Advances in Social Sciences and Management
November 2023, Vol-1, No-11, pp. 58-77

Digitizing the Cultural Capital: Harnessing Digital Humanities for Heritage Preservation in Bujumbura, Burundi

James Hutson[1] (ORCID: 0000-0002-0578-6052), Pace Ellsworth[2] (ORCID: 0009-0008-5926-6571), Matt Ellsworth[2] (ORCID: 0009-0005-6387-5984), & Jean Bosco Ntungirimana[3]

1. Department Head of Art History and Visual Culture, College of Arts and Humanities, Lindenwood University, Saint Charles, MO, USA
2. Preserving Records Everywhere (PRE), USA
3. ABADBU Burundian Association of Librarians Researchers and Archivists

Abstract:
In an era where the erosion of cultural heritage is increasingly prevalent, there exists a critical imperative to explore and implement innovative methods for the preservation and revitalization of cultural identities, as exemplified by the urgent situation in Bujumbura, Burundi. Central to this study is the exploration of innovative digital methodologies for archiving a wide spectrum of cultural artifacts, including both notable and everyday heritage elements, in Bujumbura. Traditional approaches to biographical and historical profiling have predominantly focused on official records and significant events, often neglecting the richness of personal experiences and everyday interactions that substantially shape cultural identities. To address this, the study introduces a multi-disciplinary approach that incorporates an array of archival materials, such as government records, personal correspondences, and family memorabilia. These diverse data forms are digitized and semantically classified using biotags, chronotags, and geotags, constructing a comprehensive "space-time continuum" for individuals and communities. The innovative Holistic Archival Personality Profiling Model (HAPPM) is utilized, leveraging large language models for a deeper understanding of personalities and cultural contexts. This methodology not only preserves but also revives cultural heritage, including recreating spoken languages and local dialects through historical linguistic principles. The study exemplifies how technology can bridge the gap between past and present cultural expressions, reinforcing native languages and enriching the appreciation of diverse human identities. This approach is pivotal in offering a more inclusive and holistic method for preserving the global tapestry of cultural heritage, as demonstrated in the Bujumbura case study.

Keywords: Cultural preservation, Digital humanities, Semantic analysis, Linguistic diversity, Archival data aggregation

INTRODUCTION

The recent history of Bujumbura, Burundi, and its neighbors during the Rwandan genocide is marked by a series of tumultuous events that significantly impacted the sociopolitical landscape of the region and contributed to the challenge of incomplete and undigitized records. In late April 1972, a Hutu attack in southern Burundi triggered a military response leading to the deaths of approximately 200,000 Hutus over three months (Russell, 2015). This period of ethnic bloodshed was followed by political upheavals, including the 1987 coup by Major Pierre Buyoya (1949-2020) and subsequent violent confrontations between the Tutsi ruling class and the Hutu majority, resulting in around 150,000 deaths (Saidi & Oladimeji, 2015). Following these events, the

assassination of Burundian Hutu president, Melchior Ndadaye (b.1953), in October 1993, further destabilized the region, igniting a civil war that claimed tens of thousands of lives and displaced hundreds of thousands. This event, coupled with the assassination of Rwandan President Juvénal Habyarimana (b.1937) in April 1994, catalyzed the Rwandan genocide, exacerbating violence in Burundi (Reyntijens, 2006). The influx of refugees and activities of armed groups in both countries created a complex humanitarian crisis. The memory of the 1994 Rwandan genocide extends beyond Rwandan borders, with neighboring countries, including Burundi, grappling with the aftermath and its effects on regional stability (Jaji, 2017).

The late twentieth century was marked by lawlessness, human rights violations, and a lack of effective governance for the region, which contributed to incomplete and poorly maintained records. Furthermore, the history of political instability, ethnic tension, and violence in Burundi resulted in a fragmented administrative system (Nkurunziza, 2022). The lack of a comprehensive archival system was compounded by the challenges of digitizing and preserving records during these turbulent times. The destruction of records, displacement of people, and the subsequent lack of reliable data management systems led to significant gaps in historical and administrative records (Bizimana, 2020). The situation underscores the need for innovative approaches to reconstruct, digitize, and preserve historical and cultural records to ensure the continuity and accessibility of the rich cultural heritage and administrative history of the country.

Unfortunately, the field of biographical profiling, integral to historical and psychological research to assist in reconstituting the country's history, has traditionally depended on a methodology constrained by the limitations of its data sources (Hadzigeorgiou, 2021). This conventional approach, heavily reliant on official records and notable events, has formed the foundation for constructing personality profiles and historical narratives (Smith, 2012). However, as highlighted by Neale, Henwood, and Holland (2012), this reliance on limited and fragmented datasets poses significant challenges. The essence of human experiences, particularly the daily occurrences and personal interactions that fundamentally shape an individual's personality, often remains unexplored and undocumented in these profiles. This methodological oversight results in a superficial portrayal of lives, emphasizing prominent or recorded events while neglecting the rich tapestry of everyday experiences (Clark & Watson, 1988; Dhruvarajan, 1981; Haehner et al., 2023; Reis, 2013).

The primary issue with this traditional method lies in its inherent bias towards events and records deemed 'significant' in a historical or societal context. Official records, such as birth and death certificates, marriage records, and notable achievements, form the core of this data (AbouZahr et al., 2021; Makinde et al., 2023; Suthar et al., 2019). However, these records capture only milestones, overlooking the myriad of everyday experiences that shape character, beliefs, and worldview (Wilkens & Sanford, 2009). Historical figures like Abraham Lincoln or Marie Curie are typically profiled based on their significant contributions and publicly known facts (Pasachoff, 1996; Thomas, 2008). Yet, often missing are elements of their daily lives, such as personal letters, diaries, and social interactions, which provide a more nuanced understanding of their personalities and decisions (Jolly & Stanley, 2005; Stanley, 2004).

To overcome these limitations, the proposed approach extends beyond merely expanding the dataset. The Holistic Archival Personality Profiling Model (HAPPM) redefines the boundaries of historical and biographical research, incorporating a more inclusive exploration of human experiences (Xu et al., 2023). Central to this approach is the digitization and transformation of

diverse data forms into machine-readable text, enabling semantic classification using biotags, chronotags, and geotags, effectively organizing it within specific spatial and temporal contexts (Davoudi & Traviglia, 2023; Hussain et al., 2022; Tasse et al., 2017; Ariano et al., 2017; Hu et al., 2022). To facilitate this expansive research initiative, researchers collaborated with ABADBU and the Burundian Association of Librarians, Researchers, and Archivists to digitize an extensive array of over 100,000 physical records with potentially more than one million in the future phases. These records encompass a diverse range of documentation, capturing the multifaceted nature of administrative and personal life in Burundi. The scope of materials digitized is comprehensive, encompassing everything from government edicts, passport applications, letters inquiring about retirement pensions, and election records, to architectural plans for building a clinic. This exhaustive collection represents the gamut of administrative records owned by the government. A critical aspect of this project was securing licensing rights to publish the metadata and to extract the text for broader accessibility and research purposes.

