

# SAVING ADS

Assessing and Improving Web Archives' Holdings of Online Advertisements

Christopher Rauch (cr625@drexel.edu), Travis Reid (treid003@odu.edu)

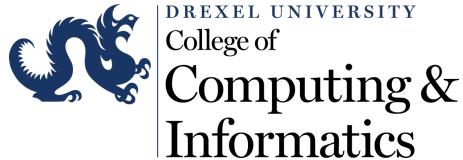
Advised by Profs. Mat Kelly, Alex H. Poole, Michael Nelson, Michele C. Weigle

mkelley@drexel.edu

ahp56@drexel.edu@drexel.edu

mln@cs.odu.edu

mweigle@cs.odu.edu





# Introduction

---

- The World Wide Web is a staging ground for transformations in marketing, advertising, and cultural representation.
- Online advertisements capture the essence of their times by reflecting prevailing economic trends, societal norms, and cultural shifts.
- The shift from traditional to digital marketing introduced challenges in preserving inherently ephemeral digital artifacts.



# The Saving Ads Project

---

- Collaborative initiative between Drexel University and Old Dominion University
- Supported by a grant from the Institute of Museum and Library Services (IMLS)
- Focuses on archiving digital marketing materials in their broader, networked context
- Examines the evolution of online advertising, the complexities of archiving such fluid content, and the strategies developed to address these challenges
- Highlights the significance of archiving digital marketing to document this era of digital 'small' history for future generations
- Underscores the broader implications of these archival efforts for subsequent scholarship

# Does it Matter?

**BBC**

**Top News story**  
Greece faces final coalition push

Greece's president faces the task of trying to forge a deal into an emergency government, after days of political wrangling and deepening economic fears.

**News**

**Chavez cancer treatment 'success'**  
Venezuelan President Hugo Chavez returns from Cuba, saying his radiotherapy treatment for cancer is over and has been successful.

**Business**

Extra capital for Spanish banks  
Eurozone economy 'to contract'

**Sport**

We can still win the title, says Eva

**Weather**

London

Saturday	Sunday	Monday	Tuesday
Max: 16°C Min: 8°C	Max: 17°C Min: 9°C	Light Rain Max: 13°C Min: 7°C	Light Rain Shower Max: 12°C Min: 4°C

**More from BBC News**

**Entertainment & Arts**

Kapoor demands Orbit ticket price

Artist Anish Kapoor gives a preview of his latest work, the Orbit tower sculpture, but admits the £15 ticket price is "half a lot of money".

Mercury "to appear" in October

30 Rock series to come to an end

**Technology**

US panel approves new HIV pill

A panel of US health experts recommends lamivudine and Truvada, a drug to prevent HIV infection in healthy people.

Dutch court bars Pirate Bay links

Games age rating rules changed

**Science & Environment**

Facebook tests "pay to post" tool

A panel of US health experts recommends lamivudine and Truvada, a drug to prevent HIV infection in healthy people.

Vesta is "last of a kind" rock

"Sonic weapon" deployed for Games

**Future**

Radiation rage

Why the Incredible Hulk has made a monster out of gamma rays

Ctr-Alt-Del

Is the complexity of our gadgets to blame for constant crashes?

**BBC**

**Top News Story**  
Greece faces final coalition push

Greece's president faces the task of trying to forge a deal into an emergency government, after days of political wrangling and deepening economic fears.

**News**

**Guilty verdicts in Hudson murders**  
Ajae Chikwe, son of the former brother-in-law of Oscar-winning actress Jennifer Hudson, guilty of murdering three members of her family.

**Saudi king sacks key conservative**

JPMorgan shares plunge on loss

**Most Popular in News**

- One-minute World News
- Egypt's first presidential debate
- Mystery over Indonesia plane crash
- Royal Navy pilot's 'Top Gun' training

**Spotlight**  
Africa Beats

**Chavez cancer treatment 'success'**  
Venezuelan President Hugo Chavez returns from Cuba, saying his radiotherapy treatment for cancer is over and has been successful.

**Sport**

We can still win the title, says Eva

**Weather**

London

Saturday	Sunday	Monday	Tuesday
Sunny Max: 16°C Min: 8°C	Sunny Max: 17°C Min: 9°C	Light Rain Max: 13°C Min: 7°C	Light Rain Shower Max: 12°C Min: 4°C

**More from BBC News**

**Entertainment & Arts**

Hudson demands Orbit ticket price

Artist Anish Kapoor gives a preview of his latest work, the Orbit tower sculpture, but admits the £15 ticket price is "half a lot of money".

Mercury "to appear" in Queen show

30 Rock series to come to an end

**Health**

US panel approves anti-HIV pill

A panel of US health experts recommends lamivudine and Truvada, a drug to prevent HIV infection in healthy people.

Most child deaths preventable

"Cell shield" for cancer patients

Games age rating rules changed

**Technology**

Facebook tests "pay to post" tool

A panel of US health experts recommends lamivudine and Truvada, a drug to prevent HIV infection in healthy people.

UV safety land grab guidelines

The United Nations says countries that pay a small fee to make some ports more accessible to developing nations.

Mobile court bars Piracy Bay links

Vesta is "last of a kind" rock

"Sonic weapon" deployed for Games

**Science & Environment**

UV safety land grab guidelines

The United Nations says countries that pay a small fee to make some ports more accessible to developing nations.

