



State-of-the-Art: A Study of Big Data in African Historical Research

Opeyemi Anthony Iroju ^{a*}

^a Department of History, Obafemi Awolowo University, Ile-Ife, Nigeria.

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: <https://doi.org/10.9734/air/2025/v26i41384>

Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://pr.sdiarticle5.com/review-history/138163>

Original Research Article

Received: 15/04/2025

Accepted: 17/06/2025

Published: 23/06/2025

ABSTRACT

African historical research usually relies on traditional sources of data such as oral traditions, historical records, ethnographic and colonial archival documents. These sources of data are often limited in quantity and scope, usually fragmented and may not include contemporary events in real time. They may also lack a consistent format and may be geographically restricted. Nevertheless, in recent times historians in African studies are gradually deploying big data tools such as the Slave Voyages Project and Endangered Archives Programme (EAP) to digitise, preserve and provide easy access to African history. This approach is aimed at raising marginalised voices in African historiography.

Aims: This study provides an in-depth study of big data projects in African historical research. The study also investigates the potentials and problems of big data in African historical research.

Methodology: The study relies primarily on secondary data which include scholarly books, journal articles and conference proceedings obtained from diverse electronic academic databases.

Results: The result of the study revealed that while big data enhance African historical scholarship by creating a more understanding of the past, lack of skills in big data analytics, language barrier as well as loss of cultural, historical and social context of data due to identification of trends across large datasets are some of the challenges of big data in African historical research.

*Corresponding author: E-mail: irojuopeyemi@gmail.com, irojuopeyemi@oauife.edu.ng;

Conclusion: This study concludes that while big data offers efficient approach to researches in African history, its attendant challenges emphasises the need to balance this digital methodology with scholarship in traditional historical research.

Keywords: African historical research; African historiography; big data; historical records.

1. INTRODUCTION

The advent of big data in the last decade has significantly transformed the way African historical scholars interpret, analyse and write about the past by offering novel approaches to investigate the complex past of the continent whose historical narratives are shaped by fragmented records, scarcity of written sources, biased Eurocentric perspectives and over reliance on oral traditions. Hence, African history is often subjected to misrepresentation in global discussion (Afolabi, 2012). Nevertheless, the integration of big data in African historical research marks a significant shift in addressing the misrepresentation of African history in global discourse by confronting established historical narratives through new evidences and perspectives from massive volumes of historical data such as digital archives and databases, online libraries and social media (Muleka, 2024). This affords African historical scholars to explore African history in a comprehensive manner, empowers them to preserve the history of African communities as well as reconstruct African history with greater accuracy and a high depth of historical investigations. Hence, the introduction of big data in African historical research is gradually prompting historical scholars to reconsider traditional methodologies by embracing this data-driven approach to research.

Succinctly, data is generally defined as a collection of observations, figures and facts about entities and events. Lately, massive amount of data is generated rapidly from multiple sources at a very high rate. Hence, the emergence of the term big data. Big data according to Iroju and Ojerinde (2016), refers to a collection of extremely large and complex sets of data which are difficult to manage using common database management tools or traditional data processing applications but continues to increase intensely over time. From the humanities point of view, Eijnatten et al. (2013) defines big data as huge quantities of digitised information that can be analysed using data-intensive methods as opposed to conventional humanist methods which are geared towards the interpretation of limited

number of texts, images and datasets. Big data are usually measured in terabytes, petabytes, exabytes and yottabytes. They can be structured or organised in a fixed format, unstructured without a predefined format and semi-structured. They are usually characterised by seven V's which describe their key features (Saeed & Husamaldin, 2021). These include volume, velocity, variety, veracity, value, variability, and volatility. These 7V's are significant to African historical research because they make it easy to interpret and analyse the massive amount of fragmented data generated from diverse sources such as oral traditions, colonial documents and artifacts.

