404 PAGE NOT FOUND: THE CRITICAL IMPORTANCE OF PERSONAL DIGITAL ARCHIVING IN JOURNALISM

Dan Valeriu VOINEA ⬤ , Ph.D., University of Craiova, Romania

Abstract

This paper examines the critical issue of digital content preservation in modern journalism, focusing on the imperative of personal digital archiving practices among journalists. As news organizations transition from print to digital platforms, the ephemeral nature of online content poses significant challenges to the preservation of journalistic work and the maintenance of historical records. Through a comprehensive analysis of current preservation practices in newsrooms and an exploration of theoretical frameworks including digital preservation theory, information lifecycle management, and professional identity development, this study reveals widespread inadequacies in institutional archiving efforts. The research employs the SPOT (Simple Property-Oriented Threat) Model to evaluate preservation practices, highlighting the vulnerabilities of digital news content. Findings indicate a alarming trend of content volatility, with only a fraction of surveyed newsrooms fully preserving their news content. This study argues that personal digital archiving is not merely a contingency plan but a critical professional practice for modern journalists, essential for maintaining comprehensive portfolios, ensuring the longevity of their contributions to the public record, and safeguarding against the volatility of digital platforms. The thesis explores the technical, ethical, and practical dimensions of personal digital archiving in journalism, proposing recommendations for improving archiving practices. It concludes by emphasizing the need for a paradigm shift in how journalists approach the preservation of their work in the digital age, calling for further research into effective models for integrating preservation practices into digital news production workflows and the potential of emerging technologies in enhancing digital news preservation.

Keywords: *Digital journalism, content preservation, personal digital archiving, information lifecycle management, professional identity, digital obsolescence, SPOT Model, metadata management, news archives, digital preservation theory*

Introduction

In the digital age, journalism faces an unprecedented challenge: the ephemeral nature of online content. As news organizations transition from print to digital platforms, a growing concern emerges about the preservation of journalistic work. The phrase "Page Not Found" and the error code 404 (*404 Not Found - HTTP | MDN*, 2024) have become an ominous specter haunting the corridors of digital newsrooms, symbolizing not just technical glitches but the potential loss of vital historical records and journalists' portfolios.

Recent years have witnessed a disturbing trend of digital content disappearing from the internet. High-profile cases, such as the shutdowns of MTV News, Deadspin and The Awl, (Tameez, 2024) have brought this issue into sharp focus. These incidents are not isolated; they represent a broader pattern of content volatility in the digital journalism landscape. The implications are far-reaching, affecting not only the accessibility of information for the public but also the professional lives of journalists who rely on their published work as testament to their skills and experience.

The current state of digital archiving in newsrooms presents a complex and often inadequate picture. A 2021 study by the Reynolds Journalism Institute found that only 7 out of 24 surveyed newsrooms were fully preserving their news content. (McCain et al., 2021) This statistic underscores a significant gap between the rapid production of digital content and the infrastructure needed to preserve it. Many newsrooms, particularly smaller digital-only outlets, lack the resources or protocols to implement comprehensive archiving systems. Even larger legacy news organizations, despite having more resources, face challenges in keeping pace with evolving technology and the changing nature of digital journalism.

In this precarious environment, the importance of personal digital archiving for journalists cannot be overstated. As institutional safeguards prove insufficient, journalists must take proactive measures to

preserve their work. This thesis argues that personal digital archiving is not merely a contingency plan but a critical professional practice for modern journalists. It is essential for maintaining a comprehensive portfolio, ensuring the longevity of their contributions to the public record, and safeguarding against the volatility of digital platforms.

This paper will explore the multifaceted nature of this issue, examining the technical, ethical, and practical dimensions of personal digital archiving in journalism. By analyzing current practices, challenges, and potential solutions, we aim to underscore the urgent need for a paradigm shift in how journalists approach the preservation of their work in the digital age.