Mobile court bars Piracy Bay links

Vesta is "last of a kind" rock

"Sonic weapon" deployed for Games

**Future**

Radiation rage

Why the Incredible Hulk has made a monster out of gamma rays

Ctr-Alt-Del

Is the complexity of our gadgets to blame for constant crashes?

## Spot the Differences

The image shows two nearly identical screenshots of the CNBC website homepage from April 25, 2024. The left screenshot is at the top of the page, and the right one is slightly lower down, showing more content.

**Top Navigation:**

- SET HOME PAGE | UNITED STATES | ASIA PACIFIC | EUROPE
- REAL-TIME QUOTES | SEARCH
- Symbol / Company | Go | Symbol Lookup | Go
- Welcome, Guest | Register | Sign In

**Main Content Area (Left Screenshot):**

### Stocks Close Lower, Led by Energy; Banks Gain

Dow Jones Industrial Average

Market	U.S.	Europe	Asia	Oil	Gold
DJIA	16436.41	▲ 76.23	0.47%	—	—
NASDAQ	—	—	—	—	—
S&P 500	1670.55	▲ 4.74	0.25%	—	—

**What Investors Should Know:**

- After-Hours Buzz: FB, LGF & More
- Why Apple May Boost Dividend Soon
- Best Trade of the Past 20 Years?
- Google's Sound, Apple's More Attractive
- With ETFs, Is Too Much Too Much?
- Currency War? Here's How to Hedge It
- Egypt Is Ravenous for Energy
- ECB's Weidmann: No Euro Intervention
- Which ETFs Are Investors Buying?
- Your Safety Information for Sale
- Frugal Year of the Snake Ushered In
- Weird ETFs, Revisited
- Dow 30 After-Hours Quotes

**Bottom Content Area (Left Screenshot):**

### HIGHLIGHTS

- Pity the Manager
- Has Volkswagen Found the Holy Grail?
- Celebrity Love Nests
- Dew in the Morning
- Carnival Rescue?
- Fido is Movin' On Up

**Top News and Analysis (Right Screenshot):**

### 4% Interest, Without Fees: Too Good to Be Checking?

Decent rates are not a thing of the past, thanks to a group dedicated to fighting for your money and even chance against the big boys.

Banks Still Raising Rates

**What Investors Should Know:**

- American-U.S Air Deal Could Be End of Airlines Stock Run
- + American-U.S Airways: What It Means for Frequent-Fliers
- Stocks Better Investment Than Homes: Shiller
- Wall Street Enforcement Lost in Revolving Door: Study
- Why Si Swimsuit Has Buzz Despite Leak—Katie Upton
- Bourbon Controversy: Just Add Water
- Pope Benedict to Resign for Health Reasons
- Next Pope? Ghana Cardinal Turkson Takes Lead
- One Man's Quest: A Beer for Diabetics
- Starbucks' New Logo and Apple CEO Say What? Bill Gates Gets Personal on Reddit
- Can US Companies' Chinese Results Be Trusted?
- What Immigration Reform Means for the Economy
- Highlights From NY Fashion Week
- Models on Twitter: Tech Hits the Runway
- 5% of Credit Reports Contain Costly Errors: FTC
- Fed Taking 'Forceful Action' on Economy: Yellen
- Pentagon Cuts Only Possible Under Sequestration: Dean Chasen
- Big Payoff, Savers Lose on Complex Bets
- Stars of Nerd Philanthropy Dominate Top 50 Givers

**Bottom Content Area (Right Screenshot):**

### HIGHLIGHTS

- Pity the Manager
- Has Volkswagen Found the Holy Grail?
- Celebrity Love Nests
- Dew in the Morning
- Carnival Rescue?
- Fido is Movin' On Up

**Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements**  
**IIPC Web Archiving Conference**  
**April 25, 2024**

cr625@drexel.edu

# Case Study: USA Today

The screenshot shows the USA Today homepage with several news articles and advertisements. At the top, there are three tabs for 'Breaking News' and 'USA TODAY'. The main navigation bar includes links for 'For You', 'U.S.', 'Sports', 'Entertainment', 'Life', 'Money', 'Tech', 'Travel', and 'Opinion'. A 'SUBSCRIBE NOW' button is visible on the right.

**Advertisement:** MSC Industrial Supply

**Top Headlines:**

- LIVE** Mother of survivor of Mexico shooting says daughter was traveling for surgery (Video)
- Tucker Carlson hated Trump 'passionately.' Takeaways from lawsuit docs
- FOR SUBSCRIBERS Why Memphis didn't burn after Tyre Nichols video release
- Women's history month: Celebrating historical moments for women in Congress
- Controversial plan: Haul hot crude oil along new railroad
- Bindi Irwin and the power of celebrities revealing health diagnoses
- A GOP war on 'woke'? Poll finds most Americans see term as positive
- Biden's budget proposal endgame: Forcing Republicans' hand

**Advertisement:** NFL

**News Stories:**

- MOTHER OF SURVIVOR** Mother of survivor of Mexico shooting says daughter was traveling for surgery (Video)
- NATION** Tucker Carlson and Jan. 6 tapes divide McConnell and McCarthy on Trump
- POLITICS** Watch: Jill Biden honors women from around the globe at White House event
- POLITICS** This sickens me: Michael Irvin addresses hotel worker's allegation
- POLITICS** A stranger filmed me without consent. In the end, I found community.
- VOICES** COVID lab theory debated by