Big data offers a wide range of opportunities for African historical researchers. For instance, Muleka (2024) reveals that big data aid historians to discover hidden patterns, trends and correlations across time and space that are often ignored by traditional research methods. A typical example of a big data project in African historical research that reveals hidden patterns is the Slave Voyages Database Project also referred to the Trans-Atlantic Slave Trade Database. This project which analyses slave trade voyages from 1514-1866 reveals that about 12.5million people were put on the Ship in Africa during the Trans-Atlantic Slave Trade while only 10.7million arrived in America (Daryle, 2018). Furthermore, the Slave Voyages Database Project adopts visualisation techniques like interactive map to reveal where slaves were taken from and the destination that they were transported to and how the slaves resisted their captivity (Misevich et al., 2017). In addition, since African history spans thousands of years and involves large amount of data from diverse traditional sources, big data help to integrate, digitise and store these data, thereby increasing their accessibility for global access. For instance, the Timbuktu Manuscript, is a big data project consisting of vast collection of Arabic Text that provides information about the history of Timbuktu, a country in West Africa (Stewart, 2021). These no doubt help African historical researchers to reconstruct their history with evidenced based narratives which ultimately promote Africa's cultural heritage. Nevertheless,

difficulty in digitising sources of historical data such as oral traditions and artifacts, dominance of colonial documents, bias in data collection, inadequate involvement of local communities as well as ethical concerns are some of the hinderances facing big data in African Historical Research. Consequently, big data in African historical research can be associated with foreign narratives which marginalises African indigenous experiences, incomplete and biased historical narratives as well as the denial of Africans control of their history which can ultimately distort African historical narratives.

It is therefore against this background that this study explores some projects in African historical research that employs big data, the potentials of big data technologies in collecting, storing, analysing and preserving African historical records as well as the challenges of adopting big data methodologies in African historical research. These are with a view to supporting innovating methodologies that can enhance African historical studies.

2. RESEARCH OBJECTIVES

The aim of this research is to examine the impact of big data in African historical research. The specific objectives of this research are to:

- i) investigate big data projects in African historical research;
- ii) examine the significance of big data in African historical research; and
- iii) explore the limitations of big data in African historical research.

3. RESEARCH QUESTIONS

The following research questions were formulated in line with the research objectives:

- i. What are the big data projects employed in African historical research?
- ii. What are the significance of big data in African historical research? and
- iii. What are the limitations of big data in African historical research.

4. RESEARCH JUSTIFICATION

The history of African countries has been influenced by colonial-based records since the 15th century when the European began to document the history of various countries on the continent. This no doubt affected the way the

histories of these countries are presented because the Europeans wrote based their own perspectives. By the mid-20th century, researchers deployed oral traditions and ethnography to reconstruct the history of African countries (Cieslik & Margocsy, 2022). However, data obtained through African historiography were often subjected to distortion, mythology, alteration and loss of data over time (Muleka, 2024). Furthermore, written African historical records including those found in various correspondences, court documents, newspapers and letters are often scarce and fragmented, thereby making the task of reconstructing African history a herculean task (Afolabi, 2012). Hence, the need to employ big data as a means of African historiography to challenge western dominance in African history, digitise and preserve African documents for the future generation and provide a comprehensive understanding of African history. Furthermore, there is limited research on this subject matter, hence, the need for this study.

5. METHODOLOGY

The objectives of this study were achieved through an extensive appraisal of relevant and related literatures. Studies that were considered significant to how big data empower historians, limitations of big data in historical practices and big data repositories that facilitate the analysis of huge quantities of data from several digital sources in innovative ways were considered for this study. Electronic databases such as African Journals Online (AJOL), Sabinet, Science Direct, JSTOR, Scopus and Google scholar were used as secondary sources of data for obtaining related documents on the subject matter. In addition, information on websites of big data projects for African studies also served as sources of information for this research.

6. THEORETICAL FRAMEWORK

This study is centered on two theories which include datafication and digital humanities theories. The theory of datafication was propounded by Mayer-Schonberger and Cukier (Mayer-Schonberger & Cukier, 2013). This theory is driven by the advancement in Information and Communication Technology (ICT), Artificial Intelligence (AI) and the recent rise in the concept of big data. The theory of datafication focuses on the transformation of human activities, practices and human behaviour

into quantified or digital data for the purpose of making decisions, real-time tracking, social use and predictive analysis. The theory suggests that the process of quantifying objects and activities is not neutral but shaped by power structures, politics, historical legacies, human choices and social contexts (Reutter, 2022). This evidently implies that the digitisation of data can be biased, incomplete and inaccurate. The theory also emphasises that data is not equivalent to information but must be processed, analysed and interpreted before it becomes information that can be used for decision making. The theory also states that the quantification and digitisation of data can absolutely lead to loss of crucial perspectives such as emotions and cultural nuances which in turns leads to the misunderstanding and misrepresentation of data (Cieslik & Margocsy, 2022). Datafication however contributes to digital amnesia, a term that describes the inability of individuals to easily recall events due to the overreliance on digital devices. This theory is related to this research because data in the African continent have been captured in massive heterogenous databases globally and these databases have been used to study African history. Hence, this theory helps to understand how big data have been used to reshape researches in African history. The theory also helps to reveal that big data reechoes narratives of the Europeans because the collection of big data is not neutral, as it often depends on colonial archives at the expense of oral traditions, local perspectives and the indigenous knowledge of individuals in African communities.