Literature review

Personal digital archiving in journalism is essential for preserving the historical record and ensuring information accessibility for future generations. The advent of digital technologies has transformed how journalists create, disseminate, and store information, highlighting the importance of effective digital preservation strategies (Gaur & Tripathi, 2012). As journalism increasingly relies on digital platforms for content creation and distribution, the significance of preserving digital archives has become crucial to uphold the integrity and authenticity of journalistic content (Reading, 2014). Digital preservation and metadata are vital components of personal digital archiving, guaranteeing the long-term accessibility and usability of digital assets. Metadata offers valuable contextual information about digital objects, aiding in their organization, searchability, and retrieval. In journalism, metadata can improve the discoverability of archived news articles, photographs, and multimedia content, allowing researchers and journalists to trace the provenance and authenticity of information (Gilliland-Swetland, 2002). The shift to digital formats has enabled journalists to archive large amounts of content more efficiently than traditional paper-based methods. However, the dynamic nature of digital technologies presents challenges in ensuring the longevity and usability of digital archives. Digital preservation strategies must tackle issues such as file format obsolescence, hardware and software dependencies, and data integrity to prevent the loss of valuable journalistic content (Gaur & Tripathi, 2012). Journalists serve as memory assemblers, shaping collective memory through their reporting and documentation of historical events. The digitization of newspaper archives, news broadcasts, and photojournalism has expanded the impact of journalism in preserving cultural heritage and historical narratives (Reading, 2014). Through engaging in personal digital archiving practices, journalists contribute to the preservation of societal memory and the dissemination of knowledge across generations.

In scholarly publishing, open access has gained prominence as a means of enhancing the accessibility and visibility of research outputs. Open access initiatives empower researchers to self-archive their publications, making them freely available to the public and promoting the democratization of knowledge dissemination. The principles of open access align with the goals of personal digital archiving in journalism, emphasizing the importance of preserving and sharing information for the benefit of society (Martín-Martín et al., 2018).

Efforts in digital preservation in journalism are further emphasized by the need to address challenges of information overload and digital obsolescence. As the volume of digital content continues to grow exponentially, journalists must adopt robust archiving practices to ensure the longevity and accessibility of their work. The evolving landscape of digital scholarship and knowledge dissemination requires ongoing innovation in digital preservation strategies to meet the changing needs of journalists and researchers (Cramer et al., 2022).

The historical context of archiving in journalism is deeply intertwined with the evolution of information preservation practices over time. Archives have long served as repositories of knowledge, capturing the essence of societal events, cultural heritage, and historical narratives (Mazumder, 2024). In journalism, the archival function plays a crucial role in documenting news stories, preserving journalistic content, and providing a historical record of events (Severson, 2018). Understanding the historical context of archiving in journalism involves recognizing the significance of archives as sources of information that shape our understanding of the past and present (Moberg, 2023).

Challenges in digital preservation for news organizations stem from the dynamic nature of digital technologies and the rapid evolution of file formats and storage systems. News archives face issues such as format obsolescence, hardware dependencies, and data integrity concerns that threaten the long-term accessibility of archived content (Gaur & Tripathi, 2012). Additionally, the sheer volume of digital content

generated by news organizations poses challenges in terms of storage capacity, metadata management, and ensuring the authenticity of archived materials (Cramer et al., 2022). News organizations must navigate these challenges to preserve their digital archives effectively and ensure the continuity of journalistic heritage.

Emerging trends in personal digital archiving are shaping the way individuals preserve and manage their digital assets. Personal digital archiving involves the preservation of personal documents, photographs, emails, and other digital content for future access and reference. With the proliferation of digital devices and online platforms, individuals are increasingly adopting strategies to organize and archive their digital footprint (Gilliland-Swetland, 2002). Cloud storage, digital asset management tools, and metadata tagging are emerging as popular methods for personal digital archiving, enabling individuals to curate and preserve their digital memories effectively.

The historical context of archiving in journalism underscores the importance of archives as custodians of knowledge and historical records. Current practices in newsroom archiving reflect a transition towards digital preservation methods to manage journalistic content efficiently. However, news organizations face challenges in digital preservation due to technological complexities and data management issues. The emerging trends in personal digital archiving highlight the growing importance of individuals preserving their digital legacy through innovative archiving strategies. By understanding the historical context, current practices, challenges, and emerging trends in archiving, journalists and individuals can contribute to the preservation and accessibility of valuable information for future generations. By utilizing digital preservation techniques, journalists can safeguard the integrity of their work, contribute to the collective memory of society, and facilitate knowledge dissemination across diverse audiences. As digital technologies shape the journalism landscape, effective personal digital archiving practices will be essential in upholding journalistic standards and preserving the rich tapestry of human experience for future generations.

