

A Vision for Trustworthy Web Archiving

Michael L. Nelson

[@phonedudemln.bsky.social](https://bsky.app/profile/phonedudemln.bsky.social)

with: Sawood Alam, Vikas Ashok, Mohamed Aturban, John Berlin, Justin Brunelle, Kritika Garg, Hussam Hallak, Himarsha Jayanetti, Mat Kelly, Mohan Krishna, Michele C. Weigle

[@WebSciDL.bsky.social](https://bsky.app/profile/webscidl.bsky.social)

DPC Members Forum and Networking Event - Americas

2025-04-24

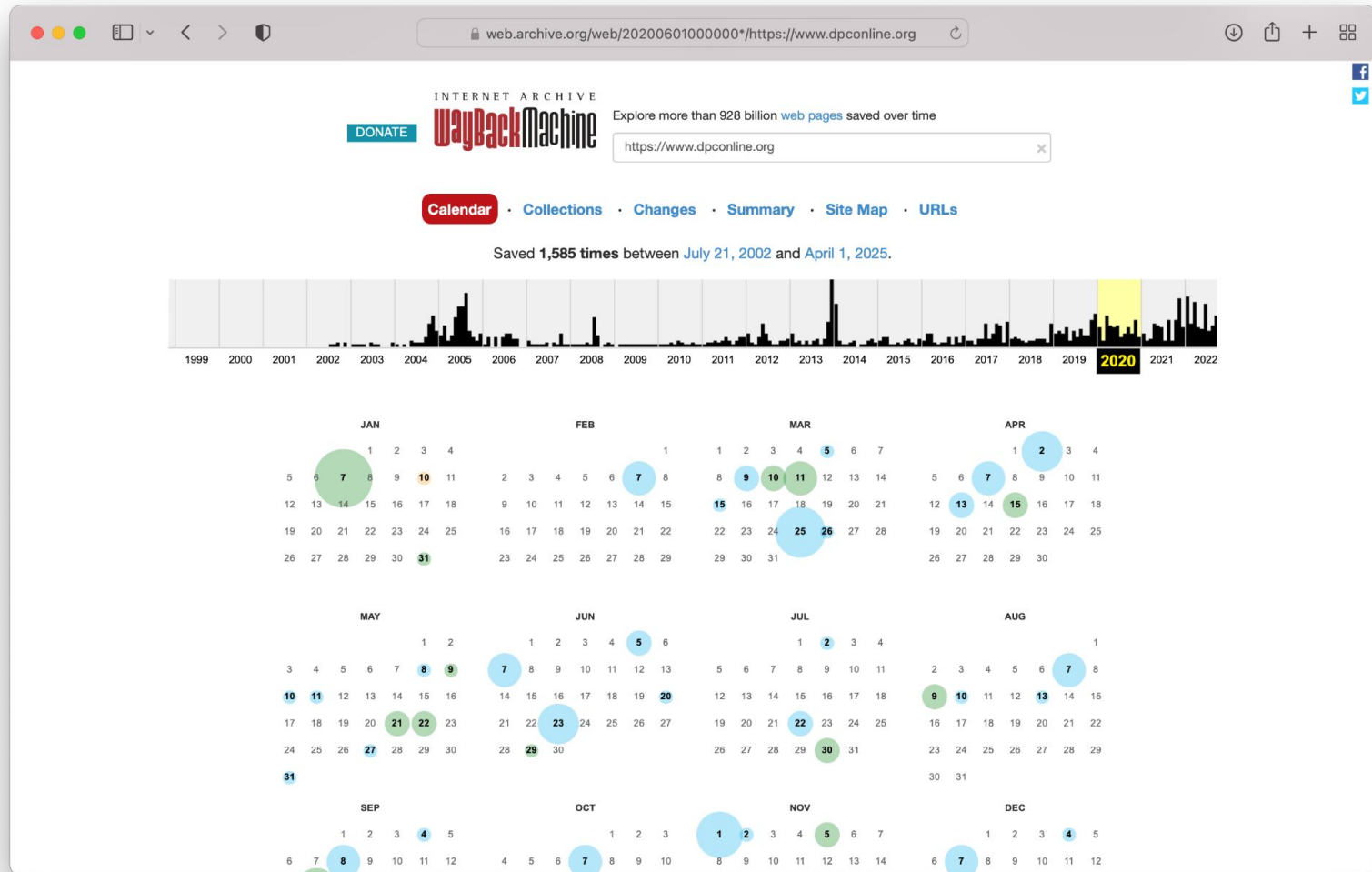
bit.ly/Nelson-DPC2025

Web archiving: experiencing the pages of the past

dpconline.org *today*

The screenshot shows the homepage of the Digital Preservation Coalition (DPC) website. The browser address bar displays 'www.dpconline.org'. The website has a green header with the DPC logo and the text 'Digital Preservation Coalition'. Navigation links include 'ABOUT', 'NEWS', 'DIGITAL PRESERVATION', 'EVENTS', and 'BLOG'. A 'Login' link is in the top right. Below the header, there is a search icon and the text 'ENHANCED BY Google'. The main content area features three columns: 1. 'Advocacy' with a background image of a woman and text: 'We work towards a climate of public and institutional policy which is better informed and better inclined towards digital preservation.' and a 'FIND OUT MORE' button. 2. 'We are building:' with a bulleted list: 'A digital preservation community which provides an effective platform for meaningful and sustained professional exchange.', 'A climate of public and institutional policy which is better informed and better inclined towards digital preservation.', 'Competent and responsive workforces that are ready to address the challenges of digital preservation.', and 'Greater maturity in digital preservation through knowledge exchange.'. 3. 'Digital Preservation Handbook' with text: 'The Handbook identifies good practice in creating, managing and preserving digital materials. By providing a strategic overview of the key issues, discussion and guidance on strategies and activities, and pointers to key projects and reports, the Handbook provides guidance for institutions and individuals and a range of tools to help them identify and take appropriate actions.'

Discovering versions of dpconline.org as archived at the Internet Archive's Wayback Machine



https://web.archive.org/web/20200601000000*/https://www.dpconline.org

Earliest version of dpconline.org: 2002-06-06

The screenshot shows a web browser window displaying the earliest version of the Digital Preservation Coalition (DPC) website. The browser's address bar shows the URL: <http://www.dpconline.org/graphics/index.html>. The Wayback Machine interface is visible at the top, indicating the capture date as 06 Jun 2002. The website header features the DPC logo and the text "Digital Preservation Coalition". The main content area is titled "Welcome to the Digital Preservation Coalition website" and provides an overview of the organization's mission and services. A vertical orange sidebar on the left contains a list of links, each marked with a question mark icon. The footer includes a "Bulldog Clip" section with a small image and a larger blue abstract graphic on the right.