The theory of digital humanities cannot be attributed to a singular individual, however, key figures that are instrumental to the development of this theory include Father Roberto Busa, often referred to as the founding father of digital humanity, John Unsworth and Stephen Ramsay (Steven, 2016). The digital humanity theory focuses on how digital technology can be utilised for studying, creating, analysing, interpreting and representing culture, humanistic knowledge, thoughts, human behaviour and communication. Hence, digital humanity theory explores how digital tools and techniques can be used to promote researches in humanities such as History, Literature and Philosophy through the digitisation of historical documents, literary texts, cultural artifacts and practices. The theory also emphasises that digital artifacts are shaped by hardware and software. This theory is relevant

to this research because it offers innovative ways of exploring researches in humanities in this digital age.

7. CONCEPTUAL REVIEW

This research is based on the following concepts briefly discussed below: big data and African historical research.

7.1 Overview of Big Data

Generally, big data is defined as a collection of large datasets that cannot be easily collected, processed, stored, managed and analysed with traditional data management technologies. Big data is usually obtained from heterogenous sources which include digital archives, social media, sensors and Internet of Things (IoT), online transactions, mobile devices as well as government databases. Big data are largely characterised by diverse authors by the 3V's which are referred to as Velocity, Volume and Variety as depicted in Fig. 1. Volume refers to the massive amount or quantity of data obtained from different sources such as social media, sensors and satellites. Big data is usually measured in terabytes, yottabytes and zerabytes. Velocity on the other hand is the speed at which data is generated, processed and analysed while variety refers to the different sources from which the data are generated. Big data are usually generated from different sources which are characterised by heterogenous formats which are briefly described below:

7.1.1 Structured data format

A structured data can succinctly be defined as a form of data that is organised in a pre-defined format, often in rows and columns (Iroju & Ojerinde, 2016). Typical examples include data stored in spreadsheets, relational databases and Common Separated Values (CSV).

7.1.2 Unstructured data format

According to Iroju and Ojerinde (2016), unstructured data format refers to data that do not follow any specific or pre-defined format. Common examples of unstructured data include text documents, multimedia files such as audio, video and images, social media contents, electronic mails and web contents. Interestingly, 80% of the data in the world is unstructured (Chief Data Officer Magazine, 2024).

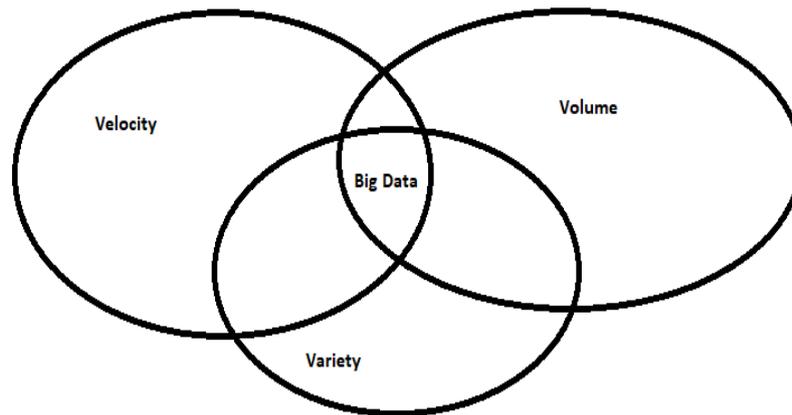


Fig. 1. The 3V's of Big Data (Bhalla, 2019)

7.1.3 Semi-structured data format

These are forms of data that do not adhere strictly to a specific or pre-defined format but possess organisational properties like tags, keys or markers that facilitates data processing (Iroju & Ojerinde, 2016). Typical examples include electronic mails with meta data, No Structured Query Language (NoSQL) databases and web pages with tags.

