Preserving Digital Footprints: Strategies for Safeguarding Ephemeral Online Data

Tolulope Balogun*

ARTICLE INFO

ABSTRACT

This paper discusses the issues related to ephemeral nature of online data and explores the strategies to safeguard digital heritage, particularly focusing on the nexus between ephemeral online data and web archiving. Document analysis was adopted to select peer-reviewed articles, reports, and publications from reputable academic databases and grey literature sources, covering the period from 2010 to 2024. The research reveals the ephemeral nature of online data, highlighting the challenges it poses for digital heritage preservation. Web archiving emerges as a crucial tool for preserving ephemeral online data, offering opportunities to capture and retain digital content for future generations. Recommendations are provided to address technical, legal, and ethical challenges facing web archiving, emphasizing the importance of interdisciplinary collaboration, increased awareness, and the development of clear digital preservation frameworks that incorporates ephemeral online data. The findings have practical implications for stakeholders involved in digital preservation, including archivists, researchers, policymakers, and technology developers.

1. Introduction

The internet has revolutionized the way people access, create, and share information. With the rapid growth of online content, there is an increased need to preserve digital heritage materials for future generations (Parent et al., 2021). This is especially crucial in the digital era where information dissemination through digital platforms has become a significant source of information for researchers, policymakers, and the public. However, the preservation of online content presents unique challenges. Unlike traditional forms of media, online content constantly changes, and the lifespan of digital content is often short-lived, with online data disappearing after a short period (Hendry & Stock, 2014; Duncan et al., 2016). This ephemeral nature of online data has led to the loss of significant amounts of digital history, including valuable information on social media, blog posts, online news articles, and online publications. It is important to preserve data for the sake of keeping accurate records. The capture and preservation of web content and social media accounts of protesters, govern-

ment officials, and institutions is important to build a complete documentary archive of a historical political event like the USA 2020 election (UNESCO, 2021). For example, the capture of @POTUS account tweets are important to ensure a complete archival record of the Donald Trump presidency (UNESCO, 2021).

Web archiving has emerged as a solution for preserving ephemeral data, information, and records (Hendry & Stock, 2014). It involves the collection, storage, and retrieval of web pages to create an archive of online content (Vlassenroot et al., 2019). The archived pages can then be accessed later, preserving the original context and content. Web archiving is an important tool for digital preservation, providing an opportunity to preserve ephemeral online data for future generations.

In the digital age, the ephemeral nature of online data presents significant challenges for the preservation of our digital heritage. The internet and social media platforms have become integral to our lives, shaping how we communicate, access information, and interact with the world around us. However, the dynamic and ever-changing nature of the web means that digital content is at risk of being lost or inaccessible over time. Despite the growing awareness of the ephemeral nature of the web and the critical role of web archiving in preserving digital history, there remains a gap in understanding the effectiveness and accessibility of web archiving initiatives, particularly in the context of developing countries. For example, there is scarce attention on research related to preservation of ephemeral online data and web archiving in Africa. Therefore, this paper discusses the ephemeral nature of the web in the context of potential loss to digital history and the role of Web archiving in preserving ephemeral online data. Ephemeral online content in the context of this study refers to temporary digital materials, often appearing briefly before disappearing or being altered. This study focuses more on social media platforms due to their heightened ephemerality and the potential for vast amounts of data to become outdated or lost. It covers social media platforms, dynamic websites, and short-lived digital publications. This paper also aims to stimulate research interest on the preservation of digital history available online, especially for researchers in developing countries. The main objectives are to:

(1) Discuss the ephemeral nature of the web.
(2) Discuss the nexus between ephemeral online data and Web archiving.
(3) Highlight ephemeral online content preservation initiatives.

2. Methodology

This is a qualitative study based on documentary analysis involving the analysis of relevant scholarly articles and online publications. Documentary analysis is a systematic method used by researchers to examine, review, and interpret written materials found in databases or the public or private domain (Mogalakwe, 2006; Dey, 2005). This method is particularly suitable for this study as it allows for an in-depth understanding of the issues related to ephemeral online data, digital preservation, and web archiving. The criteria for selecting documents or focusing on extracts reflect the specific issues the researcher seeks to address (Dey, 2005). Documents were selected based on their relevance