Internet Archive Wayback Machine 223 captures 6 Jun 2002 - 8 Apr 2025

Text Version Members Section

Welcome to the Digital Preservation Coalition website

The Digital Preservation Coalition (DPC) was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base.

The contents of this site include:

- About the Coalition:** an introduction to our mission, programme of work, members and how to participate.
- Events:** details of past and forthcoming DPC members forums and events.
- Handbook:** the online edition of 'Preservation Management of Digital Materials - A Handbook by Maggie Jones and Neil Beagrie, now maintained and updated by the DPC.
- What's New** in digital preservation: a regular summary of selected recent activity compiled as a joint service between the DPC and Preserving Access to Digital Information (PADI) gateway.
- Links** to selected external resources and partners of the DPC.
- Contact:** how to contact the DPC

Further information and news on the Coalition will be disseminated via the digital-preservation email list on JISCmail (to subscribe to the list or view its message archive see the Web pages at <http://www.jiscmail.ac.uk/lists/digital-preservation.html>).

To get the best from this and other sites we recommend that you upgrade your browser to the latest version. [Internet Explorer 6](#), [Netscape 6.2](#), [Opera](#), [Mozilla](#). The Digital Preservation Coalition is not responsible for the content of external internet sites.

Bulldog Clip

<https://web.archive.org/web/20020606183145/http://www.dpconline.org/graphics/index.html>

DPC 2025-04-24 @phonedudemln.bsky.social, @WebSciDL.bsky.social

Earliest version of dpconline.org: 2002-06-06

The screenshot shows a browser window displaying the Wayback Machine interface for the URL <http://www.dpconline.org/graphics/index.html>. The interface includes a search bar, a calendar showing the date 06 JUN 2002, and a histogram of captures. The main content area features the Digital Preservation Coalition logo and a welcome message. A red box highlights the Wayback Machine interface at the top, and two callout boxes provide annotations.

metadata banner inserted, with archive time, total # of copies, archive histogram, etc.

all the links are rewritten to point back into the archive

Earliest version of dpconline.org: 2002-06-06

Internet Archive Wayback Machine 223 captures 6 Jun 2002 - 8 Apr 2025

http://www.dpconline.org/graphics/index.html

Members Section

Text Version

Digital Preservation Coalition

Welcome to the Digital Preservation Coalition website

The Digital Preservation Coalition (DPC) was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base.