Along with Burundi, other valuable information was retrieved from their neighbors to the west. Given the status of the Democratic Republic of the Congo (DRC) as a precredit economy, records from the country included detailed voter information—names, dates of birth, places of birth, and locations of voter registration—posted on the internet before being taken down. This data, representing approximately 10% of voters, lacked personal addresses, phone numbers, and emails, but allowed researchers to aggregate familial connections and trace genealogies. For example, the data could be analyzed to determine parental relationships and was meticulously organized in reverse alphabetical order by family name, with a focus on appearances in records before 1950. Additionally, geographical analyses were conducted to ascertain the prevalence of specific family names in various villages prior to 1950, offering a unique historical insight into familial and demographic patterns over 72 years from 2012. This comprehensive archival effort also extended further afield. In Monrovia, Liberia, researchers encountered a significant collection of theses and dissertations from university graduates. These academic works, bound individually and displayed on walls, were not digitized, rendering their contents inaccessible except through physical examination. The lack of digital accessibility highlighted the importance and urgency of digitizing and preserving such academic and cultural artifacts, underscoring the broader implications and potential of the HAPPM in transforming the landscape of historical, cultural, and academic research.

The impetus for this study emerges from a profound necessity to preserve and revitalize the rich tapestry of cultural heritage, particularly in postcolonial contexts like Burundi. The article foregrounds a real-world application of scalable cultural preservation techniques, with a meticulous focus on archival records from Bujumbura, Burundi. This innovative approach transcends traditional methodologies, employing the HAPPM to digitize and semantically analyze a wide array of archival materials. Another innovative feature of the methodology lies in its application of large language models (LLMs). When these models are integrated with the enriched dataset, they facilitate a unique form of interaction, allowing users to "converse" with the past (Hutson & Ratican, 2023). This form of engagement transcends traditional fact recollection, offering a dynamic interaction with the personalities and experiences of historical figures. The integration of LLMs adds a new dimension to the data, introducing a rich semantic layer that enhances the depth and quality of information available for analysis and interpretation (Varshney et al., 2023; Wu et al., 2023).

The endeavor will not only enable Burundians to access and reconnect with their own history but also provides a valuable resource for global researchers. By integrating varied and often overlooked records, ranging from official government documents to personal correspondences, this model offers a more nuanced, comprehensive view of Burundian past. This methodology is particularly pivotal in reconstructing a more accurate postcolonial narrative of Africa, one that is richly layered and reflective of both the monumental and the mundane aspects of history. The potential benefits of this study are manifold, extending beyond academic research to impact cultural understanding, identity formation, and the preservation of collective memory. It represents a significant stride in safeguarding the cultural heritage of Burundi and contributes meaningfully to the broader discourse on postcolonial historical narratives in Africa.

LITERATURE REVIEW

The pursuit of understanding personalities and constructing historical biographies has historically been grounded in traditional methodologies, each characterized by unique approaches and theoretical foundations. In the realm of personality profiling, conventional methods have primarily relied on psychometric evaluations and psychoanalytic theories. These methodologies aim to decipher the complexities of human personality through structured instruments and interpretive frameworks (Post, 2014). These traditional approaches encompass a spectrum from Freudian psychoanalytic theory to the Myers-Briggs Type Indicator, each contributing essential insights into the multifaceted nature of personality (Parker, 1998). Simultaneously, the discipline of historical biography has traditionally been characterized by a narrative-centric, chronologically organized approach to recounting the lives of individuals. This approach typically emphasizes public achievements and well-documented events (Meister, 2018). It often focuses on the role of key figures within the larger context of historical events, thereby shedding light on their contributions and influence (Banner, 2009). Though personality profiling and historical biography differ in focus and methodologies, they converge in their aim to illuminate the human experience, albeit through varying perspectives.

The evolution of personality assessment techniques has progressed from ancient methods based on physical characteristics to more empirical, scientifically grounded approaches. Historically, techniques such as humorism, phrenology, and physiognomy linked physical attributes to personality traits, laying the early groundwork for personality assessment. These approaches transitioned significantly with the exploration of the lexical hypothesis by Sir Francis Galton (1822-1911) in the late nineteenth century, which marked a move towards systematic, language-based analysis of personality traits (Cheung et al., 1998). The twentieth century further shifted the focus from typological to dimensional measures of personality, with notable developments like the Rorschach test and the creation of the Minnesota Multiphasic Personality Inventory (MMPI) (Craik, 1986; Butcher, 2010). In contemporary times, personality assessment methods have diversified to include self-report inventories, projective tests, behavioral observations, and brain imaging techniques, aiming to provide a comprehensive view of personality traits, with an emphasis on linking brain activity to different personality characteristics. This evolution reflects a significant shift towards more nuanced, sophisticated, and empirically validated techniques in understanding human personality (Mosca et al., 2023; Selzer et al., 1987).

Psychometric tools represent another approach in personality assessment and encompasses a diverse array of techniques that provide vital insights into individual personalities. Self-report inventories, pivotal in this domain, involve individuals responding to statements about themselves to assess various personality traits, based on the premise of accurate self-

descriptions, and are widely used in both clinical and research settings (Boyle et al., 2014; Christensen et al., 2019). Projective techniques, such as the Rorschach Inkblot Test and the Thematic Apperception Test, present ambiguous stimuli to individuals and interpret their responses as reflections of underlying personality traits, including unconscious motives (Bender et al., 2014). Behavioral observation, another key method, relies on observing and recording behavior in various situations to assess personality traits, proving especially useful where direct self-reporting is not feasible (Larkin et al., 2021). Additional methods include observational measures, ratings by others, and objective performance tests, underscoring the need for a comprehensive approach that combines various techniques for a more accurate understanding of personalities (Al-Dajani et al., 2016). In organizational psychology, these tools are crucial in the hiring process, with assessments like cognitive ability tests, personality and behavioral assessments, and situational judgment tests providing objective insights into suitability of candidates, thereby enhancing recruitment and team effectiveness (Mussel et al., 2016). The field of personality assessment has seen significant advancements with these varied psychometric tools, highlighting their importance across clinical and organizational contexts.

Along with psychometric methods, the utilization of official records in personality profiling has been significantly enhanced by digital technology, leading to the rise of digital footprints as a key source of data. Research by Lambiotte (2014) and Buraya (2018) underscores the potential of digital records, like social media activity and online interactions, in revealing personality traits, offering insights into private traits, motivations, and emotional makeup. However, caution is necessary in interpreting these records. Alison (2002) highlights the need for nuanced understanding of the relationship between behavior and background characteristics, cautioning against overreliance on naive trait approaches, especially in criminal investigations, and pointing to ethical and legal implications. Given the vulnerability of such populations, research further emphasizes the importance of addressing privacy and ethical concerns when using digital records for personality assessment (Alexander III et al., 2020). Additionally, accomplishment records, such as work performance evaluations, are valuable for personality assessment, aiding in predicting job performance and satisfaction (Connelly et al., 2022). These records, often analyzed through personality self-report inventories, involve assessing agreement with statements designed to measure stable personality traits, underscoring the multifaceted nature of official records in understanding and predicting individual behavior in various contexts (Gao et al., 2023).