to the topics of ephemeral web data, digital preservation, and web archiving. This ensures a comprehensive and focused analysis of the literature. A systematic search strategy was employed using specific Boolean operators and search strings. Keywords such as "ephemeral web," "ephemeral online data," "digital preservation," "web archiving," and "web archiving initiatives" were used in academic databases like Scopus, Web of Science, and Google Scholar. Searches included terms like ("ephemeral web" OR "ephemeral online data") AND ("digital preservation" OR "web archiving"). These keywords were chosen for their direct relevance to the research questions. In total, 1,250 articles were retrieved from the initial search. The study utilized Scopus, Web of Science, and Google Scholar due to their extensive coverage of academic literature and their reliability as sources for scholarly research. Studies published between 2010 and 2024 were included to ensure that recent developments and research findings were incorporated into the study. The initial search period was from 2014, but it was expanded to 2010 to capture useful publications and research carried out in the field. From the 1,250 articles retrieved, a detailed screening process was conducted to evaluate the relevance and quality of each article. After the initial screening, 300 articles were deemed highly relevant based on their abstracts and keywords. Further full-text review narrowed this down to 70 articles that were directly related to the research objectives. In addition to peer-reviewed articles and reports from recognized institutions, other relevant online resources were manually selected regardless of their year of publication and publication medium. This manual selection aimed to include comprehensive and valuable materials that might have been excluded or not be indexed in the academic databases utilized. To preserve academic integrity and scholarly ethics, proper citation and acknowledgment of sources were maintained throughout the review process, along with transparent reporting of methodology and conclusions.

3. Discussion

3.1 Ephemeral nature of the web

The internet has played a key role in the communication infrastructure of most societies since inception, prompting researchers to increasingly study it (Brügger, 2016). The importance of websites, as sources of data for conducting research, and as a faction of a nation's cultural heritage, has made it very necessary to ensure the preservation of portions of websites, to ensure future access to information (Bainotti et al., 2020; Parent et al. 2021). Several studies have affirmed that the Internet and web contents are extremely ephemeral (John et al., 2024; Slania, 2013; Zeitlyn et al., 2015; Duncan & Blumenthal, 2016; Donovan & Haberle, 2018). This includes loss and decay of cultural heritage like art ephemera (Slania, 2013; Duncan & Blumenthal, 2016), important academic materials (John et al., 2024; Donovan & Haberle, 2018) and social media (Zeitlyn et al. 2015; Pehlivan et al., 2021). For instance, Zeitlyn et al. (2015) note that 11% of social content disappears after a year, increasing to 30% after two years. The preservation of social media archives presents a significant challenge, necessitating a long-term strategy like web archives but complicated due to the dynamic nature of social media content (Pehlivan et al., 2021). With the growing recognition

of social media as a vital resource for researchers and archivists, particularly evident in the aftermath of events like the 2011 Egyptian Revolution (SalahEldeen & Nelson, 2012), there is both enthusiasm for its potential and concern regarding the adequacy of preservation methods (Pehlivan et al. 2021). Despite these challenges, social media data has firmly entrenched itself as a crucial source across various domains of research (Pehlivan, 2021). Efforts have been made by institutions to incorporate social media content, predominantly from platforms such as Twitter, into their collections are underway, although much of this material remains inaccessible to the public (Pehlivan, 2021). The comprehensive archival coverage of social media platforms, as observed in previous research, highlights the evolving digital preservation landscape (Helmond & van der Vlist, 2019).

Furthermore, the impact of the web on our society cannot be underestimated (Brügger, 2016), but the ephemeral nature of the web requires that proactive steps should be taken to enable the recreation of the web experience for future analysis (Cannelli & Musso, 2022). Extending beyond traditional webpages, web archives serve as invaluable resources for dissecting the propagation of misinformation across the web and social media networks (Kumar & Shah, 2018). In recent times, many websites are considered culturally valuable, and the dissemination of important academic works has also been done on the web, even though there is an increase in the number of inactive URLs (Donovan & Haberle, 2018; John et al, 2024). On the other hand, studies in recent years have been able to prove that the problems associated with the vanishing of URLs can be solved through different web archiving tools (John et al, 2024; Loan et al., 2023; Singh & Devi, 2024).

With the ephemeral nature of online data, digital preservation in the form of web archiving is essential to make them available in the future (Cannelli & Musso, 2022). The uniqueness of certain information available, coupled with the web's ephemeral nature, has made the long-term preservation of ephemeral online data an important issue (Vlassenroot et al., 2021; Cannelli & Musso, 2022). In recognition of the importance of digital preservation, UNESCO (2003) reports that the loss of digital information can impoverish the heritage of all nations. Therefore, it is very important to preserve these data for historic research (Habibzadeh, 2013; Hendry & Stock, 2014). Using web archiving to capture the past for posterity is as important as preserving other aspects of our cultural heritage, regardless of whatever form it is presented in (Hendry & Stock, 2014). Permanent archiving of URLs has been suggested as one of the ways decays in web content can be mitigated (Habibzadeh, 2013).

3.2 Ephemeral Online Data and Web archiving: The Nexus

The ephemeral nature of online data poses significant challenges for web archiving (Davis, 2014; Duncan et al., 2016). In fact, web ephemera jeopardizes human knowledge (Major, 2021). Ephemeral online data refers to digital content that is volatile, rapidly changing, and has a short lifespan. This includes social media posts, news articles, and other digital content that is often not preserved through traditional archival methods. The rapid evolution of technology and the internet presents a challenge for preserving ephemeral online data (Brügger et al., 2017). As technology evolves, digital content is often presented in new and innovative ways, making it challenging to capture and preserve the content accurately.