The contents of this site include:

- About the Coalition:** an introduction to our mission, programme of work, members and how to participate.
- Events:** details of past and forthcoming DPC members forums and events.
- Handbook:** the online edition of 'Preservation Management of Digital Materials - A Handbook' by Maggie Jones and Neil Beagrie, now maintained and updated by the DPC.
- What's New** in digital preservation: a regular gateway.
- Links** to selected external resources and pages.
- Contact:** how to contact the DPC

Further information and news on the Coalition will be disseminated via the digital-preservation email list on JISCmail (to subscribe to the list or view its message archive see the Web pages at <http://www.jiscmail.ac.uk/lists/digital-preservation.html>).

To get the best from this and other sites we recommend that you upgrade your browser to the latest version. [Internet Explorer 6](#), [Netscape 6.2](#), [Opera](#), [Mozilla](#). The Digital Preservation Coalition is not responsible for the content of external internet sites.

Bulldog Clip

there are often omissions/errors at crawl time and/or replay time

<https://web.archive.org/web/20020606183145/http://www.dpconline.org/graphics/index.html>

DPC 2025-04-24 @phonedudemln.bsky.social, @WebSciDL.bsky.social

The IA's Wayback Machine is the first, biggest, etc.

INTERNET ARCHIVE

WayBackMachine

Explore more than 928 billion web pages saved over time

Enter a URL or words related to a site's home page

etsy.com
Jun 01, 2013 01:38:52

Nearly 1 Trillion web pages!

Tools

Subscription Service

Collection Search

Save Page Now

Wayback Machine Availability API

Chrome Extension

Firefox Add-on

Safari Extension

Visit Archive-It to build

This service is based on indexes

<https://web.archive.org>



dpconline.org as archived at loc.gov

The screenshot shows the Internet Archive interface for the URL <https://www.dpconline.org/>. The page title is "LIBRARY WEB ARCHIVES" with the Library of Congress logo. A search bar contains the text "Search for a URL in the web archive". Below the search bar, it states "Saved 37 times between May 10, 2011 and August 12, 2021." A bar chart shows the frequency of saves over time, with a prominent peak in 2018. Below the chart is a calendar grid for the years 2011 through 2022. The calendar for 2018 is expanded, showing months from January to August. In the April 2018 calendar, the 4th, 7th, and 22nd are circled in blue. The text "OpenWayback 2.3.0" is visible in the bottom right corner of the browser window.