At the same time, classical methods in historical research, essential for understanding and interpreting the past, are closely intertwined with the discipline of personality profiling. As outlined by Grigg (1991), the importance of systematic search procedures in historical research is paramount, especially given the diversification of scholarship and the expansion of usable sources. Fundamental tools such as palaeography, diplomatics, chronology, and epigraphy serve as cornerstones in the historical sciences, aiding historians in the evaluation and analysis of source materials. The growing accessibility of archives and the proliferation of sources have necessitated more refined and systematic approaches to historical research (Friedrich et al., 2017). In classical studies, the philological method, pivotal to the study of ancient texts since the late 19th century, plays a critical role. Kazarov (2019) highlights the essentiality of philological and historical research methods in classical studies, where interpreting ancient texts and inscriptions forms the crux of scholarly inquiry. This approach enables a deeper comprehension of historical texts through the linguistic nuances of ancient writings. Additionally, Wadhwani (2017) discusses the application of historical methods in organization and management studies, where interpreting the past involves abductive reasoning, analyzing historical data from a contemporary perspective.

This dynamic nature of historical research, where the past is viewed through the lens of present understanding, is further evidenced by historians like Wolf (2018), who liken their research approach to that of scientists. Historians rely on a comprehensive array of primary and secondary sources, including oral histories, to construct accurate and reliable narratives of the past, reflecting the complexities and nuances of historical events and phenomena.

With these approaches to reconstructing histories, perhaps the best known is still a biographical one. Biographical writing in history encompasses a range of techniques and approaches to enrich the narratives about individuals. This genre typically employs a narrative structure following the subject's life from birth to death, embedding historical context and posthumous events as necessary. The multifaceted research process involves gathering information from diverse sources such as anecdotes, interviews, letters, public records, and historical documents, with biographers often documenting their research process to provide insights into their methodologies and challenges encountered (Krasny, 2006; Renders, 2013). Some authors opt for historical fiction, blending factual information with fictional elements to create more engaging narratives (Daniels, 2023). Moreover, collective biographies, focusing on groups of individuals, offer broader perspectives on historical events and societal dynamics, enabling a more comprehensive view of the societal and historical contexts of the subjects (Rollyson, 2013; Evans & Reynolds, 2012). This biographical approach uses individual life stories as a lens to understand broader historical contexts and societal changes, underscoring the impact of individual lives on historical events and offering a unique perspective on how individuals are shaped by and shape their times. Incorporating narrative elements like character development, setting, and plot progression is crucial, making the story engaging and informative, and providing a deeper understanding of the subject's life and the historical context. Effective use of these narrative elements in biographies can offer valuable insights into the lives of historical figures and influential individuals, playing a significant role in understanding and interpreting the past (Mann, 2009; Oakley, 2010).

Finally, in the domain of digital humanities, contemporary approaches to documentation, archiving, and contextualizing narratives of various eras and civilizations have witnessed a transformative shift, driven by technological advancements and interdisciplinary collaboration. This paradigm shift is characterized by the integration of digital technologies in the process of documenting and preserving historical records, thereby enabling more dynamic and interactive modes of engaging with the past. For instance, one such approach to documentation leverages advanced technologies like optical character recognition (OCR) and natural language processing (NLP) to convert historical texts and manuscripts into digital formats. This digitization process not only preserves the physical artifacts but also makes them accessible to a global audience. Scholars like Murry (2020) and Lydon (2021) have emphasized the role of digitization in democratizing access to historical documents, noting how digital archives have become indispensable resources for researchers worldwide.

Moreover, digital archiving in the humanities extends beyond mere digitization. It involves the creation of comprehensive digital repositories that provide contextual information about the artifacts. These repositories often include metadata, annotations, and interpretative layers that enrich the primary sources. For instance, Locke and Wright (2021) highlight the use of interactive digital platforms that allow users to explore historical artifacts in a more engaging and informative way. These platforms often include features like zoomable images, 3D reconstructions, and hyperlinked annotations, which offer a deeper understanding of the cultural

and historical context of the artifacts (Howland et al., 2020; Jacobs, 2022). On the other hand, contextualizing narratives of different eras and civilizations through digital humanities approaches also involves the use of big data analytics and machine learning algorithms. These technologies enable the analysis of large-scale historical data to uncover patterns, trends, and insights that were previously inaccessible. Researchers like Lombardi and Marinai (2020) have demonstrated how machine learning (ML) can be employed to analyze patterns in historical texts, providing new perspectives on social, cultural, and political dynamics of past civilizations.

Furthermore, digital humanities projects often emphasize collaborative and interdisciplinary work, bringing together experts from various fields such as history, linguistics, computer science, and archival studies. This collaboration fosters a more holistic approach to understanding and interpreting historical narratives. As noted by Su (2020), interdisciplinary projects in digital humanities offer unique opportunities for synthesizing diverse methodologies and perspectives, leading to more nuanced and multifaceted interpretations of historical events and phenomena. In essence, digital humanities approach to documentation, archives, and contextualizing narratives represents a confluence of technology and traditional humanities scholarship. These approaches not only preserve and make historical records more accessible but also provide innovative tools for analyzing and interpreting these records. The digital humanities field thus plays a crucial role in shaping our understanding of different eras and civilizations, offering new ways of engaging with and interpreting the past. As we continue to advance in the digital age, the importance of digital humanities in preserving and understanding our cultural heritage becomes increasingly evident, underscoring its significance in contemporary scholarship.

The construction of historical biographies, a multifaceted endeavor, has traditionally utilized a range of methodologies, including personality assessment techniques, psychometric tools, personality profiling, classical methods of historical research, and, more recently, digital humanities methodologies. These approaches collectively contribute to a comprehensive understanding of historical figures, blending psychological analysis with historical context. Techniques like self-report inventories and projective methods in personality assessment, along with classical methods such as philological analysis, have provided deep insights into individual personalities and their historical significance. However, these approaches often encounter challenges when dealing with fragmentary and non-digitized records, as seen in the Burundian case study. Such records, while rich in information, may be inaccessible or incomplete, limiting the scope of traditional biographical research.

METHODOLOGY

The Holistic Archival Personality Profiling Model (HAPPM) (Table 1), in this context, represents a significant advancement in historical biography construction. By integrating methodologies from both traditional and digital humanities disciplines, HAPPM addresses the limitations inherent in fragmentary non-digitized records. Leveraging ML, NLP, and LLMs, HAPPM offers a novel approach to deciphering and contextualizing historical narratives. This model builds upon previous methods, enhancing the analysis of personalities and historical contexts with the depth and precision offered by digital tools. The strengths of HAPPM lie in its ability to process and analyze large volumes of diverse data, including fragmentary and non-standardized records. This capability is particularly beneficial in cases like Burundi's, where conventional records are incomplete or inaccessible. By applying ML and NLP, HAPPM can extract and interpret data from these challenging sources, providing new insights into historical figures and eras. However, the reliance on digital tools also brings inherent challenges, such as the potential for over-reliance on

algorithmic interpretations and the need for constant updates to technology and methodologies. Thus, while traditional approaches to constructing historical biographies have provided foundational insights, the integration of digital humanities methodologies, as exemplified by HAPPM, offers a more robust and nuanced understanding of historical narratives, especially in the face of fragmentary and non-digitized records. This integrated approach not only preserves the integrity of traditional historical research but also opens new avenues for exploration and understanding in the field of historical biography.