Web archiving has become an essential tool for preserving digital history (Brügger et al., 2017). The concept of web archiving originated in the mid-1990s, with the first web archiving project undertaken by the Internet Archive in 1996 (Grotke, 2017; Gratzinger, 2021). Since then, web archiving has become an established practice, with numerous organizations around the world engaged in preserving online content (Columbia University Archives, 2019; Donovan & Haberle, 2018). Web archiving is the collection of websites and its contents to preserve them for future use. It ensures that the contents on these websites are not erased over time. Web archiving enables organizations to access some legacy information that they may wish to take off their active websites over time, but which they consider relevant. This will generally contribute to the positive image of an organization's effectiveness in managing information (The National Archive, 2011). Brügger (2011: 25) also explains web archiving as "any form of deliberate and purposive preserving of web material". The web archiving technology "enables the capture, preservation, and reproduction of valuable content from the live web in an archival setting, so that it can be independently managed and preserved for future generations" (Pennock, 2013: 1). Szydlowski (2010) and Hendry & Stock (2014) give practical examples of ephemeral online data and the need for Web archiving.

It is important to note that archived websites differ from other archived media (Brügger, 2012). According to Brügger (2010: 7), the archiving process "creates a unique version rather than a copy, and it is a version of an original, which we can never expect to find in the form it usually took on the web; we can neither find an original among the different versions nor reconstruct an original based on different versions". Like traditional archives, web archives are collected and placed under the care of archivists', known as 'web archivists' (The National Archives, 2011). Therefore, the archived website is considered a unique version and not just a copy of what was once on the online web (Brügger, 2012). Access to the archived websites' contents is made available for use by researchers, organizations, government, and the public (The National Archives, 2011).

However, Web archiving faces several challenges, including technical, legal, and ethical issues (Adoghe et al., 2013; Bingham et al., 2021). One of the primary technical challenges is ensuring the authenticity and integrity of archived web pages (Balogun & Kalusopa, 2021). Web pages are complex entities composed of various media types, such as text, images, and videos, which can be challenging to capture accurately. Additionally, web pages are constantly changing, making it challenging to ensure that archived pages represent the original content accurately. The legal issues relate to copyright, privacy, and data protection laws (Adoghe et al., 2013; Davis, 2014). For instance, archiving copyrighted content without permission is illegal and can lead to legal action. Copyright issues are a major concern in web archiving, alongside technical challenges and questions of authenticity (Kim, 2007). The implementation of legal deposit for web materials creates a complex relationship between copyright and digital heritage preservation (Cadavid, 2014). These legal challenges extend to ethical considerations in web archiving practices, such as determining what heritage is captured or neglected (Bingham & Byrne, 2021). To address these issues, there is a need to modify copyright legislation to foster research activities and support digital humanities (Cadavid et al., 2014). Additionally, memory institutions must redefine their archival strategies to navigate the challenges of contemporary collecting in the digital age (Bingham & Byrne, 2021). Privacy and data protection laws also require archivists to consider how personal information on web pages is stored, accessed, and used (Vavra,

2018). Legal frameworks differ between countries, affecting how national libraries approach web archiving programs (Glanville, 2010). To address these issues, many web archives limit access to their holdings or provide opt-out mechanisms, which can reduce their usefulness (Rauber et al., 2008). The ethical issues relate to the impact of web archiving on individual and collective rights, such as privacy concerns (Graham, 2017; Mackinnon, 2022). Researchers are exploring automated methods to identify personal information in web pages as a potential solution (Rauber et al., 2008).

Despite these challenges, web archiving remains crucial for preserving digital cultural heritage (Gomes & Costa, 2014; Hockx-Yu, 2011). Closer collaboration with web science researchers and utilization of technologies developed for the live web could advance web archiving efforts (Hockx-Yu, 2011). Web archives provide an opportunity to preserve online content, which is often ephemeral and easily lost. They also enable researchers and scholars to access and analyze historical web content, providing insights into the evolution of digital culture and society. Looking towards the future, web archiving will continue to evolve as technology advances and new challenges arise.

3.3 Ephemeral online content preservation initiatives

The preservation of ephemeral online data including social media content has emerged as a crucial endeavor for understanding societal and cultural transformations (UNESCO, 2020). Various institutions have established specialized initiatives to archive ephemeral online data, facilitating scholarly research and historical documentation. The Library of Congress Web Archives initiative was dedicated to archiving social media content from platforms like Twitter, Facebook, and YouTube (Fondren & McCune, 2018). This initiative enabled researchers to explore different topics like political campaigns to cultural phenomena, thus facilitating inquiries into the societal impact of social media over time. The partnership between the Library of Congress and Twitter further exemplifies the ambition to preserve the entirety of public tweets since the platform's inception, highlighting the significance of capturing ephemeral social media conversations for scholarly and historical analysis (Bruns, 2018; Library of Congress, 2017).