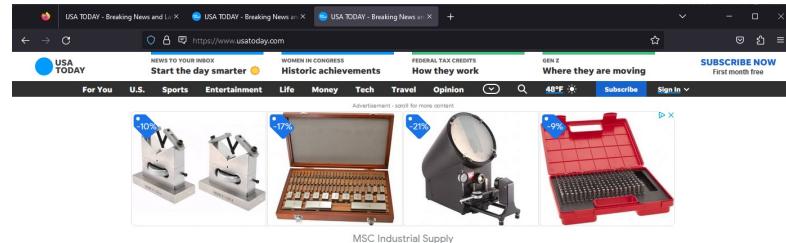
**Advertisement:** Brother Office USA

Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements  
IIPC Web Archiving Conference  
April 25, 2024

cr625@drexel.edu

# Case Study: USA Today

Live



container + data-item for ad collection

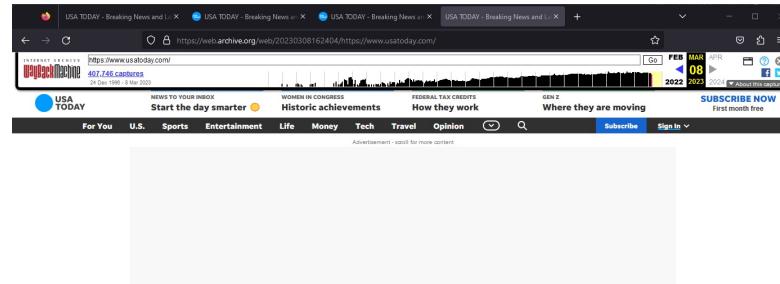
```
<aside aria-label="advertisement" class="gnt_m gnt_x gnt_x_lbl gnt_x_hi">
  <div id="ad-slot-7103-usatoday-high_impact-homepage-2" class="gnt_x_sl gnt_x_hi"
    data-google-query-id="CPKqy4GLzf0CFUzKswodfc8B0g">
    <div id="google_ads_iframe_7103/usatoday/high_impact/homepage_0_container_"
      style="border: 0px none; min-width: 728px; min-height: 90px; width: 970px; height: 250px; display: inline-block;">
      <div class="gnt_x_sl_gpt"><iframe style="border: 0px none; vertical-align: bottom;">
        <div style="border: 0px none; width: 100%; height: 100%; position: relative; background-color: white; margin-bottom: 10px;">
          <div style="border: 0px none; width: 100%; height: 100%; position: absolute; left: 0px; top: 0px; background-color: white; z-index: 1; display: flex; align-items: center; justify-content: center; font-size: 10px; color: #ccc; opacity: 0.5;">
            <span>3rd party ad content</span>
          </div>
          <div style="border: 0px none; width: 100%; height: 100%; position: absolute; left: 0px; top: 0px; background-color: white; z-index: 2; display: flex; align-items: center; justify-content: center; font-size: 10px; color: #ccc; opacity: 0.5;">
            <span>Advertisement</span>
          </div>
        </div>
        <div style="border: 0px none; width: 100%; height: 100%; position: relative; background-color: white; z-index: 3; display: flex; align-items: center; justify-content: center; font-size: 10px; color: #ccc; opacity: 0.5;">
          <span>Advertisement</span>
        </div>
      </div>
      <div style="border: 0px none; width: 100%; height: 100%; position: relative; background-color: white; z-index: 4; display: flex; align-items: center; justify-content: center; font-size: 10px; color: #ccc; opacity: 0.5;">
        <span>Advertisement</span>
      </div>
    </div>
  </div>
</div>
```

Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements  
IIPC Web Archiving Conference  
April 25, 2024

cr625@drexel.edu

# Case Study: USA Today

Internet  
Archive



## Sized container but no data-item

```
<aside aria-label="advertisement" class="gnt_m gnt_x gnt_x_lbl gnt_x_hi">
  <div id="ad-slot-7103-usatoday-high_impact-homepage-2" class="gnt_x_sl gnt_x_hi">
    <div id="google_ads_iframe_7103/usatoday/high_impact/homepage_0_container_"
      style="border: 0pt none; width: 1000px; height: 0px; min-width: 728px; min-height: 90px;" class="gnt_x_sl_gpt">
    </div>
  </div>
</aside>
```

Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements  
IIPC Web Archiving Conference  
April 25, 2024

cr625@drexel.edu

# Case Study: USA Today

The screenshot shows the USA Today homepage with a navigation bar at the top. The main content area features several news articles and a 'Top Headlines' sidebar.

**Top Headlines:**

- LIVE: The steepest Apple deals on iPads, Watches, AirPods
- 'Am I safe being in public?: Fear follows mass shootings
- Michigan State paid coach full bonus marked for him, entire staff
- After draw with England, how can the USMNT advance at the World Cup?
- FOR SUBSCRIBERS: Don't boycott the World Cup. But don't ignore the human rights issues. Columnist Carl Person
- FOR SUBSCRIBERS: When malpractice occurs at community health centers, taxpayers pay
- Ancient coin could be proof that Roman emperor was real, researchers say
- FOR SUBSCRIBERS: Kentucky lawyer's own book could free man imprisoned in her kidnapping
- Get the Daily Briefing in your inbox

**Advertisement:** (Sponsored by pywb)

Retrieved from common crawl

# Out of Context

- not possible to retrieve the containing page (small chance at reproduction)



# Out of Context

- Vendor archives

## Facebook

● Inactive  
Jul 3, 2023 - Jul 4, 2023  
Platforms:   
Categories:   
Estimated Audience Size: Not available   
Amount spent (PHP): <PHP100  
Impressions: <1K  
This ad ran without a disclaimer   
ID: 140456920342095

[See ad details](#)

**Author's Tranquility Press**  
Sponsored  
When democracy faces its biggest test, who holds the power?   
In a world where every vote counts, one elector's decision turns the tables. Explore the powerful narrative of "One Vote" by J. Stewart Willis, where a faithless elector puts the nation's fate at stake...  
  


