

REVOLUTIONIZING ARCHIVAL RESEARCH AND DIGITAL PRESERVATION - EXPLORING THE POTENTIAL OF ARTIFICIAL INTELLIGENCE

* Dr. Seema Gholap

* University of Mumbai.

Abstract:

This study glances at the Maharashtra State Archives in Mumbai's existing archive management and preservation situation as well as the possible ways artificial intelligence (AI) may improve archival procedures. According to the report, the Maharashtra State Archives in Mumbai have a wealth of historical documents, but they also suffer difficulties with accessibility, preservation, and discovery. The study offers suggestions for putting AI-powered solutions into practice and implies that archival management and preservation might be enhanced by utilizing AI-powered tools and procedures.

Key Words - Maharashtra State Archives, preservation, AI-powered tools and procedures.

Copyright © 2025 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial Use Provided the Original Author and Source Are Credited.

Objectives

1. To examine the current state of archival management and preservation at the Maharashtra State Archives, Mumbai.
2. To investigate the potential applications of Artificial Intelligence (AI) in archival management and preservation at the Maharashtra State Archives, Mumbai.
3. To analyse the challenges and opportunities of implementing AI-powered tools and techniques in archival management and preservation.

Introduction:

Maharashtra state archives are located in Bombay (presently known as Mumbai) Pune and Kolhapur, in addition with two more regional offices lately opened in Aurangabad (presently known as Sambhaji Nagar) and Nagpur. They have rich and huge records of Maratha, Peshwa from whom the East India Company took over the charge. No other state of India has such records of an old era. As per the Handbook in Mumbai Archives that are about- “ 5 lacs bound volume and files this files are consist information a related to files dealing with political upheavals, military movements, education, judicial and ecclesiastical matters, rural economy, public health, development of ports, transport, trade and industry, water-supply, etc., and covering the period from 1630 to 1955. They are preserved for posterity because of their continual utility for the necessary governmental processes, for protection of public and private rights and above all for research by scholars and students. The historical sense, which is a public necessity in every country, can only be developed by the study of historical

evidence at first hand. The Maharashtra Archives facilitates researches in its records for proper purposes. The private records acquired from the old families, which throw light on many unknown events in the Maratha History, are outstanding among the new acquisitions of the Bombay Archives during last fifteen years.”ⁱⁱ the data shows that Maharashtra state archives is wealthy in case of information. it has wide range of information which can provide knowledge to every walk of life.

1. Overview of the Maharashtra State Archives, Mumbai: History, mandate, and current collections. -

The Maharashtra state Archives was established in the year 1821ⁱⁱ. At present the Maharashtra state Archives is located in the ‘Elphinstone college’ premises at South Mumbai. This Archives earlier was known as the ‘Government house’ⁱⁱⁱ. From 1665 it was established during the time when Britishers got Bombay from Portuguese^{iv}. In August 1821 to preserve the record of the secretary office and other offices in the presidency, Bombay record office was set up. This was basically for the administrator and the historians’ references. The Medan record keeper in charge Mr. William Wissenecraft, recorded or preserved the official records from 1630 to 1955^v. Digital copies of the catalogue to the Archives are easily available as into (the handbook of Bombay Archives) This hand book is very good for the beginner to get proper direction to dig records and reach the desired documents. however, the hand book provides just an outline to start research, where in one can find the in-depth data after finding and turning various files and great number of pages. It is an ocean of information found in these age-old files and records. Scholars or any person who wish to know more details about a certain topic, specific information can have a feast to his knowledge.

2. Challenges in Archival Management and Preservation:

Management-

A. Accessibility- There are difficulties at the entry level as it is difficult for a person to locate the Maharashtra state Archives since it has no separate building but it is situated in the campus of historical college name ‘Elphinstone college’ at South Mumbai.