Complementing this effort is Archive-It, a web archiving service provided by the Internet Archive that helps organizations curate their web archives specific to their specific missions and collections. The Internet Archive's Wayback Machine stands out as one of the largest web archives globally, housing billions of archived web pages since inception of the Internet (Lutkevich, 2023) This repository includes diverse digital content like news articles, blog posts, and social media updates, serving as a valuable resource for understanding the evolution of the internet. Daily updates from photographer Humans of New York's Instagram feed, fiction writer Neil Gaiman's Tumblr blog, and Donald Trump's Facebook page have all been archived by the Internet Archive since 2016 (Major, 2021). The British Library's UK Web Archive also focuses on preserving the UK's digital heritage including websites, blogs, and social media content reflective of British culture, history, and society (Vlassenroot, 2021). While the British Library started preserving social media in 2010, some Twitter, Facebook, and YouTube content had previously been recorded before then (Espley et al., 2014; Vlassenroot, 2021). For example, its collection includes archives of Twitter accounts from 2008 (Espley et al., 2014).

In preserving political history, initiatives like the Trump Archive (Internet Archive, n.d.; National

Archives, n.d.) and the Obama White House Social Media Archive (CivicPlus, 2024; National Archives, n.d.) provide extensive records of presidential communications and social media engagements, shedding light on the intersection of digital media and governance. For instance, the Trump Archive, launched in 2017 collects non-commercial collection of TV news shows related to Donald Trump and preserves historical records for posterity (Internet Archive, n.d.), including Trump's Twitter archive (National Archives, n.d.). Meanwhile, national initiatives like the PANDORA Archive in Australia have also tried to safeguard online publications and websites of cultural significance, mitigating the risk of digital content disappearance due to technological changes. The PANDORA Archive content is part of the larger Australian Web Archive and can be searched in Trove (National Library of Australia, n.d.). Academic institutions like Stanford Libraries have also established web archiving programs dedicated to capturing social media content related to academic research, cultural events, and political movements. These initiatives contribute to scholarly discourse by documenting online conversations and debates across various social media platforms, providing insights into contemporary issues and trends.

In addition to established efforts to archive ephemeral online data, several innovative tools and collaborative initiatives contribute significantly to the preservation and analysis of ephemeral online data. For instance, Rhizome's Conifer Webrecorder is a tool for crawling web content, including social media posts and interactive websites (Kreymer, 2020; Digital Preservation Coalition, 2018a). This tool allows users to capture browsing sessions in real-time and preserve the full features and functionality of web pages before they change or vanish. Similarly, Documenting the Now (DocNow) offers researchers a platform to collect, preserve, and analyze social media content related to specific events or topics of interest (Summers, 2018; Jules et al., 2018; National Humanities Alliance, 2024). Doc Now's focus on social media data enables scholars to track the evolution of online conversations across different platforms, providing insights into societal dynamics and cultural shifts (Jules et al., 2018). Collaborative initiatives like DocNow and the RESAW Network exemplify efforts to advance the study of archived web archives (Jules et al., 2018; Gluhovic, 2023). DocNow, a collaborative project between multiple organizations, empowering researchers with tools and resources to capture and analyze social media conversations, and facilities interdisciplinary inquiries into ephemeral online data (Jules et al., 2018). The RESAW Network also promotes collaboration and knowledge sharing between researchers, archivists, and technologists in exploring the use of web archives across various disciplines (Brügger, 2023). The RESAW community, established in 2012, aims to facilitate a collaborative European research infrastructure for archived web materials (Gluhovic, 2023). By supporting projects that leverage web archives for research, the RESAW Network enhances our understanding of the challenges and opportunities related to the preservation of ephemeral online data.

Furthermore, initiatives like Perma.cc, developed by the Harvard Library Innovation Lab in collaboration with various organizations, address the challenge of link rot by enabling users to create permanent links to web pages (Perma.cc, n.d.; MIT Libraries, n.d.). This service ensures the accessibility of online sources, particularly for scholarly research and legal proceedings, even if the original URL changes or the content is removed. While Perma.cc attempts to prevent link rot by preserving web page content, it is not considered as a web archiving tool (MIT Libraries, n.d.)