https://webarchive.loc.gov/all*/https://www.dpconline.org/

dpconline.org as archived at Arquivo.pt

ARQUIVO.PT

dpconline.org

1991 6 Ago

2025 23 Abr

Páginas Imagens Narrativa

Pesquisar

Pesquisa avançada

Cerca de 53 resultados desde 1991 até 2025

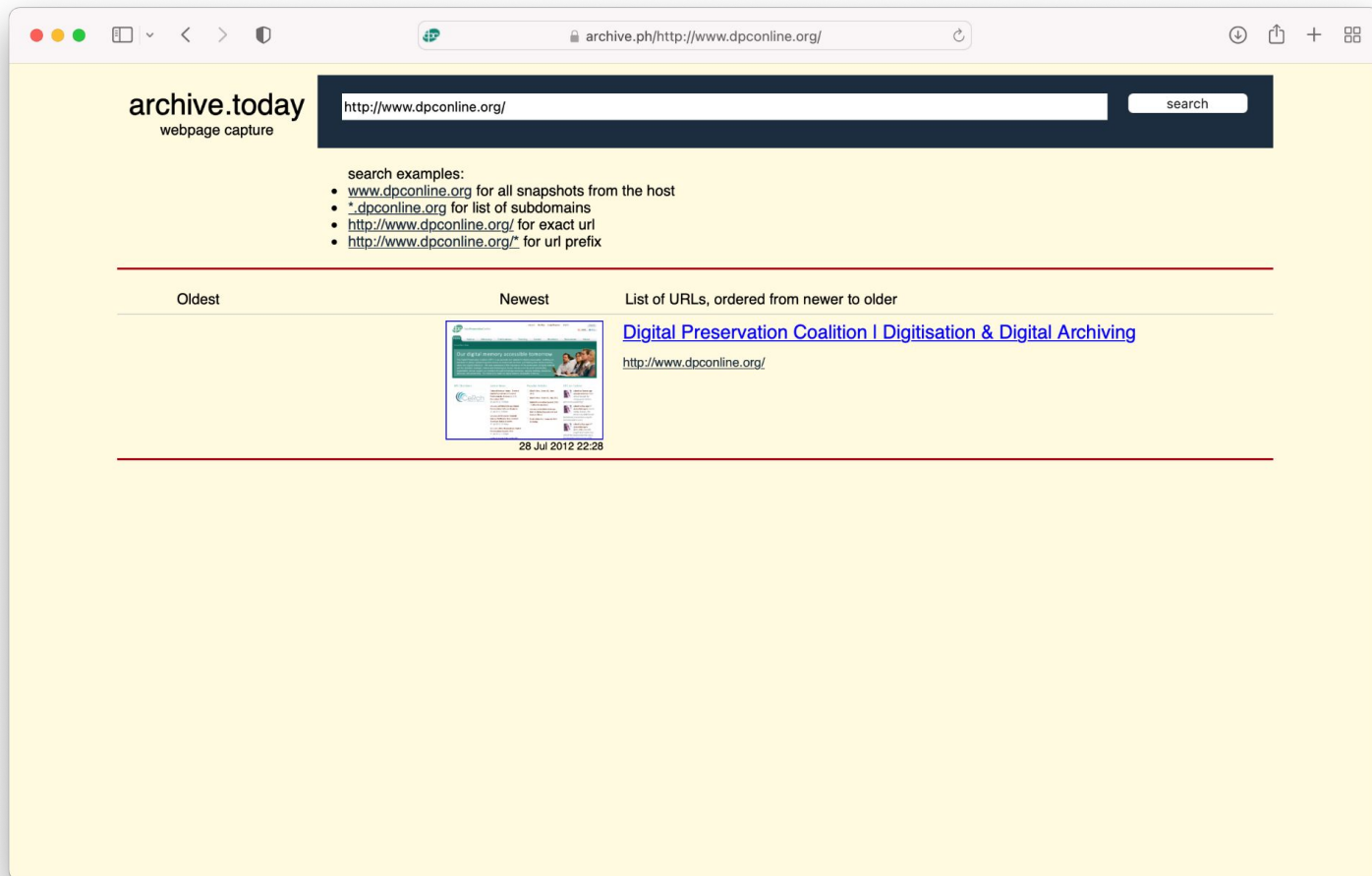
Tabela Lista

2014	1 versão
2016	3 versões
2017	1 versão
2018	2 versões
2019	7 versões
2020	2 versões
2021	9 versões
2022	12 versões
2023	5 versões

Run script "void(0)"

https://arquivo.pt/url/search?q=dpconline.org&l=pt&from=19910806&to=20250423&trackingId=2d0d3b5a7978c74a46dd_32b0980508ceac15139a&adv_and=dpconline.org