Other key characteristics of big data include veracity, value and variability. Veracity refers to the accuracy and how truthful the data is while value denotes the usefulness, relevance and benefits that can be derived from data. Variability on the other hand refers to the inconsistency and irregularities that are associated with big data. This implies that the meaning and quality of big data can change across diverse sources over time.

7.2 African Historical Research

African historical research is a field of study that evaluates the past of Africa as a continent and the different countries across the continent with a view to understanding the continent's socio-economic, cultural and political developments over time. African historical research usually adopts an interdisciplinary model which encompasses oral traditions, written documents, archeological and archival materials to analyse the socio-economic and political developments of the continent. The ultimate goals of African historical research involve the following:

7.2.1 Correction of eurocentric perspectives

African history has long suffered from discrimination from the Europeans who believed

that there was no such thing as Africa before colonisation and that the history of Africa began with the history of Europeans activities in Africa (Ade Ajayi & Alagoa, 1974). Hence, this European's view of Africa created an erroneous gap in the periodisation of African history. Furthermore, the lack of written documents and holistic archeological evidence left African scholarship in the hands of foreign adventurers, Arab traders, missionaries as well as sailors and amateur historians who had little regard for African oral traditions and cultural narratives. Hence, their accounts of the history of African societies were not an adequate representation of the societies (Ade Ajayi & Alagoa, 1974). Nonetheless, one of the goal of African historical research is to reclaim and reinterpret African history from the African perspective with the aim of correcting Eurocentric biases.

7.2.2 Preservation of african indigenous knowledge

African historical research promotes the preservation of African indigenous knowledge about governance, values, festivals, migrations and settlement patterns through oral history, folktales and folklores (Ade Ajayi & Alagoa, 1974). Indigenous knowledge of the African continent is also preserved through researches on artifacts like sculptures, textiles and architecture. This is significant for the transmission of valuable cultural heritage from one generation to another.

7.2.3 Fostering African identity and unity

African historical research provides a foundation for understanding the indigenous political institutions of diverse societies in the continent,

economic policies, technology, trade networks such as the Trans Sahara Slave trade and the Agricultural systems practiced in the continent (Daryle, 2018). This no doubt fosters a sense of unity and continental identity in Africa.

7.2.4 Fostering educational development

African historical research helps to strengthen educational development in the continent by ensuring that African students learn about their own history, leaders, culture and heritage. This gives scholars the ability to think critically and analyse different historical viewpoints.

8. LITERATURE REVIEW

There is paucity of literature on the field of big data in African historical research largely due to the novelty of digital humanities in Africa, sparse incentives and training of African scholars in data analysis as well as inadequate technological infrastructure. Nevertheless, this section briefly reviews some literatures that are considered relevant and related to the subject matter.

Muleka (2024) explores the potentials and limitations of big data in the future of historical writing. The author relied on existing secondary sources, majorly related academic literatures on big data's use, impact, limitations and potential biases to achieve the research goal. The study revealed that big data can reveal hidden patterns, challenge established narratives, and create a more comprehensive understanding of the past in historical researches. However, the study showed that reduction in data quality, data incompleteness as well as lack of data accessibility are some of the challenges of big data in historical researches. This study was not focused on African historical research, neither did it reveal the big data projects deployed in African historical studies.

Gutmann et al. (2018) examines the use of big data in economic history. The authors emphasised that economic historians have been able to create big datasets by pooling datasets from diverse censuses, genealogies and population centers. This process was facilitated by the North Atlantic Population Project and IPUMS-International project which harmonized the records of 614million individuals from 820 countries since 1960. This data has been used to conduct several researches relating to demographic processes such as fertility, mortality

and nuptiality in different countries specifically China, Italy and Japan.

Fourie (2016) examined the extent to which data revolution is consequential to the field of African economic history. The author argued that data processing software have enabled historians to capture historical statistics on a larger scale and also rewrite African history through new methodological tools. The author also emphasised that the digitisation of colonial and post-colonial records will advance the knowledge about Africa's economic past. It should be noted that Economic history is a branch of history which uses methodological tools in the field of Economics, history as well as statistics to analyse economic phenomenum of human activities and the society at large. Big data is now revolutionising Economic history by allowing historians to digitise large collection of documents such as trade ledgers, tax records, census data as well as financial transactions. These tool assist historians to now have access to large amount of data relating to economic history from diverse sources.