Despite efforts from several organizations to preserve ephemeral online data, social media content

especially remains challenging due to its interactive nature and reliance on embedded media, complicating traditional archiving methods (Digital Preservation Coalition, 2018b; Thomson, 2016; Rocha, 2023). This is evident in the challenges that have led to some initiatives being unsuccessful and thus terminated. For instance, the Library of Congress announced the termination of its Twitter archiving initiative in 2017 (Library of Congress, 2017), sparking concerns among contemporary analysts and future historians (Bruns, 2018). Bruns (2018) laments the Library's decision, attributing it to the conservative nature of the Library and a limited understanding of emerging technologies. Emphasizing the importance of archiving Twitter, Bruns (2018) advocates for Twitter's collaboration with entities like the Internet Archive or other national libraries for successful preservation efforts. Although the Library of Congress now selectively acquires tweets, it maintains its collection as an oral history of the social media era (Bruns, 2018; Wamsley, 2017). The British Library also faced challenges in preserving ephemeral online data from social media platforms (Byrne, 2017). Previous attempts to archive platforms like Instagram and Flickr were frustrated by technical challenges and access restrictions (Byrne, 2017). Byrne (2017) emphasizes the difficulty of archiving social media content due to its unique presentation and controlled API access. However, the issue of preserving online data remains important, particularly in preventing the loss of significant social media content to ensure preservation of a complete and accurate ephemeral record on the web (Rocha, 2023). Preserving digital data from social media platforms requires innovative archiving solutions, and institutions like Library of Congress and British Library need to collaboration with technology partners to better understanding digital landscapes.

4. Conclusion

Ephemeral online data preservation is a complex task due to the rapid evolution of digital platforms and the complexities of capturing context and dynamics. Traditional archival methods are not sufficient in capturing the volume and velocity of data generated by platforms like Twitter and Instagram, making identification and preservation difficult. Several studies have affirmed the ephemeral nature of online data, revealing the uncertainty of digital preservation due to rapid technological evolution, and shifting online landscapes (e.g. John et al, 2024; and Zeitlyn et al., 2015). In response to the ephemeral nature of online data, efforts have been made to develop and implement web archiving initiatives. One of such initiative was pioneered by the Internet Archive in the mid-1990s and has paved the way for comprehensive web archiving programs aimed at capturing and preserving a wide array of online materials, from individual web pages to entire websites and social media platforms. To address these challenges, interdisciplinary collaboration, robust infrastructure, and ongoing research are crucial. Initiatives like the Library of Congress Web Archives, the Internet Archive's Wayback Machine, the British Library's UK Web Archive, the Trump Archive, PANDORA Archive, the Obama White House Social Media Archive, and Stanford Libraries' Web Archiving Program have attempted to archive social media data for scholarly research and historical documentation.

Despite the efforts and progress made in the field of web archiving, there are still significant challenges that need to be tackled. Technical challenges, such as ensuring the authenticity and

integrity of archived web pages, remain a pressing concern for archivists and researchers. Moreover, legal, and ethical considerations, including copyright and privacy laws, raise complex questions about the rights and responsibilities of web archivists and the broader implications of web archiving for individual freedoms and democratic values. However, the significance of web archiving for preserving online data cannot be overstated. Web archives serve as invaluable repositories of digital heritage, providing researchers and scholars with unparalleled insights into the evolution of digital culture and society. By preserving online content that is at risk of disappearing or being altered, web archiving ensures that future generations will have access to a rich tapestry of digital history, enabling them to explore and analyze the complexities of the digital age. Advances in technology, coupled with ongoing efforts to navigate legal and ethical considerations, will shape the future of web archiving and its role in preserving our digital heritage for generations to come.

5. Recommendations

Based on the findings of this study, there is a need for continued research and development in the field of web archiving to address the numerous challenges posed by the ephemeral online data. While significant progress has been made in using web archiving techniques to preserve ephemeral online data, ongoing research is needed to address emerging challenges and opportunities in this field. Future research must continue to explore innovative methods for capturing and preserving ephemeral online data, while also addressing the ethical and legal challenges associated with web archiving. The ethical implications of archiving online content like social media, including issues of consent and privacy, require careful consideration and ongoing discussion among researchers and policymakers.

There is also a need for increased awareness and training on web archiving techniques for preserving digital history. Archivists and other stakeholders involved in digital preservation should be trained in the use of web archiving tools and techniques to ensure the effective preservation of our digital heritage. Archivists and other stakeholders (especially in Africa) should establish clear policies and procedures for the continuous monitoring and maintenance of archived web content. Furthermore, there is a need for the development of clear legal and ethical frameworks for the preservation of online data. Archivists and other stakeholders should engage with policymakers and legal experts to establish clear guidelines and standards for the preservation of online data.

In addition, there is a need for increased collaboration and partnerships between different stakeholders involved in digital preservation. Archivists, researchers, policymakers, and other stakeholders should collaborate and share knowledge and expertise to ensure the effective preservation of ephemeral data, information, and records. Collaboration between institutions, organizations, and stakeholders is important to the success of web archiving initiatives. For society, the preservation of digital heritage is essential for maintaining a collective memory and ensuring that diverse voices and experiences are recorded for posterity. Sustainable funding and resources are important to support the long-term sustainability of web archiving initiatives. In addition, governments, funding agencies or organizations, and private donors need to invest in the infrastructure, personnel, and technologies

needed to ensure the continued preservation of digital content, especially in developing countries.