[Send message...](#)

## Google

All topics Search by advertiser or website name

**Tesla Inc.**  
Legal name: Tesla Inc.  
Based in: the United States  
 Advertiser has verified their identity

54 ads

Sorting by most recent

**Sponsored** [Tesla Model 3—Built in America - Starting at \\$32,740\\*](#)  
Model 3: Starts at \$32,740 after federal tax credit - excludes taxes and fees. Learn more Up to 333 mi range. Charge at home or at 45,000+ Superchargers globally. Drive one today.  
[Experience Tesla](#) [Design your Model 3](#) [Schedule a Demo](#)

**Sponsored** [Electric Family SUV - Schedule a Demo Drive](#)  
Family SUV starting at \$40,240 after federal tax credit - excludes taxes and fees. Entertainment for the whole family comes as standard with every Tesla. Drive one...  
[View Local Inventory](#) [Design your Model Y](#) [Explore Events](#)

**Tesla Inc.**

**Sponsored** [Up to 333 miles of Range - Starting at \\$32,740\\*](#)  
Model 3: Starts at \$32,740 after federal tax credit - excludes taxes and fees. Learn more Charge at home...  
[Tesla Inc.](#)

**Sponsored** [No Annual Maintenance Required - Starting at \\$40,240\\*](#)  
Model Y: Starts at \$40,240 after federal tax credit - excludes taxes and fees. Learn more Explore Events.  
[Tesla Inc.](#)

● Active  
Started running on Jul 3, 2023  
Platforms:   
Categories:   
This ad has multiple versions   
ID: 287378683752008  
3 ads use this creative and text

[See summary details](#)

**Get U.S Citizenship**  
Sponsored • Paid for by Get U.S Citizenship  
While USCIS's free resources are a great starting point, customized study materials can elevate your learning journey!  
Our 100 Civics Questions Bundle, personalized with video, audio, and PDF materials, not only includes specifics about YOUR representative, senators, ...  


CIVICQUESTIONS.GETUSCITIZENSHIP.COM  
Pass Your Interview On The First Try! Risk Free!  
STOP WORRYING about the U.S. Citizenship Test!  
We've got you covered with our 100 Civics Questions... [Learn More](#)

**Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements**  
IIPC Web Archiving Conference  
April 25, 2024

cr625@drexel.edu

# Out of Context

- Industry Archives: AdBeat.com



# Cost Benefit

The figure displays three vertical screenshots of the USA Today website, illustrating its layout and content.

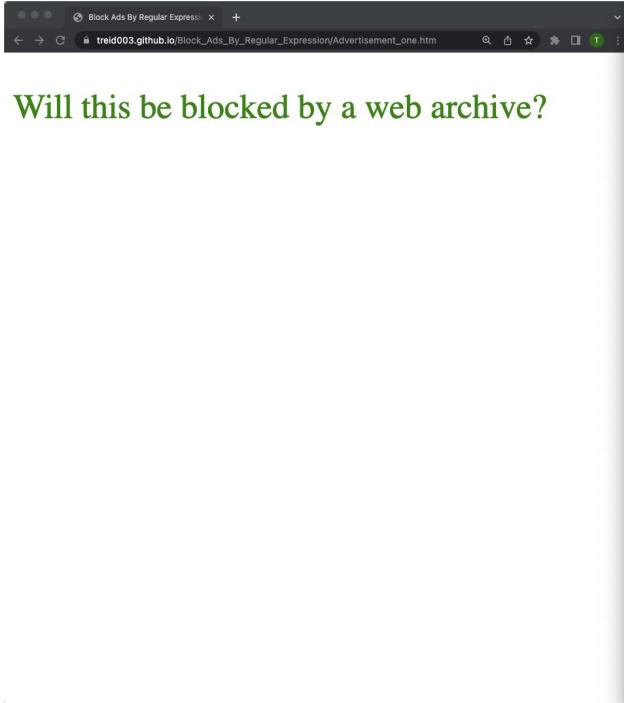
- Screenshot 1 (Left):** Shows the homepage with a prominent advertisement for "AIRKNIT X - BOXER BRIEFS". Below the ad are several news headlines and a "Featured Videos" section.
- Screenshot 2 (Middle):** Shows a full-page news article titled "NFL draft: Everything to know for tonight's first round". The page includes a large image of the NFL logo, multiple sub-headlines, and a "Top Headlines" sidebar on the right.
- Screenshot 3 (Right):** Shows a "Top Headlines" section with a grid of news cards. Each card contains a thumbnail, a headline, and a brief description. A "More Top Stories" section is visible at the bottom.

**Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements**  
**IIPC Web Archiving Conference**  
**April 25, 2024**

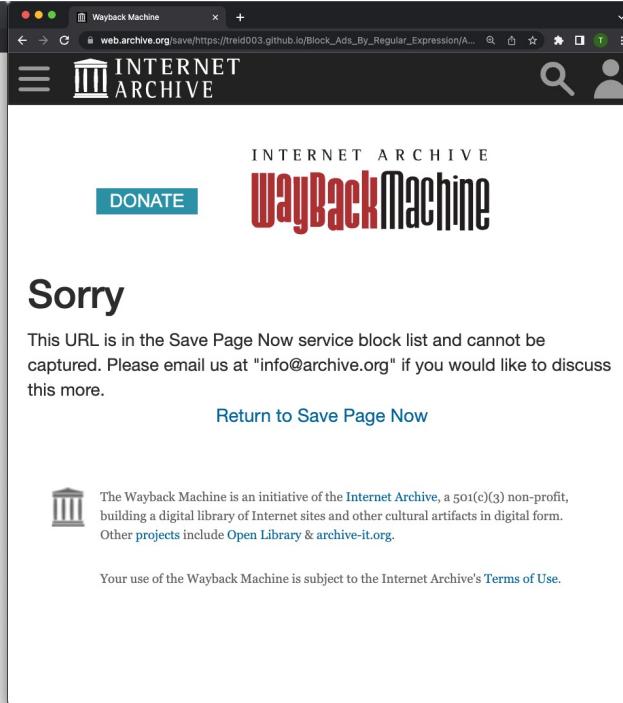
cr625@drexel.edu

# Archiving Issues

# Web Archives Blocking Ads



File name: Advertisement\_one.htm  
URI-R: [https://treid003.github.io/Block\\_Ads\\_By\\_Regular\\_Expression/Advertisement\\_one.htm](https://treid003.github.io/Block_Ads_By_Regular_Expression/Advertisement_one.htm)



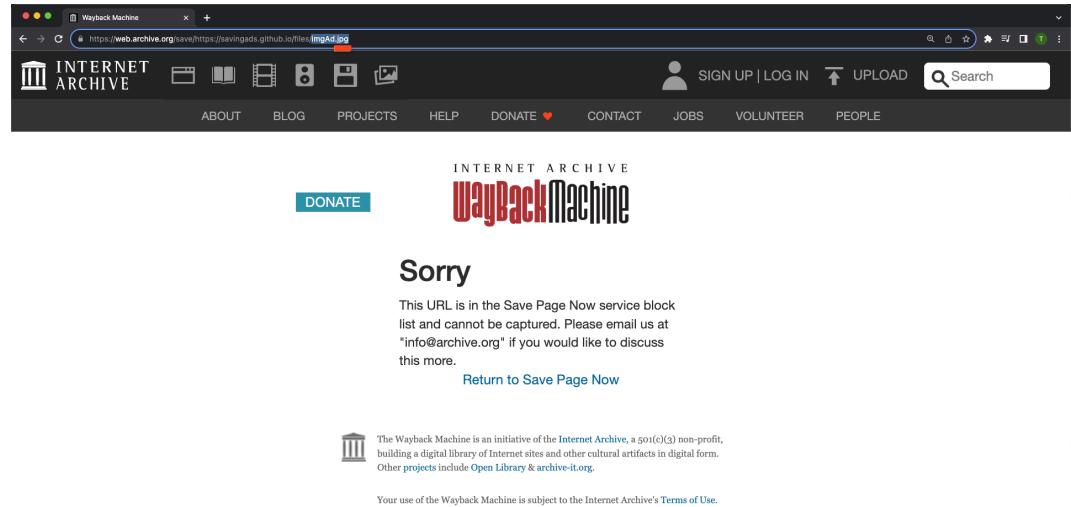
URLs that include an ad related name were blocked by Internet Archive's Save Page Now last year before August 2023

# Ad Related Filename Blocked

An ad related file name previously caused a URL to be blocked by Internet Archive's Save Page Now

- displayAds.js
- imgAd.jpg
- videoAd.mp4
- webAd.png

File name: imgAd.jpg  
URI-R: <https://web.archive.org/save/https://savingads.github.io/files/imgAd.jpg>



# Ad Related Directory Name Blocked

Example directory names that can cause a URL to be blocked:

- displayAds
- videoAd
- webAd
- ads
- Advertisement\_files

Directory name: Advertisement\_files  
URI-R: [https://savingads.github.io/Advertisement\\_files/all\\_ads\\_info.xml](https://savingads.github.io/Advertisement_files/all_ads_info.xml)