dpconline.org as archived at archive.today

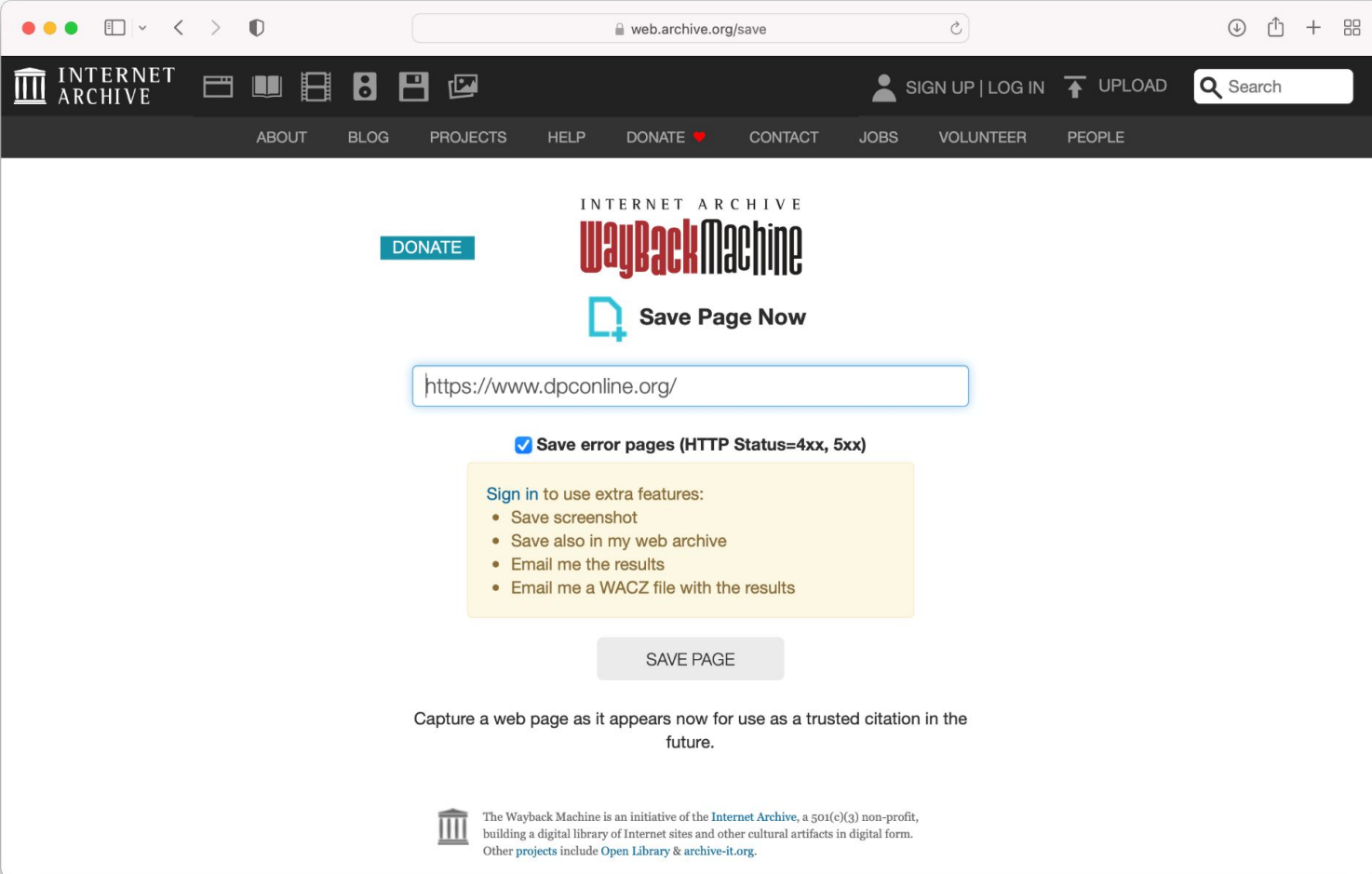


<https://archive.ph/http://www.dpconline.org/>

Web archiving is not just a spectator sport.

Archival frequency roughly correlates with popularity.
It's up to you to archive your favorite page.

You can request a copy to be made



The screenshot shows the Wayback Machine interface for saving a page. The browser address bar shows `web.archive.org/save`. The page header includes the Internet Archive logo and navigation links: ABOUT, BLOG, PROJECTS, HELP, DONATE, CONTACT, JOBS, VOLUNTEER, PEOPLE. A search bar is also present.

The main content area features the Wayback Machine logo and a "Save Page Now" button. Below this is a text input field containing the URL `https://www.dpconline.org/`. A checkbox labeled "Save error pages (HTTP Status=4xx, 5xx)" is checked. A yellow box contains the text "Sign in to use extra features:" followed by a list of options: "Save screenshot", "Save also in my web archive", "Email me the results", and "Email me a WACZ file with the results". A "SAVE PAGE" button is located below the input field.

At the bottom, a message reads: "Capture a web page as it appears now for use as a trusted citation in the future." Below this is the Internet Archive logo and a paragraph of text: "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."

A new copy of dpconline.org!

The screenshot shows the Internet Archive Wayback Machine interface. At the top, the browser address bar displays `web.archive.org/save/https://www.dpconline.org/`. The Internet Archive logo and navigation menu are visible. The main heading reads "Saving page https://www.dpconline.org/" with a green "Done!" button. Below this, a message states "A snapshot was captured. Visit page: /web/20250423033342/https://www.dpconline.org/". A list of captured elements is shown in a box, including various CSS and media files from the dpconline.b-cdn.net domain. At the bottom, there is a footer with the Internet Archive logo and text: "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."

<https://web.archive.org/save/https://www.dpconline.org/>
Similar services available at perma.cc, archive.today, etc.