Finally, there is a need for a framework for preservation of ephemeral online information and data that serves as a guideline for the preservation of digital heritage materials online in developing countries. The framework should be flexible enough to accommodate the diverse range of digital content available online, while also providing clear guidelines for archivists and other stakeholders. In developing a framework for preserving digital history, it is essential to engage with a range of stakeholders, including archivists, scholars, cultural institutions, and the public. This collaborative approach can help to ensure that the framework is comprehensive, effective, and reflective of the needs and concerns of all stakeholders. Practitioners in the field of web archiving must focus on developing more sophisticated tools and techniques, fostering interdisciplinary collaborations, and advocating for supportive legal frameworks that balance preservation needs with individual rights. Working collaboratively with stakeholders would help to develop a framework that addresses the unique challenges of preserving ephemeral online data, while also providing clear and comprehensive guidance for digital preservation experts, and other stakeholders

Declaration of Interest Statement

The author declares that there are no conflicts of interest regarding the publication of this article, and no financial assistance or funding was received for the research, authorship, or publication of this article.

References

Anthony, A., Onasoga, K., Ike, D., & Ajayi, O. (2013). Web archiving: Techniques, challenges, and solutions. *International Journal of Management & Information Technology, 5*(3), 598-603.

Bainotti, L., Caliandro, A., & Gandini, A. (2020). From archive cultures to ephemeral content, and back: Studying Instagram Stories with digital methods. *New Media & Society, 23*(12), 3656-3676.

Balogun, T., & Kalusopa, T. (2021). Web archiving of indigenous knowledge systems in South Africa. *Information Development, 38*(4), 658-671.

Barone, F., Zeitlyn, D., & Mayer-Schönberger, V. (2015). Learning from failure: The case of the disappearing web site. *First Monday, 20*(5).

Bingham, N. J., & Byrne, H. (2021, January). Archival strategies for contemporary collecting in a world of big data: Challenges and opportunities with curating the UK web archive. *Big Data & Society, 8*(1), 205395172199040.

Brügger, N. (2012). Web history and the web as a historical source. *Zeithistorische Forschungen, 9*(2), 316-25.

Brügger, N. (2016). Introduction: The Web's first 25 years. *New Media & Society, 18*(7), 1059-1065.

Brügger, N. (2023). *RESAW*. Retrieved from https://cc.au.dk/en/resaw

Brügger, Niels, Locatelli, Elisabetta, Weber, Matthew, & Nanni, Federico (2017). *Web 25: histories from the first 25 years of the World Wide Web*. In: Researchers, practitioners and their use of the archived web, 14-16 June 2017, School of Advanced Study, University of London.

Bruns, A. (2018, April). *The Library of Congress Twitter Archive: A Failure of historic proportions.* Medium. Retrieved from
https://medium.com/dmrc-at-large/the-library-of-congress-twitter-archive-a-failure-of-historic-proportions-6dc1c3bc9e2c

Byrne, H. (2017, April). *The Challenges of Web archiving social media*. Retrieved from
https://blogs.bl.uk/webarchive/2017/04/the-challenges-of-web-archiving-social-media.html

Cadavid, J. A. P. (2014). Copyright challenges of legal deposit and web archiving in the national library of Singapore. *Alexandria, 25*(1-2), 1-19.

Cadavid, J. A. P., Basha, J. S., & Kaleeswaran, G. (2014). Legal and technical difficulties of web archival in Singapore. *Rev. Prop. Inmaterial, 18*, 35.

Cannelli, B., & Musso, M. (2022). Social media as part of personal digital archives: exploring users' practices and service providers' policies regarding the preservation of digital memories. *Archival Science, 22*(2), 259-283.

CivicPlus. (2024, May 7). *The First White House Social Media Archive*. Retrieved from
https://www.civicplus.com/case-studies/sma/white-house-social-media-archive/

Columbia University Archives. (2019, April). *Web archives*. Retrieved from
https://library.columbia.edu/libraries/cuarchives/resources/webarchives.html

Davis, C. (2015, March). *Archiving the Web: A case study from the University of Victoria*. The Code4Lib Journal.

Dey, I. (2005). *Qualitative data analysis*. London: Routledge, Taylor and Francis Group.

Digital Preservation Coalition. (DPC). (2018b, September). *Preserving social media: Digital Preservation Topical Note 8*. DPC Technology Watch Publications. Retrieved from
https://www.dpconline.org/docs/dpc-technology-watch-publications/topical-notes-series/1869-dp-note-8-preserving-social-media/file

Digital Preservation Coalition. (2018a, December). *7. Web & social media archiving: Rhizome's Webrecorder*. [Video]. https://www.youtube.com/watch?v=4YbmUzoqqG0

Donovan, L., & Haberle, M. (2018). Web archiving for academic institutions. *Digital Initiatives Symposium, 4*. Retrieved from https://digital.sandiego.edu/symposium/2018/2018/4

Duncan, S., & Blumenthal, K. R. (2016). A collaborative model for web archiving ephemeral art resources at the New York Art Resources Consortium (NYARC). *Art Libraries Journal, 41*(2), 116-126.

