir Ganga Hindu Sabha,Hardwar, against the Present Arrangements for the Construction of the Headworks of the Upper Ganges Canal, Hardwar

Type Of Document	Digitized Document
Department	Home Political
Branch	A
From Year / Date (YYYY-MM-DD)	1916-10
To Year / Date (YYYY-MM-DD)	1916-10
Date Uploaded (YYYY-MM-DD)	2020-11-24
Organization	National Archives of India
Identifier	PR_000003003883
File No.	HOME_POLITICAL_A_1916_OCT_204-05KWPART-2
Location	Repository II
File Size	0.7
Pages	3
Collection	Digitized Public Records Home Political

The metadata of Delhi Archives' E-Abhilekh is quite different and it is completely in-house customized to describe its collections which quite different from normal archives collection having property records dealing with lease deeds, rent agreements, sale certificates along with Delhi state official archival records. It also uses a different layout for displaying the digital objects that combine the values from several metadata fields (figure 2). Over 24 metadata fields are filled including title, subjects/ keywords, data, etc. to describe property details of an item. Each item hosts descriptive and administrative metadata which valuable for hosting institutions. Files are organized by title, date, and sequential pages and file name Series department, year, and sequential pages.

Figure 2: Part of a metadata record for a digital object from Delhi Archive – E-Abhilekh (https://www.abhilekh-patal.in/jspui/)17

	Name	Series_Department	Year	File_No	Period	Volume	PartNo	Department	Subject
	Delhi%20Gazette	Gazettes						Training & Technical Education	Recruitment Rules for the post of Head of Department Medical Laboratory, M/s, Un
	Manuscripts	Rare Collections	1922						Gadar-Dilli Ke Afsane

bahadur shah zafar   Advance ▾   Q

Showing 1 to 2 of 2 Records

Online Access through E-portals

NAI's vision to "To help in spreading a feeling of national pride in our documentary cultural heritage and ensuring its preservation for posterity" and has the mission to "encourage greater liberalization of access to archival holdings. With these objectives, is in the continuous process of digitizing its rare and valuable records to preserve them for future generations and to ensure easy access to documents. So far 3.70 crores of pages of archival records have been digitized and out of which 28,54,931 pages are hosted on e-Abhilekh Patal (www.abhilekh-patal.in) for public access. The portal has two major collections i.e. Online Reference Media having 27,61,537 records including public records, private papers, cartographic records and oriental records; the Digitized Collection has 71,764 digital records from 15 curated collections. The statistics show the portal has been accessed having hit over 27 lakhs having 13,537 registered users.

Delhi Archives has launched E-Abhilekh Portal (http://archives.delhi.gov.in/abhilekh/) with the objective provide online information services to scholars and organisations to meet the requirements of the Right to Information Act, and to preserve the original archival records intact for a longer period. The portal has two major collections i.e. Property Records collection with 24,54,902 and Archival Records having 1,48,128 records. A total of 4 crores images including maps and another archival document way back to the year 1803 which have been digitized using state-of-the-art technology and methods. Currently, it has 1,931 registered users and 14,384 hits.

Conclusion And Recommendations

The two major archives of the country National Archives of India (NAI) and Delhi Archives (DA) are very similar in the collection development of archival material and managing records as per the acts and standard practices. But the structure of both archives is different in their deliverable as they manage and preserve different types of collection and the user's demand is very different from each other. Delhi Archives in-houses land records of Delhi's Municipal and a large number of people approach the Archive to get a copy of records for settling down their legal issues. NAI is mostly approached by scholars and it involves a strict process of registration and verification for entry and the resources which are posing threat to national security may not be accessed.

While comparing the large scale of digital archival projects carried out by both the archives are very similar in digitization methods, project management, online access through the portal website but there are differences in the structure of metadata depending upon the types of collections the archives manage and the users' delivery and demand are very different as Delhi Archives' land and lease data for common people in addition to the archival research scholars. The output of the records can be enhanced by displaying a detailed, tabulated record at a time.