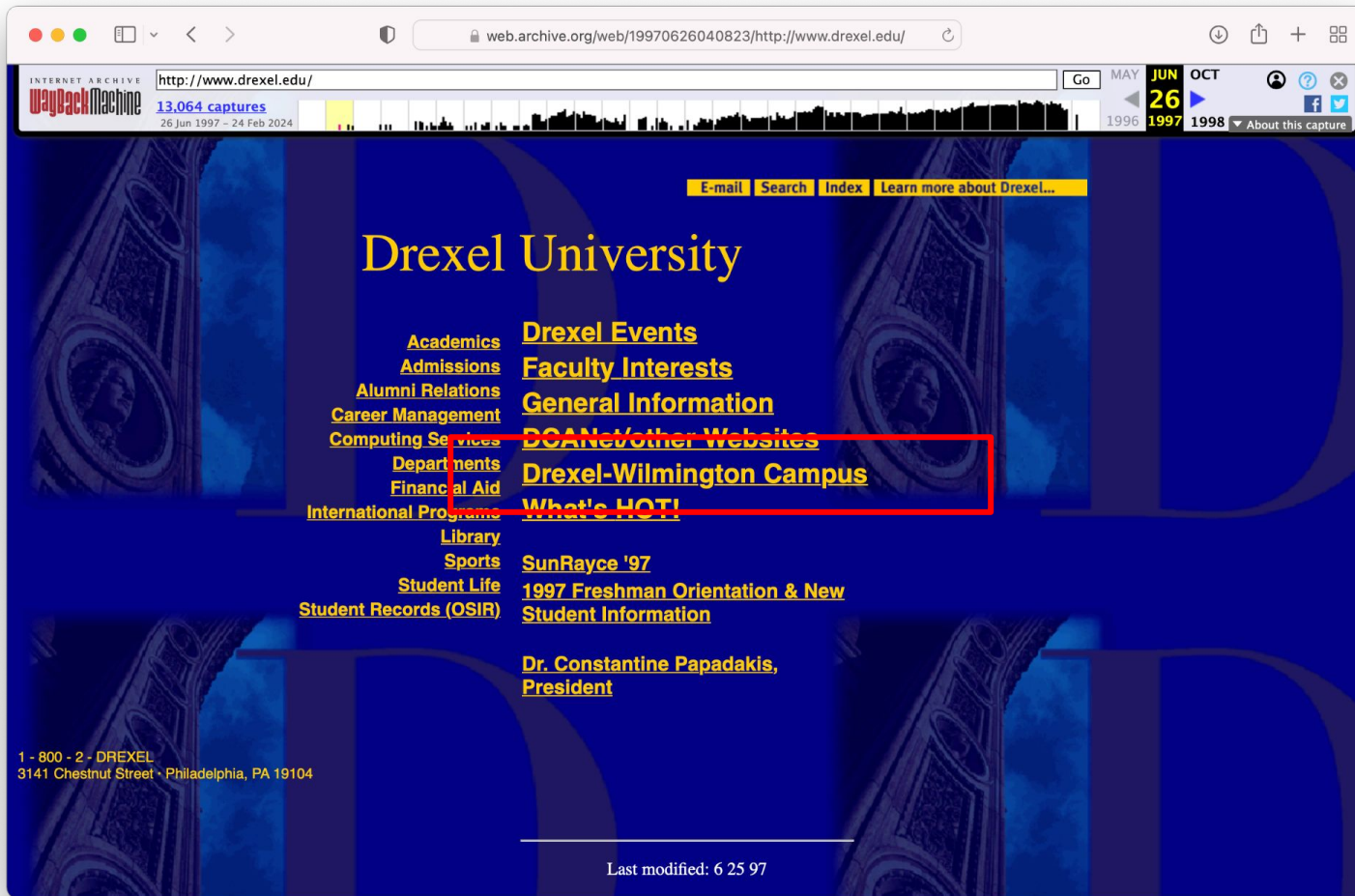
DPC 2025-04-24 @phonedudemln.bsky.social, @WebSciDL.bsky.social

Why do we archive the web?

"Is there any statistics to show that many or a good number of Web users would like to get obsolete data or resources? "

–Anonymous reviewer of our rejected submission
about web archiving to WWW 2009

Because page replay *reveals forgotten knowledge*



<https://web.archive.org/web/19970626040823/http://www.drexel.edu/>

Because the "Drexel Wilmington Campus" has all but disappeared from the live web

INTERNET ARCHIVE
Wayback Machine
13 captures
26 Jun 1997 - 24 May 2001

web.archive.org/web/19970626041147/http://www.drexel.edu/DxIWilm/

http://www.drexel.edu/DxIWilm/DrexelWilm.html

Go MAY JUN 26 OCT 1996 1997 1998 About this capture

DREXEL UNIVERSITY HOME | INDEX | E-MAIL | SEARCH

WILMINGTON CAMPUS

Drexel University is proud to announce its new off-main-campus location in Wilmington, Delaware. Course offerings at the new Wilmington, Delaware campus will begin Spring term (March 31, 1997 -June 14, 1997).

Call Drexel University - Wilmington for more information at 658-7500.
To register for the course(s) offered at our new Wilmington campus, please call 1-800-2-DREXEL

All applicants will be accepted as non-matriculated students while their credentials are being evaluated. Tuition is \$576.00 per course.

Course Offerings:

<p>Financial Accounting Foundations (3 CREDITS)</p> <p>An introduction to the preparation of the income statement and balance sheet, plus analysis and recording of business transactions and basic accounting skills.</p> <p>Wednesday 5:30-7:45 p.m.</p>	<p>PC Competency (3 CREDITS)</p> <p>A hands-on certificate program for anyone interested in developing skills in personal computing; with either a Windows or DOS-based course of study.</p> <p>Monday 5:30-7:45 p.m.</p>
	<p>Introduction to Computing (3 CREDITS)</p>