Eijnatten et al. (2013) discusses the promises and challenges of big data in historical inquiry. The authors argued that with big data, historians have the opportunity to analyse massive amount of data as well as integrating innovative methods such as data mining, machine learning, natural language processing and artificial intelligence into historical research. Furthermore, Eijnatten et al. (2013) were of the opinion that the use of big data in historical inquiries offers visual representation of data such as graphs, plots and charts which allows historians to discover hidden patterns in data.

9. RESULTS AND DISCUSSION

This section provides answers to the research questions which are presented below:

9.1 Research Question 1: What are the Big Data Projects employed in African Historical Research?

There are quite a number of big data projects in African historical researches. These projects include the Slaves Voyages Database, the African Origins Project, the Endangered Archive Programme (EAP), the Aluka project and the Timbuktu Manuscript Digitisation Project

amongst others. These projects are described below:

9.1.1 Slaves voyage database

The slaves voyage database is the largest global online repository of information on Trans-Atlantic and intra-America slave trades (Daryle, 2018). The database was developed to digitise, document, preserve, visualise and analyse the Trans-Atlantic slave trade which transported over twelve million African slaves on 36,000 slave trade voyages to America against their will between the sixteenth and nineteenth centuries using historical data (Elits, 2021). It provides information on the vessels, routes, slaves and their captors. The database draws its data from various counties across the Atlantic from libraries, colonial and government archives, shipping and port records and scholarly researches from different research institutions. The slave voyage database is therefore a significant source of information for historians whose research interest is on the Trans-Atlantic slave trade and the history of slavery in general. The key statistics of the database are depicted in Table 1.

The database uses a relational database, specifically MySQL built on Oracle Cloud Infrastructure to store the data of over 36,000 Trans-Atlantic voyages, it also possesses data visualisation tools such as time lines, interactive maps and charts which enable users to visualise the activities of the slave trade. The database also uses OpenRefine tool to clean and structure messy data in order to improve the accuracy and data analysis of the system (Daryle, 2018).

9.1.2 Endangered archive programme

The Endangered Archive Programme (EAP) is a digital archive housed by the British Library and funded by Arcadia fund (British Library, 2025).

The project focuses on the digitisation and preservation of records that are vulnerable, inaccessible, prone to destruction, neglect, or physical deterioration especially in regions with limited resources. Although, the EAP is a global initiative, it has been deployed in several African countries including Nigeria, Ghana, Kenya and Malawi (British Library, 2025). The database contains over fifteen million images, thirty-five thousand audio tracks and digital manuscripts such as the Islamic Manuscripts in Mali and Palm-leaf Manuscript in Sri Lanka, printed books and archival documents (British Library, 2025). Computing infrastructure such as digital camera and scanners are required for the capturing of high resolution images. The EAP draws its data from diverse sources ranging from hand written texts on religious, philosophical and literary works found in palaces, monasteries and mosques, colonial documents, recorded or transcribed oral traditions and cartographic records (British Library, 2025).

9.1.3 African origin project

The African Origins Project is a web-based database which is an extension of the slaves' voyage database (Elaine, 2011). This database was designed to provide information of the identities of Africans who were liberated in the nineteenth century during the Trans-Atlantic slave trade. The database contains the list of about one hundred thousand African names, places of embarkation and disembarkation, arrival date, captive fate and vessel fate amongst others. This information was obtained from historical documents such as those found in courts of mixed commission records in Havanna and Freetown during the period of the slave trade (Elaine, 2011). The information in the database assists descendants of enslaved Africans in diaspora to trace their ancestral, geographic and linguistic origin as well as their cultural background. Hence, this database supports

Table 1. Key statistics of the slave voyage database (Tran & Silva, 2002)

S/N	Database Feature	Total
1	Number of voyages	36,108
2	Number of attributes	274
3	Number of ships	9440
4	Average number of slaves per voyage	329.9
5	Average voyage duration	435.12
6	Average crew size	30.12
7	Period	1526-1844
8	Mortality rate	12%

The image shows a webpage interface for the 'People of Atlantic Slave Trade - Database of the Enslaved'. At the top, there is a red header with a hamburger menu icon on the right. Below the header, the title 'People of Atlantic Slave Trade - Database of the Enslaved' is displayed. Underneath the title, there are two buttons: 'Back to Search' and 'Contribute to origins of the name of this person'. Below this, a section titled 'Select to contribute:' contains two buttons: 'Names & Pronunciation' (highlighted in red) and 'Language Groups'. The main content area is titled 'Suggestion 1' and contains the instruction 'Suggest a modern version of the name.'. It includes a 'Spelling *' label above a text input field, a 'Comments' label above a larger text area, and a 'Recording' label above a microphone icon.