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<AllAds>
  <Ad>
    <Name>Ad One</Name>
    <Description>Description for ad 1.</Description>
  </Ad>
  <Ad>
    <Name>Ad Two</Name>
    <Description>Description for ad 2.</Description>
  </Ad>
  <Ad>
    <Name>Ad Three</Name>
    <Description>Description for ad 3.</Description>
  </Ad>
</AllAds>
```

**Sorry**  
This URL is in the Save Page Now service block list and cannot be captured. Please email us at "info@archive.org" if you would like to discuss this more.  
[Return to Save Page Now](#)

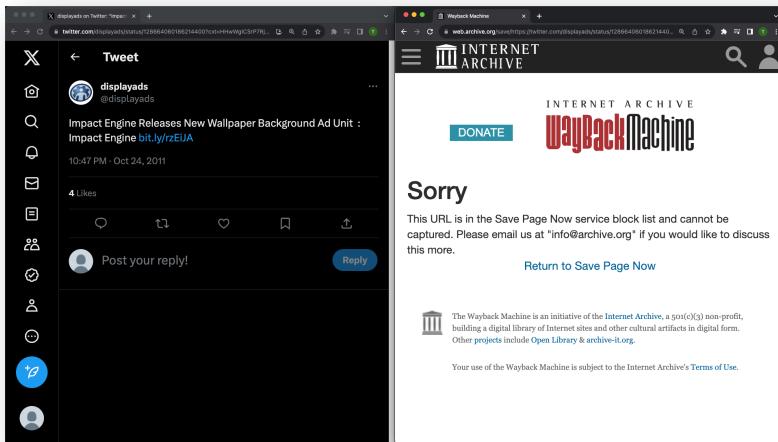
The Wayback Machine is an initiative of the Internet Archive, a 501(c)(3) non-profit, building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include Open Library & archive-it.org.

Your use of the Wayback Machine is subject to the Internet Archive's [Terms of Use](#).

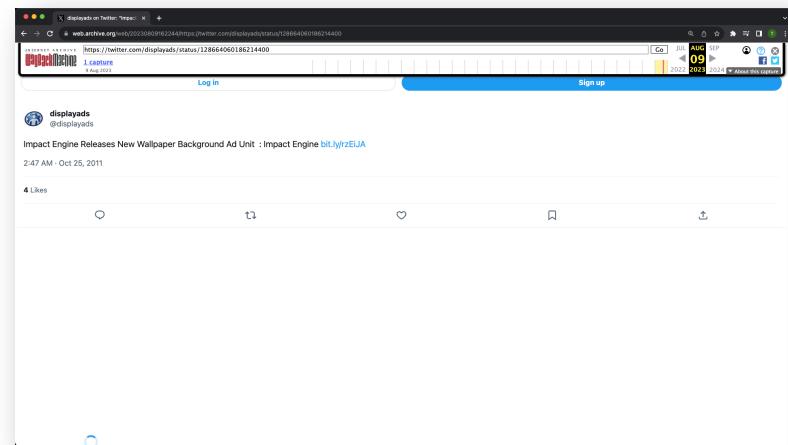