Table 1: The Holistic Archival Personality Profiling Model (HAPPM)

Component	Description	Purpose
Data Source Diversification	Inclusion of a wide range of archival materials like personal correspondences, social media footprints, and family memorabilia.	To capture a more nuanced view of individuals, encompassing everyday experiences and personal interactions.
Digitization and Data Transformation	Conversion of various data forms, including handwritten documents, into machine-readable formats.	Facilitates the extraction and analysis of comprehensive data, preserving historical documents and enhancing accessibility.
Semantic Data Classification	Use of biotags, chronotags, and geotags for organizing data within specific spatial and temporal contexts.	Enhances data organization and contextualization, providing a detailed understanding of subjects' life experiences.
Advanced Language Models	Employment of large language models for analyzing digitized and tagged data.	To simulate conversations with historical figures, offering a holistic representation of personalities.
Innovative Data Analysis Techniques	Development and application of advanced data analysis techniques for interpreting extensive data.	Ensures a more accurate representation of individuals' lives by uncovering patterns and insights from complex datasets.
Ethical and Privacy Considerations	Establishment of guidelines for ethical use of personal data and maintenance of transparency in data handling.	Respects privacy and consent, ensuring ethical research practices.
Interdisciplinary Collaboration	Collaboration between historians, data scientists, and technologists.	Enriches biographical research with diverse perspectives and expertise.
Training and Capacity Building	Provision of training in digital tools and data analysis methods.	Enhances researchers' skills in modern research methodologies.
Continuous Methodological Evaluation	Regular evaluation and refinement of methodologies through feedback and iterative improvement.	Ensures methodologies remain relevant, effective, and ethically sound

Burundi presents itself as an ideal case study for the HAPPM approach, primarily due to its unique historical context and the nature of its archival records. Unlike its neighbor Rwanda, which experienced intense and widely documented genocide, the historical trajectory of Burundi has been characterized by a lack of consistent governance, economic instability, and relative obscurity in the global context, particularly during the tumultuous periods of the 1960s and 1970s. This period was marked by looting, bombing, and a general absence of structured record-keeping, which has resulted in a fragmented and piecemeal archival landscape.

The archives in Burundi, especially when compared to those in countries like Liberia, which has a more comprehensive collection of saved newspapers and other historical documents, reveal a stark difference. The Archie P. Williams Center for Historical Information and Archives Research, for example, boasts a better collection than the University of Burundi. The archival situation in

Burundi is a patchwork of materials assembled by archivists, often relying on foreign entities for context and content. Since the 1990s, the need for context in understanding the national history has been critical, particularly for establishing financial rails and economic stability. This archival deficit is one of the factors contributing to Rwanda's comparative advancement over Burundi, with parallels observable in other regions such as Thailand and Laos.

Bujumbura, as the cultural capital of Burundi, offers an unparalleled opportunity for the HAPPM approach. The city houses over a million records that have never been digitized or analyzed on a large scale. This untapped repository includes significant information not just about Burundi but also about the broader East African region, including Rwanda. The HAPPM approach, with its integration of ML, NLP, and LLMs, is uniquely suited to tackle the challenges posed by these records. Its capability to process, digitize, and analyze large volumes of diverse, fragmented data can unveil new insights into the history, societal changes, and cultural dynamics of the region. Furthermore, the application of HAPPM in Burundi can set a precedent for similar research in other regions with under-documented histories, offering a model for how digital humanities can bridge gaps in historical knowledge and understanding.

The process of scanning and digitizing these extensive records necessitates a highly collaborative and resource-intensive effort. The initiative involves cooperation with the charity association of a private research organization that is tasked with the digitization process. The equipment used for this purpose includes flatbed scanners, laptops, desktop scanners, cell phones, and external hard drives. The agreement for accessing and digitizing these records involves a partnership with the Ministry of East African Community Affairs, Youth, Sports, and Culture. This ministry, while operating independently, works in tandem with the University of Burundi, which plays a crucial role in the scanning process.

The University of Burundi is responsible for conducting the scanning of the records, with an overarching aim to produce a study that could serve as a benchmark for similar initiatives in other regions. This involvement not only facilitates the digitization process but also empowers the university to conduct research on its own national history, leveraging the newly digitized data. The digitization process is meticulous and comprehensive, ensuring that all records are accurately scanned and uploaded to servers for subsequent analysis and research.

The methodological approach employed in the HAPPM is multifaceted, combining traditional archival techniques with advanced digital technologies to create a comprehensive framework for historical and biographical research. The methodology can be delineated into several key phases:

Data Collection and Curation
The initial phase involves the systematic collection of archival materials. For the Burundi case study, this entails collaborating with local institutions and authorities to access a wide range of records, including government documents, personal correspondences, voter registries, and various other administrative and personal records. The curation process involves categorizing these documents based on their relevance, historical significance, and the quality of information they provide.

Digitization and Data Conversion
The collected records are digitized using high-resolution scanners, ensuring that the digital copies are clear and complete. Techniques like Optical Character Recognition (OCR) are employed to

convert these records into machine-readable formats. This step is crucial for facilitating subsequent data analysis and ensuring that the records are preserved in a durable, accessible format.

Semantic Analysis and Tagging

Once digitized, the records undergo a semantic analysis process. Utilizing NLP techniques, the data is analyzed for key themes, patterns, and linguistic markers. Data is systematically classified using biotags (related to individuals), chronotags (pertaining to temporal aspects), and geotags (concerning geographical information). This semantic tagging enables the contextualization of the data within specific spatial and temporal frameworks, making it easier to analyze and understand in relation to historical events and personal histories.

Integration with Large Language Models (LLMs)

The enriched dataset is then integrated with advanced LLMs. These models are capable of "conversing" with the data, allowing researchers to interact dynamically with the historical narratives. This interaction is not limited to mere fact recollection but extends to a more nuanced engagement with the personalities and experiences of historical figures. The LLMs offer an additional layer of semantic interpretation, enhancing the depth and richness of the information extracted from the archives.

Data Analysis and Interpretation

The final phase involves the analysis and interpretation of the data. Researchers employ both quantitative and qualitative methods to extract insights from the data. This might include analyzing voting patterns, tracing familial connections, studying sociopolitical trends, or understanding the evolution of cultural practices. The interpretation phase is guided by historical knowledge, contextual understanding, and the research objectives set forth at the project's outset.

Preservation and Accessibility

Ensuring the long-term preservation of the digitized records and making them accessible to researchers, historians, and the general public is a key component of the HAPPM approach. This involves storing the data in secure, sustainable digital formats and creating user-friendly interfaces for accessing the archives.