Fig. 2. A webpage demonstrating crowdsourcing process in African Origin Project (African Origin Database, n.d.)

African historical research especially those related to the reconstruction of lost African identities and cultural heritage. The database employs Natural Language Processing tools such as morphological analysers, part-of-speech taggers and lemmatisers to process textual information of African names. The database is also designed with a standard relational database as it presents its information in columns and rows. The database employs crowdsourcing to enable users, especially Africans with expertise in African languages, ethnic groups and cultural practices to contribute knowledge about the ethnic and linguistic origins of the names listed in the records. This process is depicted in Fig. 2.

9.1.4 Aluka project

The Aluka project is a collaborative project between the University of Cape town and Aluka, a non-profit digital organisation supported by Andrew W. Mellon and Ithaka and recently integrated into JSTOR (Rajan & Heinz, 2007). The Aluka project can be described as a digital repository of scholarly resources relating to African historical research with the intent of

preserving African cultural heritage sites and landscapes, advancing human and scientific knowledge as well as revitalising African trade and craft. This digital library focuses on the collection of photographs documenting African heritage sites such as the Djingereyber mosque in Timbuktu, the rock hewn churches at Lalibela, the dry-stone ruins at Great Zimbabwe, the great mosque of Djenne and the Swahili ruins of Kilwa Kisiwan (Rajan & Heinz, 2007). It also consists of a vast collection of African plants species as well as documents and images documenting liberation struggles in South Africa, Mozambique, Zimbabwe and Namibia (Isaacman et al., 2005). This database also contains scientific and research papers, excavation reports, travelogues and monographs of various communities in Africa. This database therefore serves as a valuable tool for researchers whose interest lies in African history.

9.1.5 Timbuktu manuscript digitisation project

The Timbuktu manuscript digitisation project is a collaborative effort between the University of Cape Town, Sauvegarde et Valorisation des



Fig. 3. An example of a Timbuktu manuscript (Calvino, 2025)

Manuscripts pour la Defense da la culture Islamic (SAVAMA-DCI), Hill Museum and Manuscript Project (HMML), UNESCO and other international organisations. This project was inspired by the destruction and theft of Timbuktu manuscripts by Islamic rebels in 2013 (Rainier, 2003). The project is therefore aimed at preserving the vast collection of Timbuktu historic manuscripts in Mali through digitisation of the manuscripts in order to make them easily accessible for research, prevent the manuscripts from extreme climate, theft as well as rebel insurgencies. These collections include manuscripts in Arts, Philosophy, Science, Judicial Law, Government, Music as well as copies of the Holy Quran (Rainier, 2003). These manuscripts according to Rainier (Rainier, 2003) were mostly written in Arabic and other African languages in Arabic scripts such as Fulani, Swahili, Soninke and Songhay. The Timbuktu manuscript digitisation project consists of over four hundred thousand digitised Timbuktu manuscripts. Fig. 3 shows an example of a Timbuktu manuscript.

9.2 Research Question 2: What are the Significance of Big Data in African Historical Research?

Big data offers several benefits to African historical research. These benefits include:

9.2.1 Preservation of African historical records

Big data facilitates the digitisation of African historical records such as oral traditions and written documents which has greatly helped in

the preservation of African historical records (Muleka, 2024). Big data projects such as the Timbuktu Manuscript Digitisation Project and Endangered Archive Programme are typical examples of projects that use big data technology to prevent African cultural heritage from harsh climatic condition, loss, theft and destruction. This provides a long-term survival of vital historical records in Africa. In addition, digitised historical data are usually stored in cloud-based infrastructure which facilitates the storage of African historical records that are at the risk of being lost, stolen or forgotten thereby providing a holistic view of African historical events.

9.2.2 Integration of different sources of data

With the deployment of big data technology in African historical research, the challenge of limited sources of data that has plagued African historical research for centuries has been largely overcome (Muleka, 2024). For instance, big data projects in African historical research provides integrated data from diverse sources in form of audio and video recordings, remote sensing data, satellite images and GIS data, maps and cartographic records, time stamps and geotags. This affords researchers to conduct their researches from large volumes of data from multiple sources. This no doubt ensures accuracy and reliability of historical interpretations and also enhances the depth of African historical research. Furthermore, the integration of data from diverse sources in big data platforms improves accessibility to historical records thereby fostering a more inclusive understanding of African history.

