

iPRES 2024 • iPRES 2024 Proceedings

Archiving Digital Marketing

**Christopher Rauch¹ Alex H. Poole¹ Travis Reid² Michele C. Weigle²
Michael L. Nelson² Faryaneh Poursardar² Mat Kelly¹**

¹Drexel University, ²Old Dominion University

Published on: Aug 30, 2024

URL: <https://ipres2024.pubpub.org/pub/x38rbwuu>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Abstract – The transition to digital marketing has revolutionized advertising, reflecting and shaping societal norms and trends. The “Saving Ads” project addresses the challenges of preserving these ephemeral digital artifacts, essential for understanding the evolution of advertising and its socio-cultural impact. The initiative focuses on technical solutions for archiving dynamic online ads and enhancing access to these critical resources for future scholarship. By examining the preservation of online advertisements and suggesting improved approaches for archiving dynamic online content, this project contributes to the documentation of digital history.

Keywords – [Digital marketing](#), [Online advertisements](#), [Web archiving](#), [Cultural preservation](#), [Digital humanities](#)

This paper was submitted for the iPRES2024 conference on March 17, 2024 and reviewed by 4 anonymous reviewers. The paper was accepted with reviewer suggestions on May 6, 2024 by co-chairs Heather Moulaison-Sandy (University of Missouri), Jean-Yves Le Meur (CERN) and Julie M. Birkholz (Ghent University & KBR) on behalf of the iPRES2024 Program Committee.

Introduction

Global networked computing has fomented a sea change in marketing, advertising, and cultural representation more broadly, marking a transformation that began in the mid-1990s with the public’s increasing access to the Internet. The World Wide Web represents the primary staging ground for these transformations. Online advertisements, in particular, have emerged as critical digital artifacts since they reflect prevailing political, cultural, social, and economic events, norms, and trends. The seismic shift from analog to digital marketing media not only revolutionized advertising strategies but also introduced new challenges in preserving these complex, ephemeral digital artifacts.

Supported by a National Leadership Planning Grant from the Institute of Museum and Library Services (IMLS), "Saving Ads: Assessing and Improving Web Archives’ Holdings of Online Advertisements" (2022-2024) is a collaborative initiative between Drexel University and Old Dominion University. It focuses on archiving digital marketing materials, including their broader networked context. Its impact is two-fold. First, preserving these digital traces is essential for future scholarly exploration. Through these artifacts, scholars may examine the evolution of online advertising and how it shapes and reflects political, social, and cultural mores. Second, the project has developed strategies to address the technical challenges of archiving web ads.

Why Preserve Online Advertisements

Advertisements are not merely promotional tools but are complex molders and reflections of cultural and societal trends, norms, and values over time. [1] After the removal of regulatory barriers in 1991, the Internet saw its first marketing endeavors by 1993, becoming a recognized mass medium for advertising by 1996. This evolution is characterized by the dynamic adoption and innovation of advertising strategies, merging traditional

methods with new approaches to blur the distinction between advertising and content. Initial resistance to Internet advertising proved unsustainable over time. [2] Online advertising now captures a significant share of overall advertising expenditure.

As a novel type of documentary evidence, online advertisements are indispensable sources for histories written about the mid-1990s and afterward, enabling scholars to explore new research questions and revisit existing findings. [3] However, the transient nature of web content, including advertisements, poses a significant challenge to preservation. Failing to capture this content will lead to serious shortcomings in the documentary record. Without archiving efforts, future researchers will be unable fully to understand this era [4]. Moreover, web ads can provide insights into how firms market products to various demographics, offering a resource for studying legal and social issues related to advertising practices. [5]

Preserving online advertisements also exposes critical aspects of microtargeting, a practice where advertisements are tailored to specific audiences based on detailed user data. This is particularly relevant for scholars studying political phenomena. Research has shown that microtargeting can play a significant role in divisive issue campaigns, where ads are strategically designed to influence voter behavior in key demographics and geographic areas. For example, Kim et al. (2018) illustrate microtargeting to disseminate politically charged ads. [6]

Finally, web ads can help students and researchers understand advertising strategies and media literacy, fostering a critical view of media consumption's influence on societal values and behaviors. [7]

Challenges

Archiving online advertisements presents unique technical hurdles, from capturing dynamic content to addressing compatibility issues with archival tools. Enhancing user interfaces, supporting dynamic content, raising public awareness, fostering partnerships, creating specialized archives, improving metadata, and developing legal and ethical frameworks are pivotal steps toward improving access to web archives, thereby ensuring the preservation of digital marketing's legacy for posterity.