The process of taking the output of these steps and creating conversable AI clones of historical figures using documentation is a groundbreaking endeavor that requires a meticulous and culturally informed approach (Table 2). This process begins with the careful selection of historical figures and the compilation of an extensive array of primary source documents associated with them. These documents may include personal letters, speeches, diaries, and other writings, complemented by third-party accounts such as biographies, news articles, and scholarly research. The thorough compilation of these materials is crucial, as it forms the foundational dataset from which the AI will learn and recreate the personality and thought patterns of the historical figure. Once the historical figures are selected and the relevant documents compiled, the next critical step is a deep cultural and contextual analysis of the material. This stage addresses the challenges highlighted by research from Harvard University, which found discrepancies in AI responses based on cultural distances from the United States (Atari et al., 2023). To mitigate such biases and ensure cultural authenticity, the involvement of experts well-versed in the culture and primary sources of the subject is indispensable. These experts, whether they are historians, linguists, or

cultural scholars, provide invaluable insights into the nuances, idioms, and context-specific meanings in the source material. Their expertise ensures that the AI model is not only factually accurate but also culturally and historically representative of the figures it seeks to emulate.

Table 2: Steps to Creating AI Clones Using GPTs

Step	**Details**
Step 1: Define the Project Scope and Objectives	**Objective:** Determine the specific historical figures and periods you want to focus on. **Audience:** Decide who the target audience is (e.g., students, researchers) and tailor the content to be age-appropriate.
Step 2: Gather Primary Source Information	**Research:** Collect primary source documents relevant to the historical figures and periods of interest. This can include letters, speeches, diary entries, and other historical texts. **Digitize:** Ensure all documents are in a digital format, preferably text-based (e.g., PDF, Word, plain text).
Step 3: Building Your Custom ChatGPT/OpenAI	**OpenAI Account:** First, you need a ChatGPT Plus or Enterprise account from OpenAI. Once you have it, log in to your account on OpenAI's website (https://chat.openai.com/). **Create Your GPT:** In the sidebar on OpenAI's site, click "Explore" and then "Create a GPT." You'll be presented with a split screen where you can enter prompts to build your chatbot on one side and preview its responses on the other. **Input Instructions:** Enter your specific instructions for the chatbot in the message box of the "Create" page. The GPT builder will suggest a chatbot name, profile picture, and default conversation starters, which you can modify as needed. **Advanced Configuration:** Click "Configure" to access advanced settings. Here you can change the chatbot's name, description, profile picture, and more. Most importantly, you can upload knowledge files for your chatbot to reference, such as primary source documents about historical figures. **Save and Share Your GPT:** After configuring your GPT, save it and decide who you want to share it with. You can make it accessible only to yourself, to anyone with a link, to the public, or just within your company if you're on an Enterprise plan.
Step 4: Customizing Age-Appropriate Responses	**Custom Instructions Feature:** ChatGPT has a feature that allows you to add custom instructions, which influence how it generates responses. This feature is available for all ChatGPT plans and is a huge time-saver as it eliminates the need to constantly reiterate your preferences. **Setting Up Custom Instructions:** The custom instructions feature includes two boxes: one for adding information about yourself and your role, and another for specifying how you want ChatGPT to format its responses, including tone and style. **Response Filter:** You can start with a response filter, which helps ensure that ChatGPT provides accurate, factual, and nuanced answers. For example, you can instruct it to provide nuanced, factual answers, and flag uncertainties.

The next phase involves the processing of this data and the training of the AI model, typically a sophisticated large language model. This stage is where the intricacies of language, speech patterns, and the rhetorical style of the historical figures are translated into a format that the AI can understand and replicate. The data must be meticulously cleaned, organized, and semantically tagged to facilitate effective learning by the AI. The training process is

comprehensive, aimed at enabling the AI to accurately mimic the communication style and intellectual dispositions of the historical figures. It is a delicate balance between ensuring the AI's responses are reflective of the figure's known views and characteristics while avoiding the imposition of contemporary biases or misinterpretations.

Post-training, the clones undergo rigorous testing to verify the accuracy and authenticity of their responses. This is a critical step, as it assesses whether the AI effectively captures the essence of the historical figures in a manner that is both credible and respectful of their legacy. The cultural experts and historians involved in the project play a pivotal role in this phase, evaluating the AI's outputs against the historical and cultural backdrop of the figures. Their feedback is crucial in identifying any areas where the AI may deviate from expected responses, necessitating further refinement and retraining of the model.

Implementing continuous feedback loops is essential for the ongoing improvement and accuracy of the AI clones. As new historical data emerges or as further research sheds new light on aspects of the lives and thoughts of figures, the AI models are updated accordingly. This iterative process ensures that the AI clones remain dynamic and evolve in line with the expanding body of knowledge about the historical figures they represent. Finally, these clones are introduced to public interaction and educational contexts. This step is where the fruits of the laborious process are realized, as the clones offer unique, interactive experiences for learning about historical figures. They serve as innovative educational tools, providing students and enthusiasts with an engaging way to explore history and understand the personalities that shaped it. This interactive dimension adds a new layer to historical education, allowing for a more immersive and personalized experience of learning about the past. The clones can be utilized in various settings, from classrooms to museums, offering responses and engaging in dialogues that bring historical figures to life in a manner previously unattainable.

The adoption of this methodology globally holds significant promise for the field of historical research and education. By leveraging advanced AI technologies and ensuring cultural accuracy and sensitivity, researchers can recreate conversable clones of a wide range of historical figures, each offering unique insights into their era and society. This approach not only enhances our understanding of the past but also democratizes access to historical knowledge, allowing a broader audience to engage with history in a meaningful and interactive way. Therefore, the creation of conversable AI clones of historical figures represents a novel convergence of technology, historical scholarship, and cultural sensitivity. The process outlined above, when adopted globally, has the potential to revolutionize the way we interact with and understand history. It offers an exciting future for the field of digital humanities, where technology serves as a bridge connecting us more closely with the past and its influential personalities.

RECOMMENDATIONS

Following upon the first phase of the project, which focused on digitizing over 100,000 physical records from various sources, the groundwork has been laid for a more expansive exploration of Burundian history and culture. Through international and interinstitutional collaborations, like that of The University of Burundi, where the scanning the records presents an opportunity to model a successful digitization initiative for similar projects. To further augment these initial datasets, the research team recommend combining these digitized records with online records from the Ministry of Education, which includes the results of exams (such as the PSAT) encompassing over 200,000 names, along with dates of birth and schools attended (Sabiraguha

et al., 2023). The integration of additional information will provide a more comprehensive dataset, enabling a richer analysis of educational trends and demographics over time (Koch et al., 2021).

Moreover, it is essential to consider the data lifecycle, recognizing that the longevity of information is intrinsically linked to the robustness of the systems supporting it (Nguyen et al., 2020). Data stored on a single server or system faces the risk of being lost or becoming obsolete due to technological advancements, changes in data formats, or shifts in institutional management (Dawood et al., 2023). Consequently, the resarch team emphasize the importance of diversifying the storage of data across multiple online servers to mitigate the risk of data loss (Duggineni, 2023). While many databases are securely archived on platforms such as Ancestry.com for genealogical research, they often remain underutilized for broader analytical purposes (Granja, 2023). These databases, which include images and are equipped with software capable of identifying familial connections, offer untapped potential for linguistic and historical analysis. By incorporating similar rich data, he understanding of familial structures and social dynamics within Burundi can be better understood. Another critical resource that has been underutilized is the 1870 census held by the National Archives and Records Administration. This historical document, though licensed for use, has not been extensively explored to answer questions that AI could potentially address (McDowell, 2023). The census data can leverage AI to analyze and uncover insights into the social, economic, and demographic patterns of the period. We also suggest exploring innovative communication methods for making historical data accessible to diverse audiences, including those with disabilities.