9.2.3 Reconstruction of African history

The use of big data technologies helps Africans to contribute to the existing platforms of big data projects in African historical researches. This no doubt serves as a veritable tool for the reconstruction of African history (Afolabi, 2012). For instance, African Origin Project provides a platform for African researchers, authors and communities to decolonise African history by granting them the opportunity to contribute to African historical perspectives. With this, historians can challenge African narratives dominated by Eurocentric perspectives which have usually resulted in biased interpretations, hegemony and marginalisation of Africans from their history. Hence, it can be deduced that big data projects play a significant role in cultural history.

9.2.4 Promotion of African historical scholarship

Big data tools such as the Slaves Voyages Database and Timbuktu Manuscript Digitisation Project promote African historical scholarship and research by providing information on Afrocentric perspectives in global historical discourse. Furthermore, the integration of historical records from heterogeneous sources such as those obtained from colonial administrators and port authorities assist researchers to consult a wide range of historical data during their research activities (Tran & Silva, 2002). This facilitates a more inclusive as well as a greater accuracy and a high depth of historical investigations

9.2.5 Provision of data visualisation tools

According to Muleka (2024), big data platforms often incorporate data visualisation tools which illustrate trends and relationships in historical data and also assist historians to disseminate the outcome of their research to a wider audience. For instance, the Slave Voyages Database Project adopts visualisation techniques like interactive map to reveal where slaves were taken from and the destination that they were transported to during the period of the Trans-Atlantic slave trade.

9.3 Research Question 3: What are the Limitations of Big Data in African Historical Research?

There are quite a number of factors that are considered as challenges to this new innovative

methodology of conducting African historical research. These factors are discussed below:

9.3.1 Promotion of western narratives in African history

Big data platforms deployed in African historical research are usually projects of foreign institutions and organisations and they usually deploy data from large sources of European documents on African history which include port authorities' documents, missionary accounts and colonial archives. These sources of data may not adequately represent the historical realities of Africans.

9.3.2 Language barrier

Africa as a continent boasts of over three thousand languages which constitute about one-third of the world's languages (Statista, 2024). Most of these languages are not standardised, this makes it difficult to integrate historical data from different sources and diverse languages into big data platforms. Similarly, the translation of historical records written in non-standardised African languages can be difficult to translate into standardised global languages like English upon which big data systems are built. This can lead to the loss of cultural, historical and social context of data, which invariably results in the misinterpretation of African history. In addition, Natural Language Processing tools used in big data platforms are designed in English Language and not for African languages, this makes the automation, analysis and extraction of data from documents written in African languages difficult.

9.3.3 Ethical concerns

The use of big data in African historical research raises a lot of ethical concerns such as privacy concerns, informed consent, data control and data ownership (Muleka, 2024). For instance, most African historical records stored on big data platforms are domiciled in foreign institutions and organisations, which in turn limits the access of Africans to these systems. Hence, the question of who owns the data remains unanswered. Furthermore, the digitisation of the history of African societies without the explicit consent and inclusivity of individuals of the communities also raises ethical issues in the use of big data platforms as a source of data for African historical research. This form of data colonialism results in the continued marginalisation of African narratives from their history.

9.3.4 Lack of adequate big data analytic skills on African historians

It is a known fact that most African historians lack the basic skills in big data analytics and data science which is a major barrier to African historical research in contemporary times. Hence, Africans rely on foreigners on this innovative approach to African historical research.

9.3.5 Lack of computing resources to support big data projects in African historical research

Big data computational resources are very expensive and as a result they are not affordable to most research institutions in Africa. Hence, the inability of African historians to develop big data platforms for historical researches.