The preservation of digital ads requires not just storing files but also ensuring they can be replayed in future browsers, which may not support older media formats or web standards. Klein et al. (2019) highlight the increasing difficulty in maintaining archival quality due to the proliferation of dynamic web content, particularly content that is only accessible through the activation of JavaScript-based features. [8]

Online ads often depend on external scripts, tracking pixels, and third-party services to function correctly. These dependencies can make it difficult to archive an ad in a way that preserves its original context and functionality. For instance, certain Google ads use a Google SafeFrame with a random value in the URL's subdomain, making it challenging to replay the ad content since the random value generated during replay does not match the one generated during crawl time.

Additionally, compatibility issues can arise with the tools used for archiving. For example, recent versions of Google Chrome have been found incompatible with Brozzler, a web crawling tool, which prevents web pages from being loaded and archived successfully.

In addition, institutional archives often resist archiving online advertisements for various reasons, ranging from technical challenges to philosophical considerations about what constitutes valuable or significant historical material. But as our work suggests, these institutions would do well to reconsider the archiving of online advertisements. This might involve collaborating with copyright holders, developing more sophisticated web crawling technologies, and reevaluating the criteria for what constitutes historically significant web content. Engaging in dialogue about the cultural and historical importance of advertising could also help shift perspectives and demonstrate the value of preserving this aspect of our digital heritage. The “Saving Ads” project is a resource to encourage these conversations.

Methods

Data Collection

Different web archiving tools and services were employed to capture a wide array of advertisements. These tools included Internet Archive's Save Page Now, Arquivo.pt, archive.today, and Conifer, which were chosen because they allow users to archive web pages without restrictions on the number of pages archived without cost. Additionally, other browser-based web archiving tools such as ArchiveWeb.page, Browsertrix Crawler, and Brozzler were utilized for their capabilities in capturing dynamic and interactive elements of web pages containing advertisements.

In our study, the selection of online advertisements for archiving was guided by specific criteria to ensure a representative and diverse sample. The web pages we archived were selected from SimilarWeb's top websites worldwide (excluding the “Adult” category). The process involved creating a list of top websites from SimilarWeb's categories and then randomly sorting this list. Web pages were selected for archiving by filtering the list for web pages with advertisements.

The selected ads covered various formats, including text, image, video, and combinations thereof, encapsulated as iframe elements when embedded in pages. The primary sources of these advertisements were major web archiving services, and live web captures performed from January through June 2023. This comprehensive approach ensured that the selection process captured a broad spectrum of advertisements, reflective of the dynamic and diverse nature of online advertising content.

Data Capture and Analysis

The metrics used to assess the fidelity of the archiving process were replay success rates and the types of ads captured. This involved measuring how often ads were replayed successfully without issues and categorizing

the types of ads (text, image, video, combination).

Several technical challenges were identified, such as compatibility issues with current web archiving tools and the dynamic nature of ad content. Specific examples included problems with replaying ads that use dynamic JavaScript generated page elements or external scripts.

Strategies to overcome these challenges involved the development and integration of new archival tools and methods. This included adapting existing web-crawling technologies and developing techniques to ensure better replayability of archived content.

The replayability of archived ads was validated by testing the archived web pages in different replay systems to ensure consistency and accuracy. To determine success or failure, the web page was loaded, and interactions with the ad were performed to check if it worked the same as during the crawling session. If the ad did not load after replaying the archived web page multiple times, it was labeled "failed to replay" or "never replay," except in cases where a different version of the replay system resolved the issue. If the ad would sometimes replay after multiple attempts, it was labeled as "sometimes replay." If the ad always replayed correctly, even with different versions of the replay system, it was labeled as "always replay."

Study Results

In our study, we analyzed the replayability of archived online advertisements within their original web pages after a crawling session. Our findings reveal that 40.66% of the ads were consistently successful in replaying without any issues. However, 41.39% of ads experienced intermittent replay failures due to factors like the replay system version, errors during web page replay, such as delayed dynamic resource loading, or a change in the set of ads selected during replay. Significantly, 17.95% of ads consistently failed to replay, often because no requests for content were made for these ads during the replay process. This issue was frequently linked to the failure of ad iframes, notably Google SafeFrame or Amazon's ad iframes, to load properly, thereby preventing the ads from being displayed.