One of the major challenges in assembling a representative dataset for countries in sub-Saharan Africa is the fact that their archives, often comprising materials scrounged by locals, presents a contrast in quantity and quality compared to collections held by colonial powers like Portugal and France (Bush, 2013). The European-structured data, while in the hands of post-colonial governments, often remains unexplored due to its external perspective and lack of relevance to native narratives (Ngom & Castro, 2019). However, this data can be invaluable in unraveling the colonial context and understanding historical events from a different viewpoint. There are more records related to these regions in Europe, yet they remain unaggregated due to a lack of demand because the infrastructure and awareness necessary to utilize these resources effectively are absent (Borel-Saladin, 2017). The concept of digital repatriation needs emphasis, requiring a framework that protects digital rights management, ownership, and know-how to monetize information. European countries might not have an immediate need for these records, but they are crucial for validation and research in the originating countries. Therefore, obtaining digital copies, rather than physical ones, is the most pragmatic and efficient solution. Such repatriation efforts have been common in the United States with regards to indigenous populations in North America, but are only now becoming common regarding post-colonial powers and their previous provinces in Africa, the Middle East, and Asia (Christen, 2011; Fitch, 2013; Krupa & Grimm, 2021). This approach respects ownership while ensuring accessibility and usability of the information.

The legal systems in many post-colonial states, influenced by European jurisprudence, often lack access to precedents established in colonial courts, leading to a need to recreate legal foundations. This situation is exemplified by the case of Indian court records. The digitization of such records is not just a matter of preserving history but also of practical necessity for the functioning of contemporary legal systems (Likhoviski, 2021). In countries like Liberia, where even pages from the land registry are only accessible through specific requests and are subject to

alterations, the digitization of records assumes even greater importance (Home, 2021). It prevents the alteration of history and ensures the safe preservation of information. Therefore, this project argues for the usability and accessibility of digitized records, advocating for systems that facilitate easy digitization and retrieval. The Universal Declaration of Human Rights (UDHR), adopted by the United Nations General Assembly, underscores the fundamental rights and freedoms of all human beings. In light of this, the digital repatriation of historical documents as a legal and ethical imperative should be foregrounded. This action aligns with the principles enshrined in the UDHR and the customary international law it reflects (Tünsmeyer, 2022).

The digitization initiative in Burundi outlined here marks a critical step in preserving the cultural heritage of the nation while enhancing the legal and educational utility of historical records. This effort, rooted in principles of digital repatriation and informed by international law and ethical guidelines, addresses the challenges of accessibility and preservation. By fostering digital accessibility and usability, the project not only safeguards historical narratives but also empowers nations to reconnect with their past, thereby nurturing a sense of identity and ownership. Also, the significance of the initiative is further amplified by the planned genealogy symposium. This event aims to unite stakeholders from across East Africa, creating a forum to exhibit the results of the digitization project, discuss the methodologies and technologies employed, and deliberate on the project's broader implications for historical, cultural, and academic research. The symposium is poised to stimulate dialogue and share best practices among archivists, historians, and researchers, promoting collaborative efforts that could inspire similar projects throughout the region.

The comprehensive approach of the HAPPM in guiding the digitization of Burundian archival records is pivotal. This project not only confronts the issue of fragmented, non-digitized records but also exemplifies the application of digital humanities methodologies in revealing the intricate historical and cultural narratives of postcolonial societies. The University of Burundi's involvement in this project is instrumental, aiming to establish a model for archival digitization that can be replicated across East Africa. The genealogy symposium will serve as a crucial milestone in this journey, offering an opportunity to highlight the project's successes and foster regional cooperation in the realm of archival preservation and research.

CONCLUSION

The digitization project in Burundi, focusing on the rich tapestry of its cultural heritage, addresses an urgent need for the preservation and revitalization of historical records. The project emerged from the recognition that a significant portion of Burundi's past, like many postcolonial societies, remains undocumented or trapped in fragile, non-digitized formats. The background of and necessity for the initiative stem from a desire to safeguard these records against the ravages of time and technological obsolescence, ensuring that future generations have access to a crucial part of their cultural identity and history. The methodological approach adopted in this project is groundbreaking. The Holistic Archival Personality Profiling Model (HAPPM) represents a novel synthesis of traditional archival techniques and advanced digital technologies, including machine learning, natural language processing, and large language models. This approach not only allows for the efficient digitization of vast amounts of historical data but also enables a more nuanced analysis and interpretation of this data, shedding new light on the social, cultural, and political dynamics of Burundi’s past. At the same time, it is essential to reiterate that the digitization project in Burundi, still in its nascent stages, has shown significant promise and feasibility through the application of the profiling model. The project, already progressing with the near goal of

digitizing over one million physical records and the incorporation of various online sources, is methodically unveiling a more detailed and expansive narrative of the history of the region. This early success lays a robust foundation for continued research and educational pursuits. Looking forward, the potential of this digitization initiative is vast. Expected to surpass 5,000,000 digitized documents, with additional materials to be sourced, the project underscores an expansive and continuous effort in historical preservation. Currently, the vast collection of documents, integral to cultural and historical heritage, is accessible only through physical handling on-site. The digitization effort is thus pivotal in making this extensive body of knowledge widely available, not just for the local researchers and historians but also for the global academic community.

The forthcoming genealogy symposium is set to play a crucial role in this endeavor. The event will serve as a platform for sharing the insights gained, the methodologies applied, and the best practices discovered through this project. This event is anticipated to foster a collaborative network that will encourage similar preservation and analysis efforts across East Africa and potentially beyond. Such a network could drive a movement towards a more inclusive and comprehensive understanding of history, particularly in postcolonial societies. Thus, the digitization initiative in Burundi exemplifies the transformative power of integrating traditional historical research methods with modern digital technologies. This project transcends mere preservation; it enriches our understanding of the past and ensures that the lessons and experiences of our ancestors continue to enlighten and inspire future generations. As the project progresses, its impact is expected to extend far beyond the current scope, offering invaluable insights and preserving the rich tapestry of Burundi's history for posterity.

Data Availability
Data available upon request.

Conflicts of Interest
The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement
NA

Authors' Contributions
Conceptualization, J. Hutson; Methodology, J. Hutson; Validation, J. Hutson; Investigation, J. Hutson – Original Draft Preparation, J. Hutson; Writing – Review & Editing, J. Hutson.; Visualization, J. Hutson.

REFERENCES

AbouZahr, C., Bratschi, M. W., Cercone, E., Mangharam, A., Savigny, D. D., Dincu, I., ... & Setel, P. (2021). The COVID-19 pandemic: effects on civil registration of births and deaths and on availability and utility of vital events data. *American Journal of Public Health, 111*(6), 1123-1131.

Al-Dajani, N., Gralnick, T. M., & Bagby, R. M. (2016). A psychometric review of the Personality Inventory for DSM–5 (PID–5): Current status and future directions. *Journal of personality assessment, 98*(1), 62-81.