Theoretical Framework

This study is grounded in three key theoretical areas that provide context for understanding the challenges and importance of preserving born-digital news content: digital preservation theory, information lifecycle management, and professional identity and portfolio development in journalism.

Digital Preservation Theory

Digital preservation theory provides the foundation for understanding the unique challenges of preserving born-digital content compared to analog materials. (Moore, 2008) Digital content is inherently more fragile and at risk of loss than analog materials like print newspapers or books. Digital preservation theory emphasizes that preserving digital content requires active, ongoing intervention and management, rather than the "benign neglect" that may suffice for analog materials. Key principles of digital preservation theory relevant to journalists include: The need for multiple, geographically distributed copies of digital content to protect against loss; The importance of file format selection and potential format migration over time as technologies change; The critical role of metadata in ensuring digital objects remain findable, accessible and usable over time; The concept of "significant properties" - identifying and preserving the essential characteristics that give digital objects meaning and value

The OAIS (Open Archival Information System) Reference Model provides a conceptual framework for digital preservation systems and workflows. While news organizations may not implement full OAIS-compliant archives, the model offers useful concepts like Submission Information Packages, Archival Information Packages, and Dissemination Information Packages that inform preservation workflows. (THE CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS), 2012)

The SPOT (Simple Property-Oriented Threat) Model used to analyze preservation practices in this study also derives from digital preservation theory. The six essential properties it examines - availability, identity, persistence, renderability, understandability, and authenticity - encapsulate core digital preservation principles. (Vermaaten et al., 2012)

Information Lifecycle Management

Information lifecycle management (ILM) theory provides a framework for understanding how digital content should be managed from creation through active use, long-term preservation, and eventual disposition. For news content, the lifecycle begins with story planning and continues through reporting, editing, publishing, archiving, and potential future reuse.(Reiner et al., 2004)

Key ILM principles relevant to digital news preservation include: The   importance   of   capturing preservation metadata early in the content lifecycle; Defining retention policies and preservation actions for different content types;   Managing relationships between related content objects across their lifecycle; Implementing access controls and tracking content usage over time; Planning for potential content migration or emulation as systems change

Many news organizations struggle to implement effective ILM practices for digital content. Issues like inconsistent metadata application, loss of linkages between related content, and difficulty accessing older content all point to a need for more robust ILM strategies.

Professional Identity and Portfolio Development in Journalism

The third key theoretical area underpinning this study relates to how digital content preservation impacts journalists' professional identity and ability to build career portfolios. In the pre-digital era, journalists could rely on physical newspaper archives or broadcast tape libraries to maintain a record of their work. The ephemerality of digital news content poses challenges for journalists seeking to document their professional history and achievements.

Theories of professional identity emphasize the importance of a strong body of work in establishing credibility and expertise. For journalists, this traditionally meant a portfolio of published articles or aired stories. As news organizations struggle to maintain long-term access to digital content, journalists may find it difficult to retrieve and showcase past work.

Portfolio theory in career development highlights how professionals use curated collections of their best work to demonstrate skills and experience. The disappearance of news librarians and archivists from many newsrooms has impacted the quality and completeness of preserved content. Journalists often struggle to access their past work, especially when changing employers. Many news organizations rely on third-party services for public access to archives, which may not include all published content. The lack of comprehensive, easily accessible archives may hinder journalists' ability to build and showcase their professional portfolios.

By examining digital news preservation through these theoretical lenses of digital preservation, information lifecycle management, and professional identity/portfolio development, this study aims to highlight both the technical and human implications of current preservation practices in the news industry. The findings and recommendations are intended to address preservation challenges in ways that support not only long-term access to valuable historic content, but also the professional needs of the journalists creating that content.

Discussion

The findings of this study have significant implications for journalistic practice, the preservation of historical records, and the ethical considerations surrounding personal archiving of published work. This section will explore these implications and provide recommendations for improving personal digital archiving practices in journalism.