Our categorization of ads showed that combination ads, which combine various elements like text, images, and videos, formed the majority (57.14%) of our dataset. Image ads were also prevalent, making up 29.30% of all ads identified. When examining the replay success rate by ad category, we found that text-only ads always successfully replayed. Image ads had a more than 50% success rate of always replaying correctly. The primary challenges to replaying image ads involved ads not being selected for replay or web pages failing to dynamically load all resources, including ad iframes that did not replay successfully.

Video ads presented the most significant challenge, with 80% consistently failing to replay, mainly due to the absence of ad requests during the replay. Embedded web page ads also faced difficulties, with around 90% not loading consistently during replay sessions. These ads typically depend on ad iframes, and their replay failure

was also often attributed to issues with Google SafeFrame or Amazon ad iframes, depending on the version of the replay system used.

For combination ads, about 58.97% only sometimes replayed successfully, impacted by variability in ad selection during page loads and issues with dynamic resource loading on web pages. These findings underscore the complexity of archiving and replaying online advertisements, highlighting the need for improved techniques and understanding of how different ad formats and replay systems interact.

Solutions

Improving access to online advertisements in web archives involves addressing both technical challenges and usability concerns. A multifaceted approach incorporating technological advancements, user engagement, and collaborative efforts is essential for the effective preservation and future use of these digital artifacts.

Enhanced User Interface and Tools

Developing more intuitive and feature-rich user interfaces for web archive platforms can significantly improve accessibility. Tools that enable users to search, browse, and view archived web content and advertisements in a user-friendly manner can encourage wider use. Incorporating advanced search functionalities, such as full-text search and metadata-based filtering, can make it easier for users to find specific advertisements or web content within large archives. [9]

Increased Support for Dynamic Content

Many web pages, including those containing online advertisements, rely heavily on dynamic content generated by JavaScript, asynchronous calls, and other web technologies. Improving the capability of web archiving tools to capture and accurately replay this dynamic content can enhance the fidelity of archived advertisements. This may involve developing new methodologies [10] or enhancing existing ones, like browser-based archiving, to ensure that interactive elements and advertisements are preserved and can be interacted with upon replay.

Partnerships and Collaboration

Establishing partnerships with websites, advertisers, and ad networks to ensure that advertisements are archived in a manner that respects copyright and privacy yet remains accessible for future research can be beneficial. These collaborations can help address legal and technical barriers to archiving advertisements. The International Internet Preservation Consortium (IIPC), for example, hosted a presentation on the topic, allowing us to present our findings during their 2024 Web Archiving Conference (WAC). This platform helped communicate the value of capturing web advertisements and facilitated connections with scholars also interested in this material, fostering a collaborative environment for future advancements.

Specialized Archives for Advertisements

Creating archives specifically dedicated to preserving online advertisements could provide a centralized resource for researchers and historians interested in the study of advertising trends, consumer behavior, and cultural shifts. Such archives could implement targeted crawling strategies to efficiently capture advertisements across the web.

Legal and Ethical Frameworks

Addressing legal and ethical concerns related to copyright, privacy, and data protection is an often overlooked mechanism for improving access to web archives. Developing clear policies and frameworks that balance the need for preservation with the rights of content creators and individuals can facilitate broader acceptance and use of web archives. [\[11\]](#)

Conclusion

The goal of preserving online advertisements reveals a complex landscape where digital heritage and technological challenges intersect. The "Saving Ads" project underscores the critical importance of archiving digital marketing materials to safeguard the rich tapestry of our digital era for future generations. As we have seen, online advertisements are not just manifestations of commercial entities but are deeply intertwined with the socio-cultural and economic fabric of their times, offering invaluable insights into the evolution of consumer behavior, marketing strategies, and societal norms.

The technical challenges associated with archiving dynamic and ephemeral digital content, such as online advertisements, necessitate innovative solutions and collaborative efforts. Enhancing archival tools, fostering public awareness, and developing robust legal and ethical frameworks can help us overcome these obstacles and ensure that digital advertisements are preserved with integrity.

As we move forward, it is imperative for researchers, archivists, and technologists to continue exploring and refining the methodologies for digital preservation. The efforts to archive online advertisements not only serve the purpose of historical documentation but also provide a foundation for future scholarly research, enabling a deeper understanding of the digital artifacts that constitute the raw materials of digital history. The "Saving Ads" project, therefore, represents a significant step towards recognizing and preserving the digital footprints that define our contemporary world, ensuring that the traces of today's digital marketing landscape are accessible to the scholars and historians of tomorrow.