Alexander III, L., Mulfinger, E., & Oswald, F. L. (2020). Using big data and machine learning in personality measurement: Opportunities and challenges. *European Journal of Personality, 34*(5), 632-648.

Alison, L., Bennell, C., Mokros, A., & Ormerod, D. (2002). The personality paradox in offender profiling: A theoretical review of the processes involved in deriving background characteristics from crime scene actions. *Psychology, Public Policy, and Law, 8*(1), 115.

Ariano, G., Colavito, M., & D'Annibale, S. (Eds.). (2017). *Archeomedsites: Mediterranean perspectives.* Gangemi Editore spa.

Atari, M. Xue, M. Park, P. Blasi, D. & Henrich, J. (2023) Which Humans? *https://doi.org/10.31234/osf.io/5b26t*

Banner, L. W. (2009). Biography as history. *The American Historical Review, 114*(3), 579-586.

Bender, D. S., Morey, L. C., & Skodol, A. E. (2014). Toward a model for assessing level of personality functioning in DSM-5, part I: A review of theory and methods. *Personality assessment in the DSM-5,* 35-49.

Bizimana, S. C. (2020). E-government Readiness Assessment for Government institutions in Burundi. *Development, 3*, 4.

Borel-Saladin, J. (2017, December). Data dilemmas: availability, access and applicability for analysis in sub-Saharan African cities. In *Urban Forum* (Vol. 28, pp. 333-343). Springer Netherlands.

Boyle, G. J., Saklofske, D. H., & Matthews, G. (Eds.). (2014). *Measures of personality and social psychological constructs.* Academic Press.

Bush, R. (2013). *Publishing sub-Saharan Africa in Paris 1945-67* (Doctoral dissertation, Oxford University, UK).

Butcher, J. N. (2010). Personality assessment from the nineteenth to the early twenty-first century: Past achievements and contemporary challenges. *Annual Review of Clinical Psychology, 6*, 1-20.

Buraya, K., Farseev, A., & Filchenkov, A. (2018). Multi-view personality profiling based on longitudinal data. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings 9* (pp. 15-27). Springer International Publishing.

Cheung, F. M., & Leung, K. (1998). Indigenous personality measures: Chinese examples. *Journal of cross-cultural psychology, 29*(1), 233-248.

Christen, K. (2011). Opening archives: Respectful repatriation. *The American Archivist, 74*(1), 185-210.

Christensen, A. P., Golino, H., & Silvia, P. (2019). A psychometric network perspective on the measurement and assessment of personality traits. *Preprint.*

Clark, L. A., & Watson, D. (1988). Mood and the mundane: relations between daily life events and self-reported mood. *Journal of personality and social psychology, 54*(2), 296.

Connelly, B. S., McAbee, S. T., Oh, I. S., Jung, Y., & Jung, C. W. (2022). A multirater perspective on personality and performance: An empirical examination of the trait–reputation–identity model. *Journal of Applied Psychology, 107*(8), 1352.

Craik, K. H. (1986). Personality research methods: An historical perspective. *Journal of Personality, 54*(1), 18-51.

Daniels, J. I. (2023). *Anything But Novel: Pushing the Margins in Latin American Post-Utopian Historical Narrative*. University of Alabama Press.

Davoudi, H., & Traviglia, A. (2023). Discrete representation learning for handwritten text recognition. *Neural Computing and Applications*, 1-15.

Dawood, M., Tu, S., Xiao, C., Alasmary, H., Waqas, M., & Rehman, S. U. (2023). Cyberattacks and security of cloud computing: a complete guideline. *Symmetry, 15*(11), 1981.

Dhruvarajan, V. (1981). *Level of self-differentiation: Caste status vs. mundane experience.* The University of Chicago.

Duggineni, S. (2023). Impact of controls on data integrity and information systems. *Science and Technology, 13*(2), 29-35.

Evans, T., & Reynolds, R. (2012). Introduction to this special issue on biography and life-writing. *Australian Historical Studies, 43*(1), 1-8.

Fitch, M. L. (2013). *Native American Empowerment Through Digital Repatriation* (Doctoral dissertation, East Tennessee State University).

Friedrich, M., Müller, P., & Riordan, M. (2017). Practices of historical research in archives and libraries from the eighteenth to the nineteenth century. *History of Humanities, 2*(1), 3-13.

Gao, N., Ananthan, S., Yu, C., Wang, Y., & Salim, F. D. (2023). Critiquing Self-report Practices for Human Mental and Wellbeing Computing at Ubicomp. *arXiv preprint arXiv:2311.15496.*

Granja, R. (2023). Citizen science at the roots and as the future of forensic genetic genealogy. *International Journal of Police Science & Management,* 14613557231164901.

Grigg, S. (1991). Archival practice and the foundations of historical method. *The Journal of American History, 78*(1), 228-239.

Hadzigeorgiou, Y. (2021). Biographical profiling of Nikola Tesla for the creation of an engaging story. *Education Sciences, 12*(1), 12.

Haehner, P., Bleidorn, W., & Hopwood, C. J. (2023). Examining individual differences in personality trait changes after negative life events. *European Journal of Personality,* 08902070231156840.

Home, R. (2021). History and prospects for African land governance: institutions, technology and 'land rights for all'. *Land, 10*(3), 292.

Howland, M. D., Liss, B., Levy, T. E., & Najjar, M. (2020). Integrating digital datasets into public engagement through ArcGIS StoryMaps. *Advances in Archaeological Practice, 8*(4), 351-360.

Hu, B., Zhao, T., Wang, Y., Cheng, J., Howard, R., Chen, Y., & Wan, H. (2022, October). BioTag: Robust RFID-based continuous user verification using physiological features from respiration. In *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing* (pp. 191-200).

Hussain, I., Ahmad, R., Muhammad, S., Ullah, K., Shah, H., & Namoun, A. (2022). PHTI: Pashto handwritten text imagebase for deep learning applications. *IEEE Access, 10*, 113149-113157.

Hutson, J., & Ratican, J. (2023). Life, death, and AI: Exploring digital necromancy in popular culture—Ethical considerations, technological limitations, and the pet cemetery conundrum. *Metaverse, 4*(1).

Jacobs, H. L. (2022). SketchUp and Sketchfab: Tools for Teaching with 3D. *Journal of the Society of Architectural Historians, 81*(2), 256-259.

Jaji, R. (2017). Under the shadow of genocide: Rwandans, ethnicity and refugee status. *Ethnicities, 17*(1), 47-65.
Jolly, M., & Stanley, L. (2005). Letters as/not a genre. *Life Writing, 2*(2), 91-118.

Kazarov, S., & Petrechko, O. (2019). Philological and historical research methods of classical studies. *East European Historical Bulletin, 1* (13), 8-15.

Koch, B., Denton, E., Hanna, A., & Foster, J. G. (2021). Reduced, reused and recycled: The life of a dataset in machine learning research. a*rXiv preprint arXiv:2112.01716.*

Krasny, K. (2006). Into a New Light: Re-Envisioning Educational Possibilities for Biography. *Language and Literacy.*

Krupa, K. L., & Grimm, K. T. (2021). Digital repatriation as a decolonizing practice in the archaeological archive. *Across the Disciplines, 18*(1-2), 47-58.