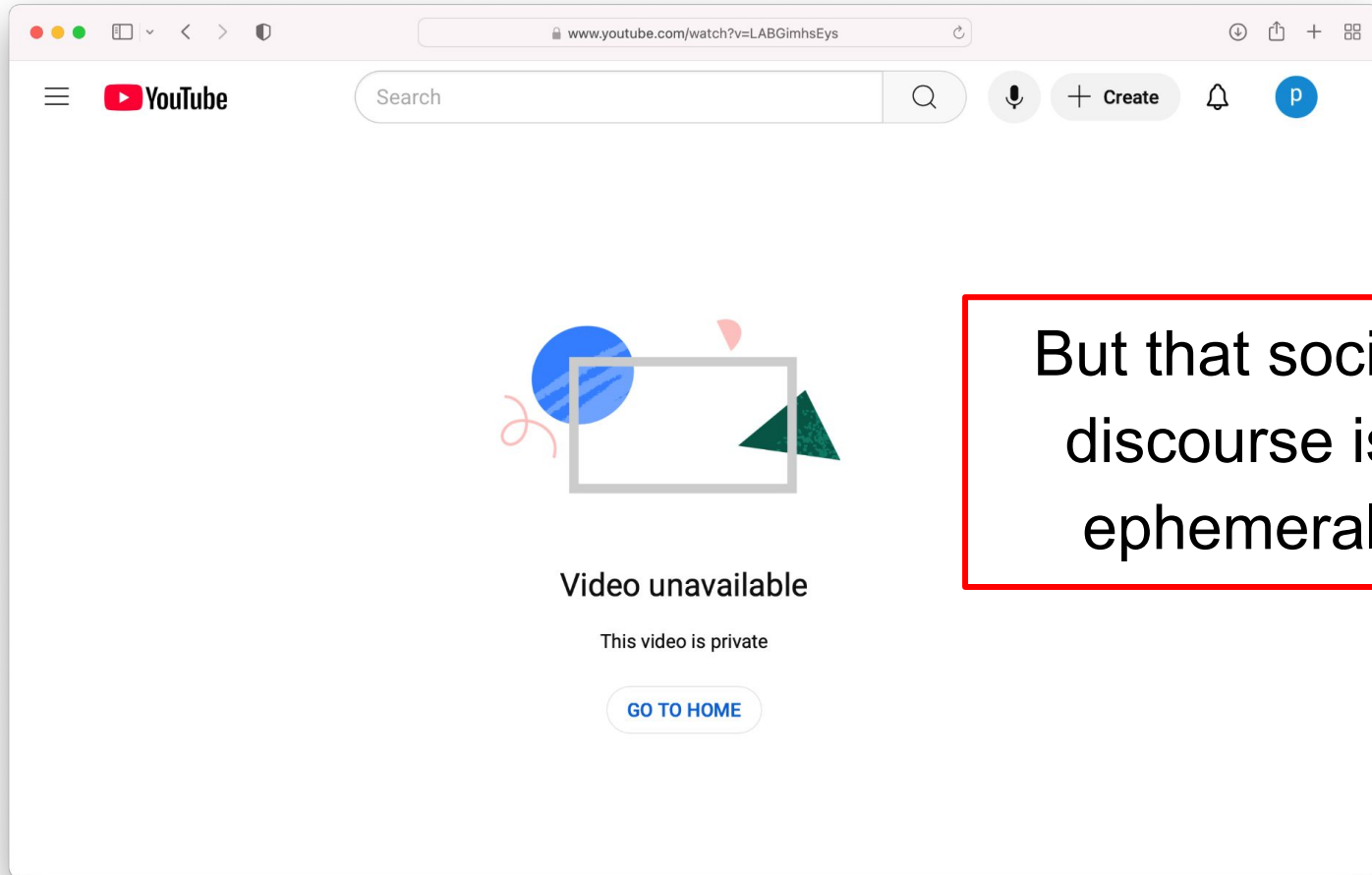
<https://web.archive.org/web/19970626041147/http://www.drexel.edu/DxIWilm/DrexelWilm.html>

Because our social discourse is *dominated by the web*



<https://www.youtube.com/watch?v=LABGimhsEys>

~~Because our social discourse
is dominated by the web~~

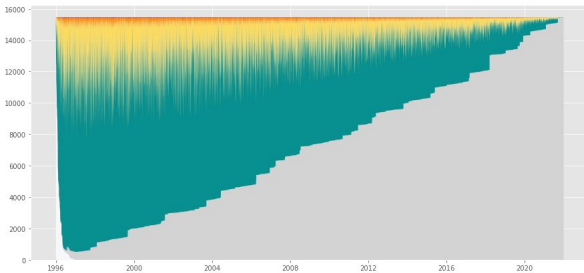


But that social
discourse is
ephemeral

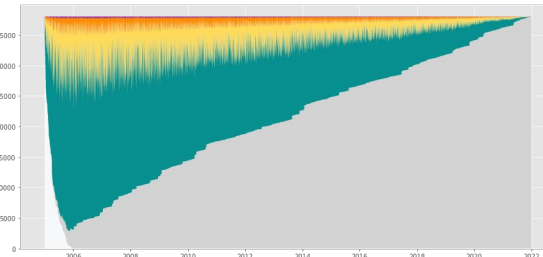
<https://www.youtube.com/watch?v=LABGimhsEys>