10. CONCLUSION

The impact of big data in African historical research has profoundly strengthened researchers whose research interest lies in African history to access, analyse and interpret large volumes of historical data on big data platforms. Hence, this research examines the significance of big data in African historical research. The study reveals that big data systems such as the Slaves Voyages database and the African Origin Project provides a platform for Africans to preserve and reconstruct their history, promote African historical scholarship and decolonise African history. However, lack of skills in big data analytics, lack of computing resources to support big data projects in African historical research, ethical concerns and language barrier are considered as some of the factors limiting the adoption of big data platforms in African historical researches. Ultimately, while big data serves as powerful and innovative source of data for African researches, the need to balance this computational approach with scholarship in traditional historical research is paramount to promote accurate, inclusive and in-depth understanding of African history.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

- Ade Ajayi, J. F., & Alagoa, E. J. (1974). Black Africa: The historians' perspectives. *Slavery, Colonialism, and Racism*, 103(2), 125–134.
- Afolabi, F. (2012). *Historiography and methods of African history*. Oxford Bibliographies.
- African Origin Database. (n.d.). *Slave voyages*. <https://www.slavesvoyage.org>
- Bhalla, R. (2019). *Introduction to big data*. Lovely Professional University.
- British Library. (2025). *Welcome to the Endangered Archives Programme*. <https://www.eap.bl.uk>
- Calvino, I. (2025). *Timbuktu: Rescuing books that bear the human soul*. <https://www.gerryco23.wordpress.com>
- Chief Data Officer Magazine. (2024). 80% of your data is unstructured and potentially unprotected—4 strategies to tackle this. <https://www.cdomagazine.tech/opinion-analysis/80-of-your-data-is-unstructured-and-potentially-unprotected-4-strategies-to-tackle-this>
- Cieslik, K., & Margocsy, D. (2022). Datafication, power and control in development: A historical perspective on the perils and longevity of data. *Progress in Development Studies*, 22(4), 352–373.
- Daryle, W. (2018). *Digital approaches to the history of the Atlantic slave trade*. Oxford University Press.
- Eijnatten, J., Pieters, T., & Verheul, J. (2013). Big data for global history: The transformative promise of digital humanities. *Low Countries Historical Review*, 128(4), 55–77.
- Elaine, J. (2011). New website to trace origins of enslaved Africans. *International Research News of Arts and Humanities, Emory State University*.
- Elits, D. (2021). The Trans-Atlantic Slave Database: Origins, developments, contents. *Journal of Slavery and Data Preservation*, 2(3), 1–8.
- Fourie, J. (2016). The data revolution in African economic history. *The Journal of Interdisciplinary History*, 47(2), 193–212.
- Gutmann, M. P., Merchant, E. K., & Roberts, E. (2018). Big data in economic history.

- Journal of Economic History*, 78(1), 268–299.
- Iroju, O., & Ojerinde, O. (2016). Big data in healthcare: Prospects, challenges and resolution. In *Proceedings of Future Technologies Conference*, San Francisco, United States (pp. 1153–1157).
- Isaacman, A. F., Lalu, P., & Nygren, T. I. (2005). Digitisation, history and the making of a post-colonial archive of South African liberation struggles: The Aluka project. *African Today*, 52(2), 55–77.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Misevich, P., Domingues, D., Eltis, D., Khan, M., & Radburn, N. (2017). An online database explores nearly 36,000 slave voyages that occurred between 1514 and 1866. <https://www.smithsonianmag.com/history/digital-archive-slave-voyages-details-largest-forced>
- Muleka, M. A. (2024). The future of historical writing: Will big data rewrite the past? *Journal of Research and Academic Writing*, 1(1), 22–30.
- Rainier, C. (2003). Reclaiming the ancient manuscripts of Timbuktu. *National Geographic News*. <https://www.archive.ph>
- Rajan, R. S., & Heinz, R. (2007). Building a digital library of scholarly resources from developing world: An introduction to Aluka. *African Arts*, 40(2), 1–4.
- Reutter, L. (2022). Constraining context: Situating datafication in public administration. *New Media & Society*, 24(4), 903–921.
- Saeed, N., & Husamaldin, L. (2021). Big data characteristics (V's) in industry. *Iraqi Journal of Research*, 8, 1–9.
- Statista. (2024). Number of languages spoken in Africa 2022, by country. <https://www.statista.com>
- Steven, E. J. (2016). *Roberto Busa, S.J. and the emergence of humanities computing: The priest and the punched cards*. Routledge.
- Stewart, C. (2021). *What's in the Manuscript of Timbuktu? A survey of the contents of 31 library*. Cambridge University Press.
- Tran, P., & Silva, A. (2002). History as a data science: Missing data imputation on the slave voyages dataset. *Association for Computing Machinery*, 1–7.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). This publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© Copyright (2025): Author(s). The licensee is the journal publisher. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:

<https://pr.sdiarticle5.com/review-history/138163>