Lambiotte, R., & Kosinski, M. (2014). Tracking the digital footprints of personality. *Proceedings of the IEEE, 102*(12), 1934-1939.

Larkin, I., Pierce, L., Shalvi, S., & Tenbrunsel, A. (2021). The opportunities and challenges of behavioral field research on misconduct. *Organizational Behavior and Human Decision Processes, 166*, 1-8.

Likhovski, A. (2021). A colonial legal laboratory? Jurisprudential innovation in British India. *The American Journal of Comparative Law, 69*(1), 44-92.

Locke, J. L., & Wright, B. (2021). History Can Be Open Source: Democratic Dreams and the Rise of Digital History. *The American Historical Review, 126*(4), 1485-1511.

Lombardi, F., & Marinai, S. (2020). Deep learning for historical document analysis and recognition—A survey. *Journal of Imaging, 6*(10), 110.

Lydon, J. (2021). Indigenous uses of photographic digital heritage in postcolonizing Australia. *Photography and Culture, 14*(3), 269-296.

Makinde, O. A., Odimegwu, C. O., Udoh, M. O., Adedini, S. A., Akinyemi, J. O., Atobatele, A., ... & Orobaton, N. (2020). Death registration in Nigeria: a systematic literature review of its performance and challenges. *Global Health Action, 13*(1), 1811476.

Mann, S. (2009). Scene-setting: Writing biography in Chinese history. *The American Historical Review, 114*(3), 631-639.

McDowell, E. A. (2023). The children of slavery. *The Black Reparations Project: A Handbook for Racial Justice,* 174.

Meister, D. R. (2018). The biographical turn and the case for historical biography. *History Compass, 16*(1), e12436.

Mosca, L. L., Continisio, G. I., De Lucia, N., Gigante, E., Guerriera, C., Maldonato, N. M., ... & Cantone, D. (2023). A scoping review on innovative methods for personality observation. *Frontiers in Psychology, 14*, 1112287.

Murray, S. (2020). *Introduction to contemporary print culture: Books as media*. Routledge.

Mussel, P., Gatzka, T., & Hewig, J. (2016). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment.*

Neale, B., Henwood, K., & Holland, J. (2012). Researching lives through time: An introduction to the Timescapes approach. *Qualitative Research, 12*(1), 4-15.

Ngom, F., & Castro, E. (2019). Beyond African orality: Digital preservation of Mandinka ʿAjamī archives of Casamance. *History Compass, 17*(8), e12584.

Nguyen, D. C., Cheng, P., Ding, M., Lopez-Perez, D., Pathirana, P. N., Li, J., & Poor, H. V. (2020). Enabling AI in future wireless networks: A data life cycle perspective. *IEEE Communications Surveys & Tutorials, 23*(1), 553-595.

Nkurunziza, J. D. (2022). *The origin and persistence of state fragility in Burundi. In State fragility* (pp. 101-140). Routledge.

Oakley, A. (2010). The social science of biographical life-writing: some methodological and ethical issues. *International Journal of Social Research Methodology, 13*(5), 425-439.

Parker, A. L. (1998). *Theoretical models of personality type in clinical practice: A critical examination of the Myers-Briggs Type Indicator*. University of Colorado at Boulder.

Pasachoff, N. (1996). *Marie Curie: And the science of radioactivity.* Oxford University Press.

Post, J. M. (2014). Personality profiling analysis. *The Oxford Handbook of Political Leadership*, 328.

Reis, H. T. (2013). Domains of experience: Investigating relationship processes from three perspectives. In *Theoretical frameworks for personal relationships* (pp. 87-110). Psychology Press.

Renders, J. W. (2013). The Limits of Representativeness. Biography, Life Writing and Microhistory. In *Theoretical Discussions of Biography* (pp. 195-210). Edwin Mellen Press.

Reyntjens, F. (2006). Briefing: Burundi: a peaceful transition after a decade of war? *African affairs, 105*(418), 117-135.

Rollyson, C. (2013). *Theoretical Discussions of Biography: Approaches from History, Microhistory, and Life Writing*.

Russell, A. (2015). Rebel and rule in Burundi, 1972. *The International Journal of African Historical Studies, 48*(1), 73-97.

Sabiraguha, A. E., Havyarimana, V., Niyongabo, P., Kamdjoug, J. R. K., Sindayigaya, I., & Niyonsaba, T. (2023). Digital in Higher Education in Burundi. *Open Journal of Social Sciences, 11(*11), 284-297.

Saidi, M. A., & Oladimeji, T. (2015). Hutu-Tutsi Conflict in Burundi: A Critical Exploration of Factors. *Available at SSRN 2689413.*

Smith, L. (2012). Biographical method. *J. Goodwin. Biographical research Volume one: Biographical research: starting points, debates and approaches*, 1-36.

Stanley, L. (2004). The epistolarium: On theorizing letters and correspondences. *Auto/biography, 12*(3), 201-235.
Su, F. (2020). Cross-national digital humanities research collaborations: structure, patterns and themes. *Journal of Documentation, 76*(6), 1295-1312.

Suthar, A. B., Khalifa, A., Yin, S., Wenz, K., Ma Fat, D., Mills, S. L., & Mrkic, S. (2019). Evaluation of approaches to strengthen civil registration and vital statistics systems: A systematic review and synthesis of policies in 25 countries. *PLoS Medicine, 16*(9), e1002929.

Tasse, D., Liu, Z., Sciuto, A., & Hong, J. (2017, May). State of the geotags: Motivations and recent changes. In *Proceedings of the International AAAI Conference on Web and social media* (Vol. 11, No. 1, pp. 250-259).

Thomas, B. P. (2008). *Abraham Lincoln: a biography*. SIU Press.

Tünsmeyer, V. (2022). Repatriation of Sacred Indigenous Cultural Heritage Under International Human Rights Law. In *Repatriation of Sacred Indigenous Cultural Heritage and the Law: Lessons from the United States and Canada* (pp. 49-141). Cham: Springer International Publishing.

Varshney, N., Chatterjee, A., Parmar, M., & Baral, C. (2023). Accelerating LLM Inference by Enabling Intermediate Layer Decoding. *arXiv preprint arXiv:2310.18581.*

Wadhwani, R. D., & Decker, S. (2017). Clio's toolkit: historical methods beyond theory building from cases. In *Routledge companion to qualitative research in organization studies.*

Wilkens, S., & Sanford, M. L. (2009). *Hidden worldviews: Eight cultural stories that shape our lives.* InterVarsity Press.
Wolf, J. H. (2018). Historical methods. *Journal of human lactation, 34*(2), 282-284.

Wu, S., Fei, H., Qu, L., Ji, W., & Chua, T. S. (2023). Next-gpt: Any-to-any multimodal llm. *arXiv preprint arXiv:2309.05519*.

Xu, W., Li, M., Liu, W., & Lin, G. (2023). Ritualizing the mundanity of holidays in usual environment. *Tourism Management Perspectives*, *47*, 101133.