# Social Media Accounts Blocked

Social media accounts with an ad related username were also getting blocked, but after talking to Wayback Machine staff these accounts can now be archived

Account name: displayads  
Tweet blocked during July 2023



Account name: displayads  
Tweet archived during August 2023



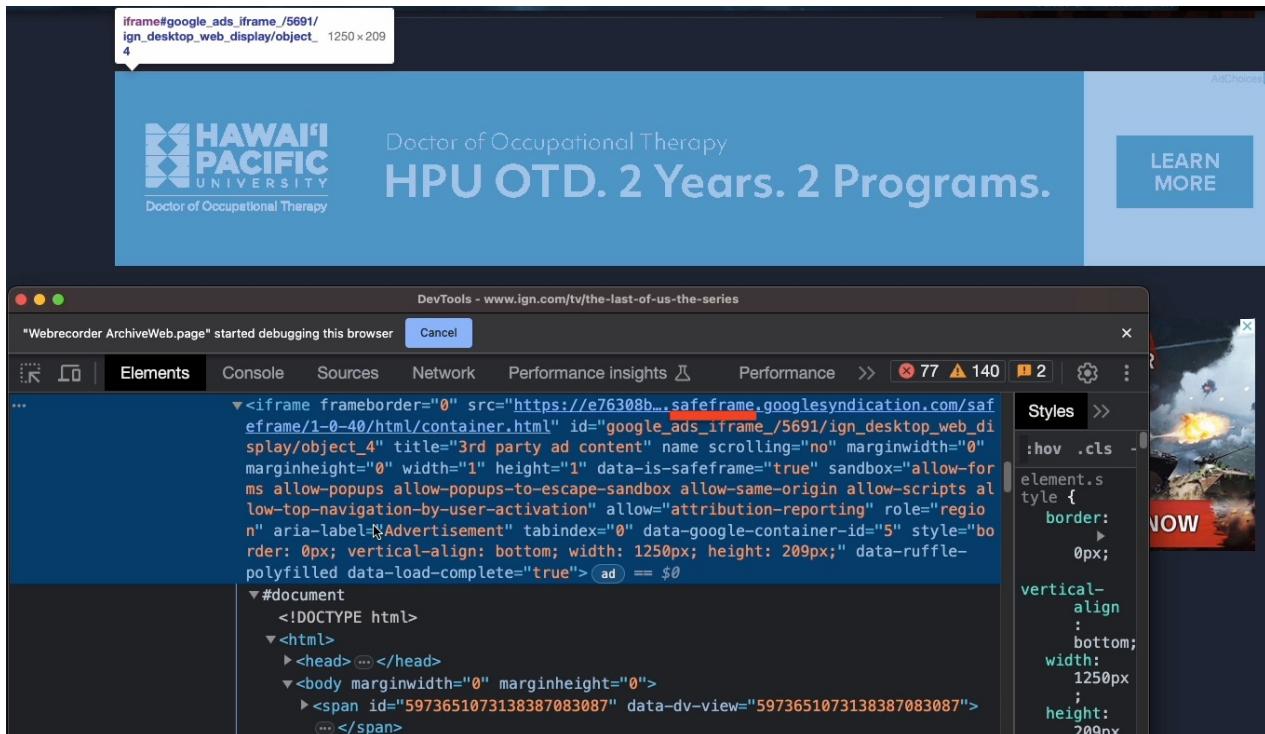
URI-R  
<https://twitter.com/displayads/status/128664060186214400?cxt=HHwWgICSp7RjckDAAAA>

**Saving Ads: Assessing and Improving Web Archives' Holdings of Online Advertisements**  
IIPC Web Archiving Conference  
April 25, 2024

cr625@drexel.edu

# Replay Issues

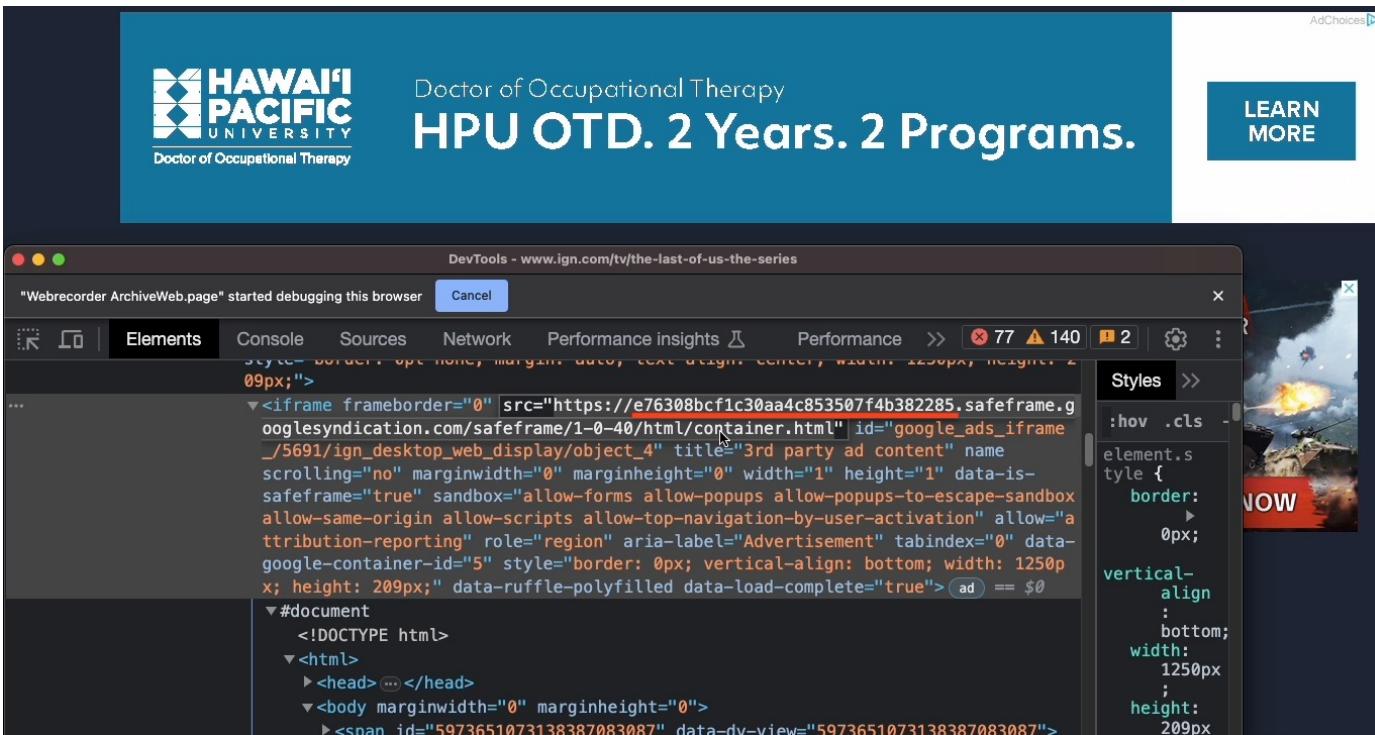
# Ad Delivery Networks and Replay (Google SafeFrame)



A Google SafeFrame is an <iframe> that is used to load web advertisements

# Google SafeFrame

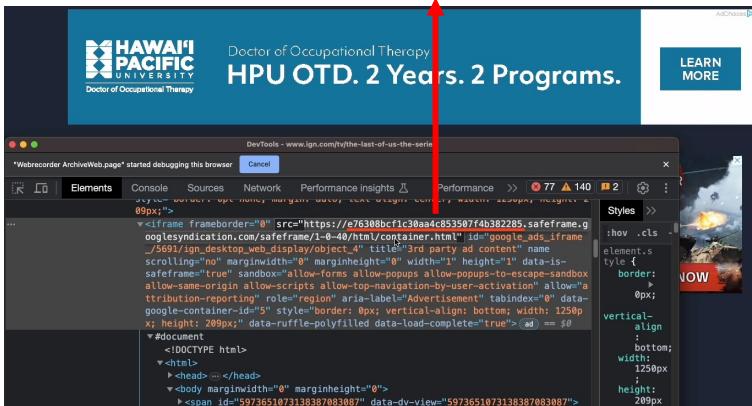
Most Google SafeFrames have a URL with a random value in the subdomain



# Google SafeFrame Random Values

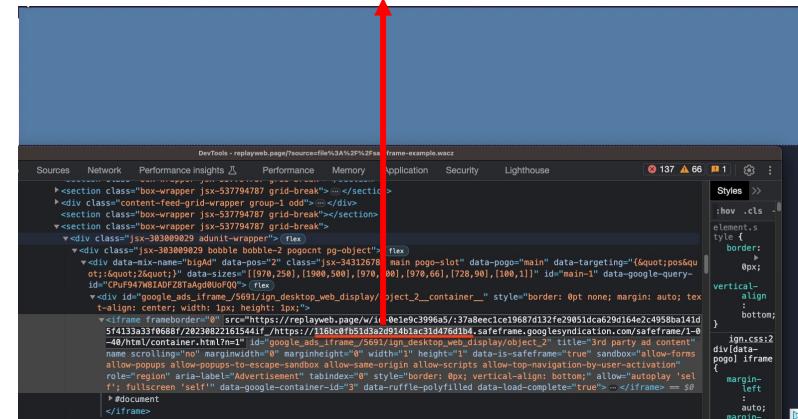
## Live Web Page

Random value in subdomain:  