Future Work

Need for Personalized Advertising Studies

The advent of digital tracking technology and prolific data gathering has ushered in an era of unprecedented personalization in advertising, a trend that significantly impacts consumer behavior, marketing strategies, and

the broader societal landscape. As online advertisements become increasingly tailored to individual preferences, behaviors, and demographics, understanding the nuances of these personalized approaches and their effects is a significant challenge. This area of study promises to shed light on the effectiveness of personalized advertising, its implications for privacy and consumer autonomy, and its role in shaping cultural and social norms.

Objectives of Future Research

1. Investigate how personalization techniques influence the effectiveness of advertising campaigns in various sectors.
2. Explore how the psychological aspects of personalization, such as feelings of being understood or intruded upon.
3. Address the growing concerns surrounding data privacy and the ethical implications of using consumer data for personalization.
4. Examine the trade-offs consumers are willing to make between personalized experiences and privacy.
5. Examine the broader cultural and social implications of personalized advertising, particularly how it affects the diversity of content presented to individuals and the potential for creating echo chambers or reinforcing stereotypes. This research will contribute to the ongoing discourse on the societal impact of algorithm-driven content personalization.
6. Examine the use of artificial intelligence (AI) and machine learning algorithms and their potential to transform personalization. Understanding these innovations will be key to anticipating future trends in digital marketing and their implications for consumers and society.

As the Web continues to include diverse new users, the exploration of personalized advertising and its multifaceted impacts will be critical in understanding both the historical context and the context of current events. This line of inquiry not only promises to enhance academic and practical knowledge in digital marketing but also to foster a deeper understanding of the interplay between technology, society, and the individual. By examining various personas and the effects of personalization in advertising, future work will contribute to a more nuanced and comprehensive discourse on the role of digital advertisements in society.

ACKNOWLEDGMENTS

This research was made possible through the generous support of the Institute of Museum and Library Services (IMLS) #LG-252362-OLS-22.

References

1. McDONALD, R.E., LAVERIE, D.A., AND MANIS, K.T. 2021. The Interplay between Advertising and Society: An Historical Analysis. *Journal of Macromarketing* 41, 4, 585–609. [↵](#)

2. BUCHWITZ, L.A. 2018. A model of periodization of radio and internet advertising history. *Journal of Historical Research in Marketing* 10, 2, 130–150. [↵](#)
3. KELLY, M., BRUNELLE, J.F., WEIGLE, M.C., AND NELSON, M.L. 2013. On the Change in Archivability of Websites Over Time. *Research and Advanced Technology for Digital Libraries*, Springer, 35–47. [↵](#)
4. ROSENZWEIG, R. 2003. Scarcity or Abundance? Preserving the Past in a Digital Era. *The American Historical Review* 108, 3, 735–762. [↵](#)
5. SWEENEY, L. 2013. Discrimination in online ad delivery. *Communications of the ACM* 56, 5, 44–54. [↵](#)
6. KIM, Y.M., HSU, J., NEIMAN, D., ET AL. 2018. The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook. *Political Communication* 35, 4, 515–541. [↵](#)
7. HWANG, T. 2020. *Subprime Attention Crisis*. Macmillan. [↵](#)
8. KLEIN, M., SHANKAR, H., BALAKIREVA, L., AND VAN DE SOMPEL, H. 2019. The Memento Tracer Framework: Balancing Quality and Scalability for Web Archiving. *Digital Libraries for Open Knowledge*, Springer International Publishing, 163–176. [↵](#)
9. JACKSON, A., LIN, J., MILLIGAN, I., AND RUEST, N. 2016. Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities. *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, Association for Computing Machinery, 103–106. [↵](#)
10. KIESEL, J., KNEIST, F., ALSHOMARY, M., STEIN, B., HAGEN, M., AND POTTHAST, M. 2018. Reproducible Web Corpora. *Journal of Data and Information Quality* 10, 4. [↵](#)
11. LEERSSEN, P., AUSLOOS, J., ZAROUALI, B., HELBERGER, N., AND VREESE, C.H. DE. 2019. Platform ad archives: promises and pitfalls. *Internet Policy Review* 8, 4. [↵](#)