Implications for journalistic practice and professional development

The state of digital news preservation revealed by this study has profound implications for how journalists work and develop their careers. The inconsistent and often inadequate preservation practices observed across many news organizations pose challenges for journalists in several ways. The difficulty in maintaining a comprehensive portfolio of work can hinder career advancement and limit opportunities for journalists to showcase their best work to potential employers or clients. The loss of easily accessible archives can impede journalists' ability to build on past reporting, potentially leading to repetition of work or missed opportunities for in-depth follow-ups on important stories. When past articles are difficult to retrieve or have disappeared entirely, it becomes challenging for journalists to fact-check their own work or for others to hold them accountable for past reporting. As preservation practices evolve, journalists may need to develop new skills in digital asset management and personal archiving to ensure the longevity of their work.

The role of personal archives in maintaining the historical record

While institutional archives have traditionally been the primary keepers of the historical record, the findings of this study suggest an increasing importance of personal archives maintained by individual journalists. This shift has several implications. With inconsistent preservation practices across news organizations, personal archives may become crucial in filling gaps in the historical record. These archives may allow for a more diverse range of voices and perspectives to be preserved, potentially challenging or complementing official institutional narratives. However, personal archives may raise questions about the reliability and authenticity of preserved content, as individual journalists may not have the resources or expertise to implement robust preservation practices. Additionally, personal archives may be less accessible to researchers and the public compared to institutional archives, potentially limiting their usefulness in maintaining a comprehensive historical record.

Ethical considerations in personal archiving of published work

The practice of personal archiving of published work raises several ethical considerations. Journalists must navigate complex issues of copyright and content ownership when archiving their published work, especially when moving between different news organizations. Personal archives may include sensitive information or quotes from sources who agreed to be interviewed for a specific story but not for indefinite preservation and potential future use. As discussions around the "right to be forgotten" gain traction, journalists must consider how to balance the preservation of their work with individuals' desires to have certain information about them removed from public access. Personal archiving practices may lead to selective preservation based on a journalist's own biases or preferences, potentially skewing the historical record.

Recommendations for improving personal digital archiving practices

Based on the findings of this study and the implications discussed above, we offer several recommendations for improving personal digital archiving practices among journalists. Journalists should create a systematic approach to preserving their work, including regular backups, consistent metadata application, and consideration of long-term storage solutions. To ensure the longevity of their work, they should employ a combination of local storage, cloud-based solutions, and potentially physical copies of important pieces. Adopting consistent metadata practices, such as using standardized tags and descriptors, can greatly enhance the searchability and usability of personal archives. It's crucial for journalists to familiarize themselves with the copyright policies of their employers and negotiate for rights to archive and use their work when changing jobs. Where possible, they should explore opportunities to collaborate with libraries, universities, or other institutions to ensure the long-term preservation and accessibility of their work. Journalists should also push for improved preservation practices within their news organizations, highlighting the importance of comprehensive archives for both the institution and individual professionals. When maintaining personal archives, journalists should carefully consider the ethical implications, particularly regarding privacy, consent, and the potential for future harm. They should stay informed about and utilize emerging technologies that can aid in personal archiving, such as blockchain for verifying authenticity or AI-powered tools for improved metadata generation. By addressing these recommendations, journalists can take proactive steps to ensure the preservation of their work, contribute to the maintenance of the historical record, and navigate the ethical considerations inherent in personal archiving practices. As the landscape of digital news continues to evolve, the importance of robust personal archiving practices will only grow, making it an essential skill for journalists in the 21st century.

Conclusions

This study has revealed several critical insights into the state of digital news preservation. We found that while most news organizations are saving some digital content, preservation practices are often inconsistent and inadequate. There's a widespread lack of awareness about digital content fragility and proper preservation techniques. Public media organizations generally outperform private sector news outlets in preservation practices. Metadata management is frequently weak, with many organizations struggling to

maintain content linkages over time. The disappearance of news librarians and archivists from many newsrooms has negatively impacted the quality and completeness of preserved content.

These findings have significant implications for both journalism and digital preservation. For journalism, the inadequate preservation of digital news content threatens the industry's ability to maintain a reliable historical record and support journalists' professional development. It challenges the fundamental role of journalism as society's first draft of history. For the field of digital preservation, this study highlights the unique challenges posed by news content, with its rapid production cycle and complex mix of text, images, audio, and video. It underscores the need for preservation strategies that can keep pace with the evolving digital news landscape.

The study also emphasizes the growing importance of personal digital archiving practices among journalists, suggesting a shift in responsibility for preservation from institutions to individuals. This shift has implications for how we conceptualize digital preservation in professional contexts and highlights the need for education and tools to support personal archiving efforts.