B. Registration- There are certain forms and norms set up for every individual to fill in before accessing the records. The individual needs to fill in information asked in the form, after getting precise information and paying certain amount as an official fee and on fulfilling the required criteria, the person gets an access to the records or books. The individual then comes across certain files, registers which acts as index to get his desirable file’s volume, book number, post which he has to fill in the requisition slip. The record lifter brings in the file, not all at once but only 3 at a time as per rule set by the Archives official. After receiving the records brought by the record lifter, the individual has to sign on the requisition slip at the received column. After such lengthy procedure an individual then can go through the records.

C. Seating Arrangements- There are few tables and chairs set up for study and if the seats are not available due to overcrowd, the scholar needs to stand to get the desired information which causes great inconvenience.

D. Discovery- A record lifter takes long time due to the backdated record finding. Often due to human error or spaces or time constraint the records are not placed in proper shelf or in orderly manner. Sometime the files are dusty and the ink used to write the heading of the file gets faded and unreadable. Few records like

single papers, pamphlets are compiled in one file to keep it safe but it's very difficult to allocate or to find. The records file name is not mention in such situation. It becomes difficult for the record lifter to find the record within time.

In such cases the record attender informs the individual to wait for two -three days. This causes great inconvenience to individual/scholar.

Record officer- The record officer guides you through your journey but due to unavailability of staff he has to limit in keeping the records information and at times he remains unavailable.

Preservation –

A. Shortage of place- The place which is allotted to the Maharashtra state Archives is located in Elphinstone college west wing. This place is now over occupied by various records and needs additional place to preserve newly added records.

3. Digitization challenges.

A. Handbook -F. Kindersley i 1921 has first established handbook of the Bombay government. Unfortunately, it has been out of print since quit a long time. The Record till 1820(few exceptions) are found in this hand book but of late there had been humongous increase in the volume of records transfer to the department of Archives and so the revise handbook is a need of an hour. The unavailability of such handbook brings researcher and official untold and uneasiness and inconvenience which results into wastage of time.^{vi}

B. Inconvenience to the foreign scholars -There are good number of foreign Scholars who wish to study about India and about its various aspects. Maharashtra state Archives, Mumbai has a great number of official records and it is considered as the primary source. There are many foreign Scholars who visit Mumbai Archives and face difficulties right from the entry level like locating the archive`s office which is established in the Elphinstone college premises in South Mumbai. After that, the registration process official formalities of finding an allocating the records and un revised handbook leads to a lot of inconvenience not only to foreign students and research scholars but also to Indian scholars especially those belonging to interior or remote place of India.

4. Potential Applications of AI in Archival Management and Preservation:

A. Transkribus- This is a comprehensive platform and AI Product which is widely used for recognising the script. It is programmed to read the faded ink in very clear manner. It also preserves the data and we can make necessary changes in the text if desired. It also gives option to choose as page, book to have that data. It basically scans the document, make pdf, word document for preserving purpose. This is a great help in preserving document which are in volumes.^{vii}

B. Google Handwriting Recognition: -This platform is widely used. This tool is basically used to recognise the text written in images, maps and the text which is written in unknown language of the user. The user can upload the image and get it translated it in desired language.^{viii}

C. **Microsoft Handwriting Recognition:** Microsoft Handwriting Recognition is a software application that is designed to identify handwritten text in documents and images.^{ix}

D. Optical Character Recognition (OCR) Tools

A. **Tesseract OCR:** This is a Google initiative and it is an open-source OCR engine.

B. **Adobe Acrobat OCR:** A device that is skilled in detecting text in scanned documents and images.

C. **Readiris:** A document management, record digital transformation, and OCR tool.^x

D. **Machine Learning Tools-** There are certain Machine learning tools which assist in data mining, classification, regression, and more. e.g. scikit-learn, Weka Microsoft Azure Machine Learning

Preservation:

Digital Preservation Tool (DPT): Software designed to safeguard digital content like documents, images, and multimedia files.