e76308bcf1c30aa4c853507f4b382285



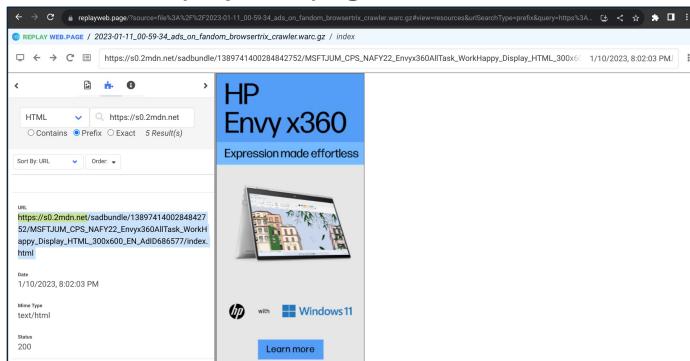
## Archived Web Page

Random value in subdomain:  
116bc0fb51d3a2d914b1ac31d476d1b4

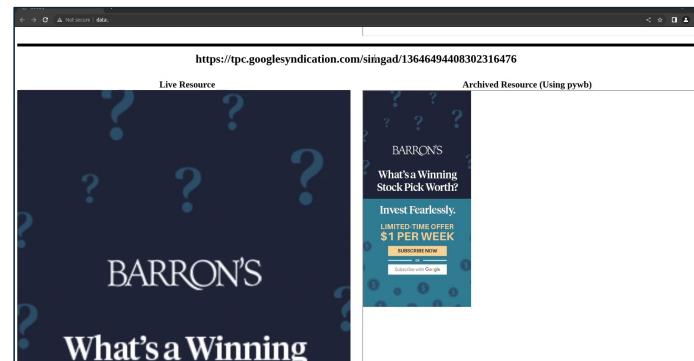


# Processes and Tools: Finding Archived Ads

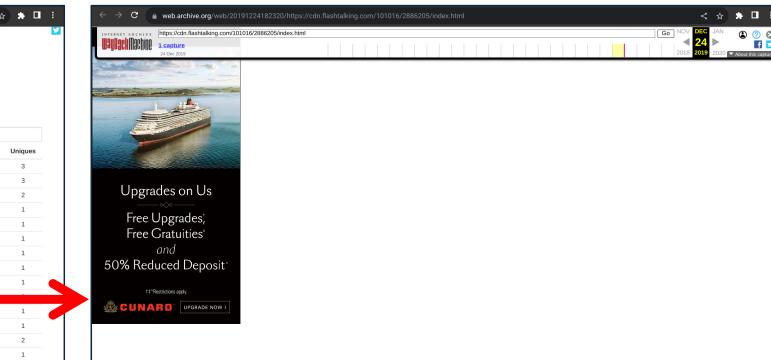
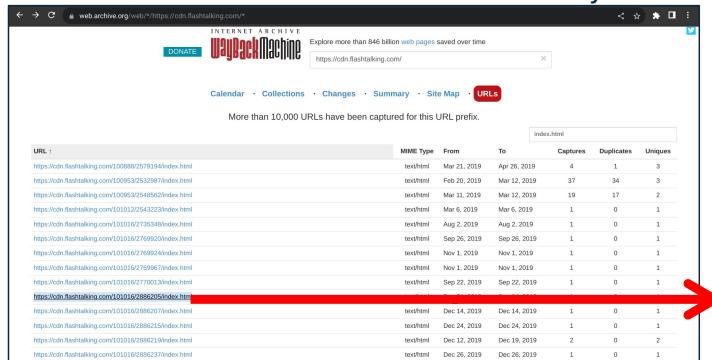
ReplayWeb.page's URL Search



Our Tool For Displaying Ads From A WARC file



Wayback Machine's URL Search



# Conclusion

<https://sites.google.com/view/treid803/home>

# Discussion/Questions

## SAVING ADS

Assessing and Improving Web Archives' Holdings of Online Advertisements

[treid003.github.io](http://treid003.github.io)

Christopher Rauch (cr625@drexel.edu), Travis Reid (treid003@odu.edu)

Advised by Profs. Mat Kelly, Alex H. Poole, Michael Nelson, Michele C. Weigle

[mkelly@drexel.edu](mailto:mkelly@drexel.edu)

[@drexel.edu](mailto:ahp56@drexel.edu)

[mln@cs.odu.edu](mailto:mln@cs.odu.edu)

[mweigle@cs.odu.edu](mailto:mweigle@cs.odu.edu)