Espley, S., Carpentier, F., Pop, R., & Medjkoune, L. (2014). Collect, preserve, access: Applying the governing principles of the National Archives UK Government Web Archive to social media content. *Alexandria, 25*(1-2), 31-50.

Fondren, E., & Menard McCune, M. (2018). Archiving and preserving social media at the Library of Congress: institutional and cultural challenges to build a Twitter archive. Preservation, *Digital Technology & Culture, 47*(2), 33-44.

Glanville, L. (2010). Web archiving: ethical and legal issues affecting programmes in Australia

and the Netherlands. *The Australian Library Journal, 59*(3), 128-134.

Gluhovic, R. (2023, April). *About: About RESAW*. Retrieved from https://cc.au.dk/en/resaw/about

Gomes, D., & Costa, M. (2014). The importance of Web archives for humanities. *International Journal of Humanities and Arts Computing, 8*(1), 106-123.

Graham, P. M. (2017, October). Guest editorial: Reflections on the ethics of Web archiving. *Journal of Archival Organization, 14*(3-4), 103-110.

Gratzinger, O. (2021). *The Internet Archive: Founded by Brewster Kahle*. Retrieved from https://archive.org/about

Grotke, A. (2017). Getting started in Web archiving. Paper presented at *IFLA WLIC 2017 - Wrocław, Poland - Libraries. Solidarity. Society*. In Session 186 - National Libraries with Information Technology. Retrieved from https://library.ifla.org/id/eprint/1637/

Habibzadeh, P. (2013). Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs small journals. *Applied Clinical Informatics, 04*(04), 455-464.

Helmond, A. (2019) A Historiography of the hyperlink: Periodizing the web through the Changing Role of the Hyperlink. In: Brügger, N, Milligan, I (eds) *The SAGE Handbook of Web History*, 153-167. London: SAGE.

Hendry, R., & Stock, G. (2014). Forget me net, not: Inside the struggle to preserve the world's data. *Newsweek Global, 168*(2), 1-6.

Hockx-Yu, H. (2011, June). The past issue of the web. In *Proceedings of the 3rd International Web Science Conference* (pp. 1-8).

Internet Archive, (n.d.). *Trump Archive*. Retrieved from https://archive.org/details/trumparchive?&sort=-publicdate&page=

John, H. C., Simisaye, A. O., & Iseyemi, T. J. (2024). Missing and recovery of URLs using Internet Archive: A case study on African Journal of Library, Archives and Information Science (AJLAIS). *International Journal of Information Science and Management (IJISM), 22*(2), 193-210.

Jules, B., Summers, E., & Mitchell, V. (2018). Ethical considerations for archiving social media content generated by contemporary social movements: Challenges, opportunities, and recommendations. *Documenting The Now White Paper*. Retrieved from https://www.docnow.io/docs/docnow-whitepaper-2018.pdf

Kreymer, I. (2020, June). A New Phase for Webrecorder Project, Conifer and ReplayWeb.page. *Webrecorder*. Retrieved from https://webrecorder.net/2020/06/11/webrecorder-conifer-and-replayweb-page.html

Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559.*

Library of Congress. (2017). Update on the Twitter Archive at the Library of Congress. Retrieved from https://blogs.loc.gov/loc/files/2017/12/2017dec_twitter_white-paper.pdf

Loan, F. A., Khan, A. M., Andrabi, S. A. A., Sozia, S. R., & Parray, U. Y. (2023, July). *Giving life to dead: role of WayBack Machine in recovery of dead URLs*. Data Technologies and Applications. Retrieved from https://doi.org/10.1108/dta-06-2022-0242

Lutkevich, B. (2023, August). *Wayback Machine*. WhatIs. Retrieved from

https://www.techtarget.com/whatis/definition/Wayback-Machine

Mackinnon, K. (2022). Critical care for the early web: ethical digital methods for archived youth data. *Journal of Information, Communication and Ethics in Society, 20*(3), 349-361.

Major, D. (2021). The Problem of Web Ephemera. In *The Past Web: Exploring Web archives* (pp. 5-10). Cham: Springer International Publishing.

MIT Libraries, (n.d.). *LibGuides: Citation management and writing tools: Perma.cc.* Retrieved from https://libguides.mit.edu/cite-write/perma-cc

Mogalakwe, M. (2006). Research Report. The Use of Documentary Research Methods in Social Research. *African Sociological Review, 10*(1), 221-230.