Preservation Manager: A system that organizes digital preservation processes, encompassing intake, manipulation, and safekeeping.

Archivematica: A comprehensive platform for digital preservation, offering integrated functionality for receiving, processing, and storing digital materials.

1 Focus on the tools' capabilities-

To analyse and understand human language, developers employ tools such as Stanford Core NLP (Natural Language Processing), spaCy, and NLTK. These libraries enable applications to perform tasks like identifying parts of speech, recognizing entities, gauging sentiment, and understanding the meaning of text.

5 Breaking it into list form, with slight modification-

Here are some common Natural Language Processing tools:

- **Stanford CoreNLP:** A Java library useful for tasks like part-of-speech tagging, named entity recognition, and sentiment analysis.
- **spaCy:** A Python library that is known for its speed and efficiency in text data processing.
- **NLTK:** A Python library with a wide range of NLP capabilities, including text processing, tokenization, and semantic reasoning.^{xi}

5. AI Adoption in Archival Institutions: At present Mumbai state archives is in primary stage in adopting AI. They provide various photocopies on demand after paying their set fees. It also provides information in pdf format after getting the formalities done. It takes care of certain security measures by scrutinising the purpose of data seeking. For the old record AI is still not adopted and there are few employees who are copying the data from old files to secure them for future purpose.

6. Challenges and Opportunities of Implementing AI-Powered Tools and Techniques: The Maharashtra state archives is government organisation needs certain approvals regarding financial aids to purchase the software's. The officials were not certain to provide information due to their own limitation. The human resources challenges which are faced in the Maharashtra state Archives cannot go unnoticeable. They have

taken measure for the staff recruitment which is inadequate in number. For collaboration and knowledge-sharing.

1. Maharashtra state archives need to install updated technical support -as in undated PC laptop along with the new trained human resource to utilise it judiciously.
2. If the organisation overcome the challenges, it can become an online hub of information and knowledge.
3. People, researchers can get the information at their home, without travelling to south Mumbai,
4. This also can reduce handling of original copy of data. For the authenticity of the records and information the researcher can get the document stamped in one visit.
4. This can save time, money and pollution which causes by vehicle used for commuting to and fro.

Conclusion:

The study comes to the conclusion that, while having a wealth of historical data, the Maharashtra State Archives in Mumbai suffers difficulties with preservation, accessibility, and discovery. The study offers suggestions for putting AI-powered solutions into practice and implies that archival management and preservation might be enhanced by utilizing AI-powered tools and procedures. The report also emphasizes how crucial it is for archival institutions to work together and exchange knowledge, as well as how crucial it is to create best practices and standards for the use of AI in archival administration and preservation.

Cite This Article:

Dr. Gholap S. (2025). *Revolutionizing Archival Research and Digital Preservation -Exploring the Potential of Artificial Intelligence.* In **Educreator Research Journal:** Vol. XII (Issue II), pp. 93–97.

ⁱ Maharashtra State Archives, Mumbai Handbook. 4

ⁱⁱ Ibid pg.2

ⁱⁱⁱ Ibid pg.2

^{iv} Ibid pg.2

^v Ibid pg.2

^{vi} Ibid pg.3

^{vii} . Mühlberger, G., & Sieglerschmidt, J. (2019). Transkribus: A Platform for Handwritten Text Recognition and Transcription. *Journal of Cultural Heritage*, 35, 53-63.

^{viii} "Deep Learning for Handwritten Text Recognition: A Review" by Journal of Intelligent Information Systems (2020) - DOI: 10.1007/s10844-020-00634-6

^{ix} "Microsoft Handwriting Recognition: A Review" by International Journal of Advanced Research in Computer Science (2020) - Published in: Volume 11, Issue 2 - DOI: 10.26483/ijarcs.2020.020214

^x Readiris vs. Tesseract Comparison on 5.3.25 at IST.13.59

^{xi} Readiris vs. Tesseract Comparison on 5.3.25 at IST.13.59