None of the archives uses any protocols for search and retrieval of records to re-use digital records at the client level. It is recommended to uses open search and retrieval protocols such as Z39.50, Search/Retrieve Web Service (SRW) and Search/Retrieve via URL (SRU), and Open Archives Initiatives – Metadata Harvesting Protocol (OAI-PMH) to make digital archival records interoperable.

The large-scale digital archives projects may be different for each archival institution depending upon their collection, available technologies, and funds, user's demand, etc. However, these projects need to have detailed documentation of digitization policy and digital collection development to make them sustainable and successful. It is anticipated that these two major archives of India will share their digitization strategies and in the development of best practices with the other archives which are in the process of digitization in order adopt uniform standards of the digitization process, metadata creation, and retrieval of records.

References

1. Dictionary.com https://www.dictionary.com/ (accessed on 20 November, 2020).

2. Schellenberg, T. R. The Appraisal of Modern Records. Bulletins of the National Archives. 1956, No.8, 1-46

3. Perry, S. R. Digitization and Digital Preservation: A Review of the Literature. School of Information Student Research Journal, 2014, 4(1), 1-12.

4. Directorate of State Archives (Government of West Bengal). http://wbsadte.gov.in/ (Accessed on 15 October, 2020)

5. National Archives of India. www.nationalarchives.nic.in (Accessed on 15 September, 2020)

6. The National Archives and Records Administration (NARA). Plan for Digitizing Archival Materials for Public Access: 2007-2016. 2014, https://www.archives.gov/files/comment/nara-digitizing-plan.pdf (Accessed on 1 November, 2020)

7. Hofman, H. The Archive. Archives, 2005, 131-158.

8. Keneley, M., Potter, B., West, B., Cobbin, P., & Chang, S. Digitizing Archival Records: Benefits and Challenges for a Large Professional Accounting Association. Archivaria, 2016, Vol.81, 75-100

9. Dietz, Steve. "Curating (on) the Web," in Museums and the Web: Museums and the Web 98: Selected Papers, eds. David Bearman and Jennifer Trant (Pittsburgh, PA: Archives and Museum Informatics) 6 Nov 1998 http://www.archimuse.com/mw98]

10. National Archives of India. www.nationalarchives.nic.in (Accessed on 15 September, 2020)

11. Gautam, M. History of Archives and Archival Sciences from 1950 Onwards Based on Indian Practices. Atlanti, 2013, 23(13), 135-147.

12. Department of Delhi Archives. http://www.archives.delhi.gov.in/ (Accessed on 17 September, 2020)

13. Northeast Documentation Conservation Center. Digital Preservation Assessment Handbook. 2019. 52p. https://www.nedcc.org/assets/media/documents/nedcc-DPA-hndbk-6.24-web.pdf (Accessed on 3 December, 2020)

14. Lapworth, E., Jones, S. and Georgieva, M. . Microfilm, Manuscripts, and Photographs: A Case Study Comparing Three Large-Scale Digitization Projects. Journal of Contemporary Archival Studies. 2019, 6(5), 1-21 (Accessed on 26 November, 2020)

15. Tolulope, B and Adjei, E. Challenges of digitization of the National Archives of Nigeria. Information Development. 2018, 1–12, doi: 10.1177/0266666918778099

16. National Archives of India - Abhilekh Patal. https://www.abhilekh-patal.in/jspui/ (Accessed on 18 September, 2020)

17. Department of Delhi Archives. Delhi Archive – E-Abhilekh. https://www.abhilekh-patal.in/jspui/ (Accessed on 18 September, 2020)**Websites**

18. Contributors

19. Mr Akhlak Ahmad is Ph.D scholar and working with the Centre for Women's Development Studies (CWDS) since 1993 and presently working Library Associate (Grade I). For this study, he is responsible for the conceptualization, methodology, investigation, data curation, writing, original draft preparation, writing review and editing.

20. Dr Swadesh Sharma presently working as Dy. Librarian cum Asst. prof. Head of Libraries, G. D. Goenka University, Gurugram Haryana. He contributed to this study by providing overall guidance in drafting and improving the contents of the paper.