National Archives, (n.d.). *Archived Social Media: Donald J. Trump Presidential Library*. Retrieved from https://www.trumplibrary.gov/research/archived-social-media

National Humanities Alliance. (2024). Documenting the Now. *Humanities for All*. Retrieved from https://humanitiesforall.org/projects/documenting-the-now

National Library of Australia, (n.d.). *PANDORA Web archive*. Retrieved from https://pandora.nla.gov.au/

Parent, I., Seles, A., Storti, D., Banda, F., Blin, F., McKenna, G., Lee, I., Murdock, S. J., Chee, J., Hagedorn-Saupe, M., Knight, S., & Roberts, W. (2021). *The UNESCO/PERSIST Guidelines for the selection of digital heritage for long-term preservation*. Retrieved from https://repository.ifla.org/handle/123456789/1863

Pehlivan, Z. (2021). Linking Twitter archives with television archives. In *The Past Web: Exploring Web Archives* (pp. 127-139). Cham: Springer International Publishing.

Pehlivan, Z., Thièvre, J., & Drugeon, T. (2021). Archiving social media: the case of Twitter. In *The Past Web: Exploring Web Archives* (pp. 43-56). Cham: Springer International Publishing.

Pennock, M. (2013, March 6). *Web-Archiving*. Retrieved from https://doi.org/10.7207/twr13-01

Perma.cc, (n.d.). *Websites change. Perma Links don't.* (n.d.). Retrieved from https://perma.cc/

Rauber, A., Kaiser, M., & Wachter, B. (2008, September). Ethical issues in web archive creation and usage-towards a research agenda. In *8th International Web Archiving Workshop (IWAW08)*.

Rocha, D. P. (2023, December). *Web Archiving: preserving the ephemeral. | Medium*. Medium. Retrieved from https://medium.com/@danielpetri1/web-archiving-b09cfb47e440

SalahEldeen, H. M., & Nelson, M. L. (2012). Losing my revolution: How many resources shared on social media have been lost?. *In International Conference on Theory and Practice of Digital Libraries* (pp. 125-137). Springer, Berlin, Heidelberg.

Singh, T. G., & Devi, K. S. (2024). Web decay analysis and digital archiving of websites of technical institutions: a view from Wayback Machine. *College Libraries, 39*(I), 1-10.

Slania, H. (2013). Online art ephemera: Web archiving at the National Museum of Women in the Arts. *Art Documentation: Journal of the Art Libraries Society of North America, 32*(1), 112-126.

Summers, E. (2018, February). *Introducing Documenting the Now*. Medium. Retrieved from https://news.docnow.io/introducing-documenting-the-now-416874c07e0#.6wp34iv6a

Szydlowski, N. (2010). Archiving the web: It's going to have to be a group effort. *The Serials Librarian, 59*(1), 35-39.

The National Archives. (2011). *Basic Web archiving Guidance*. Retrieved from https://nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf

Thomson, S. D. (2016). Preserving social media. *DPC Technology Watch Report 16-01 February 2016*. Retrieved from https://www.dpconline.org/docs/technology-watch-reports/1486-twr16-01/file

Trump Twitter Archive V2, (n.d.). Retrieved from https://www.thetrumparchive.com

UNESCO (2003). *Charter on the preservation of digital heritage*. United Nations Educational, Scientific and Cultural Organization (UNESCO). Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000179529

UNESCO (2020, October). Cutting edge: Protecting and preserving cultural diversity in the digital era. Retrieved from https://www.unesco.org/en/articles/cutting-edge-protecting-and-preserving-cultural-diversity-digital-era

UNESCO. (2021). Documentary heritage at risk: Policy gaps in digital preservation. Outcomes of UNESCO Policy Dialogue Prepared by the Preservation Sub-Committee of the International Advisory Committee of the UNESCO Memory of the World Programmed. Retrieved from https://webarchive.unesco.org/web/20211224075907/https://en.unesco.org/sites/default/files/documentary_heritage_at_risk_policy_gaps_in_digital_preservation_en.pdf

Vavra, A. N. (2018). The right to Be forgotten: An archival perspective. *The American Archivist, 81*(1), 100-111.

Vlassenroot, E., Chambers, S., Di Pretoro, E., Geeraert, F., Haesendonck, G., Michel, A., & Mechant, P. (2019, March 8). Web archives as a data resource for digital scholars. *International Journal of Digital Humanities, 1*(1), 85-111.

Vlassenroot, E., Chambers, S., Lieber, S., Michel, A., Geeraert, F., Pranger, J., Birkholz, J., & Mechant, P. (2021). Web-archiving and social media: an exploratory analysis. *International Journal of Digital Humanities, 2*(1-3), 107-128.

Wamsley, L. (2017, December 26). *Library of Congress Will No Longer Archive Every Tweet*. NPR - The Two-Way. Retrieved from https://www.npr.org/sections/thetwo-way/2017/12/26/573609499/library-of-congress-will-no-longer-archive-every-tweet

You-Seung, K. I. M. (2007). A study of legal issues for web archiving. *Journal of the Korean Society for Library and Information Science, 7*(3), 5-24.

[About the author]

Tolulope Balogun is a Postdoctoral researcher at the University of South Africa with research interest in digitization, digital preservation, Artificial Intelligence, records and archives.