This study opens up several avenues for future research. Further investigation is needed into effective models for integrating preservation practices into digital news production workflows. Research could explore the potential of emerging technologies, such as blockchain or AI, in enhancing digital news preservation. Studies examining the long-term impacts of inadequate preservation on journalistic practice and historical research would be valuable.

Additionally, research into the ethical implications of personal digital archiving in journalism could provide guidance for developing best practices. Comparative studies across different countries and media systems could offer insights into how cultural, legal, and economic factors influence digital news preservation practices.

Finally, interdisciplinary research bringing together journalists, archivists, computer scientists, and ethicists could lead to innovative solutions for the challenges identified in this study. Such collaborative efforts are crucial for ensuring that the vital role of journalism in documenting and interpreting our world is not lost in the digital age.

References

404 Not Found—HTTP | MDN. (2024, July 25). https://developer.mozilla.org/en-US/docs/Web/HTTP/Status/404

Belter, C. (2014). Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. *Plos One*. https://doi.org/10.1371/journal.pone.0092590

Cramer, T., German, C., Jefferies, N., & Wise, A. (2022). A Perpetual Motion Machine: The Preserved Digital Scholarly Record. *Learned Publishing*. https://doi.org/10.1002/leap.1494

Gaur, R., & Tripathi, M. (2012). Digital Preservation of Electronic Resources. *Desidoc Journal of Library & Information Technology*. https://doi.org/10.14429/djlit.32.4.2522

Gilliland-Swetland, A. J. (2002). Digital Preservation and Metadata: History, Theory, Practice. *The Journal of Academic Librarianship*. https://doi.org/10.1016/s0099-1333(02)00300-2

Martín-Martín, A., Costas, R., van Leeuwen, T. N., & López-Cózar, E. D. (2018). Evidence of Open Access of Scientific Publications in Google Scholar: A Large-Scale Analysis. *Journal of Informetrics*. https://doi.org/10.1016/j.joi.2018.06.012

Mazumder, A. L. (2024). The Archives as a Source of Social History: Studying Belonging of the Indian–Chinese Community. *Studies in Indian Politics*. https://doi.org/10.1177/23210230241235358

McCain, E., Nell, M., Kara, V. M., Dorothy, C., Bernard, R., Kerri, W., Sandra, S., Joe, A., & Sarah, B. (2021). *Endangered but not too late: The state of digital news preservation*. University of Missouri. https://doi.org/10.32469/10355/80931

Moberg, B. R. (2023). The Faroese Cultural Archive: The Archive as a Source and Theme in Local Historical Writing. *Scandinavistica Vilnensis*. https://doi.org/10.15388/scandinavisticavilnensis.2023.12

Moore, R. (2008). Towards a Theory of Digital Preservation. *International Journal of Digital Curation*, *3*(1), 63–75. https://doi.org/10.2218/ijdc.v3i1.42

Reading, A. (2014). *The Journalist as Memory Assembler: Non-Memory, the War on Terror and the Shooting of Osama Bin Laden*. https://doi.org/10.1057/9781137263940_11

Reiner, D., Press, G., Lenaghan, M., Barta, D., & Urmston, R. (2004). Information lifecycle management: The EMC perspective. *Proceedings. 20th International Conference on Data Engineering*, 804–807. https://doi.org/10.1109/ICDE.2004.1320052

Severson, P. (2018). The Politics of Women's Digital Archives and Its Significance for the History of Journalism. *Digital Journalism*. https://doi.org/10.1080/21670811.2018.1513336

Tameez, H. (2024). To preserve their work—And drafts of history—Journalists take archiving into their own hands. *Nieman Lab*. https://www.niemanlab.org/2024/07/to-preserve-their-work-and-drafts-of-history-journalists-take-archiving-into-their-own-hands/

THE CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). (2012). *Reference Model for an Open Archival Information System (OAIS)—CCSDS 650.0-M-2*. CCSDS 650.0-M-2. http://public.ccsds.org/publications/archive/650x0m2.pdf

Vermaaten, S., Lavoie, B., & Caplan, P. (2012). Identifying Threats to Successful Digital Preservation: The SPOT Model for Risk Assessment. *D-Lib Magazine*, *18*(9/10). https://doi.org/10.1045/september2012-vermaaten