Because the pages that carry that discourse *decay linearly*

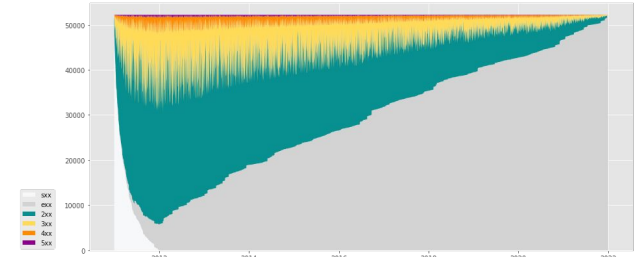
1996 Root URLs



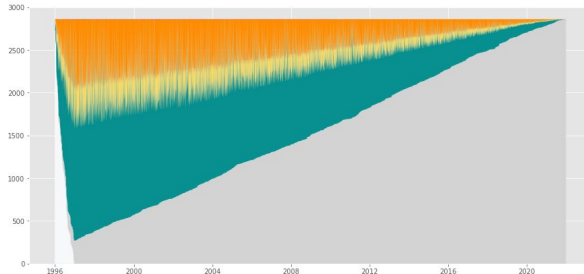
2005 Root URLs



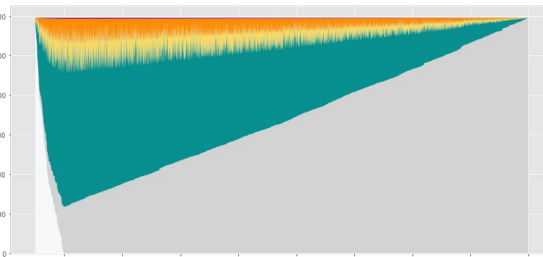
2011 Root URLs



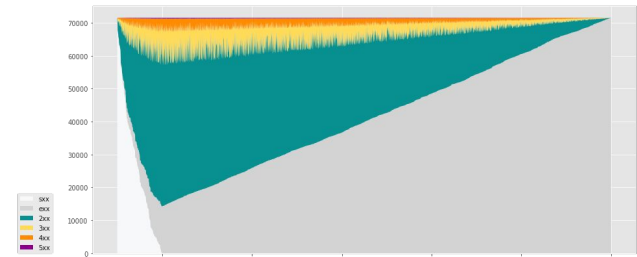
1996 Deep Links



2005 Deep Links



2011 Deep Links



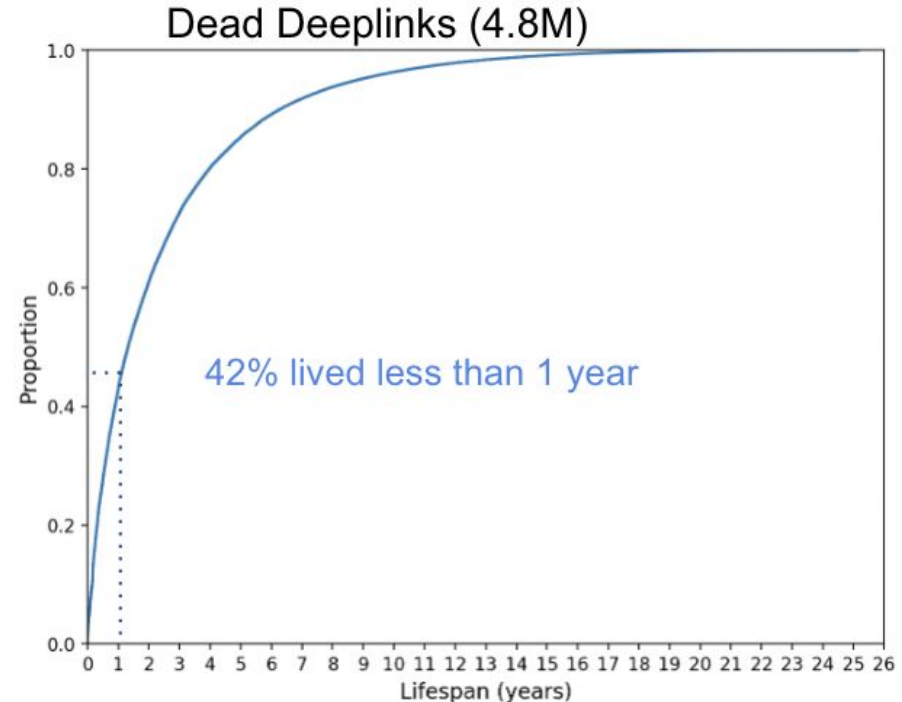
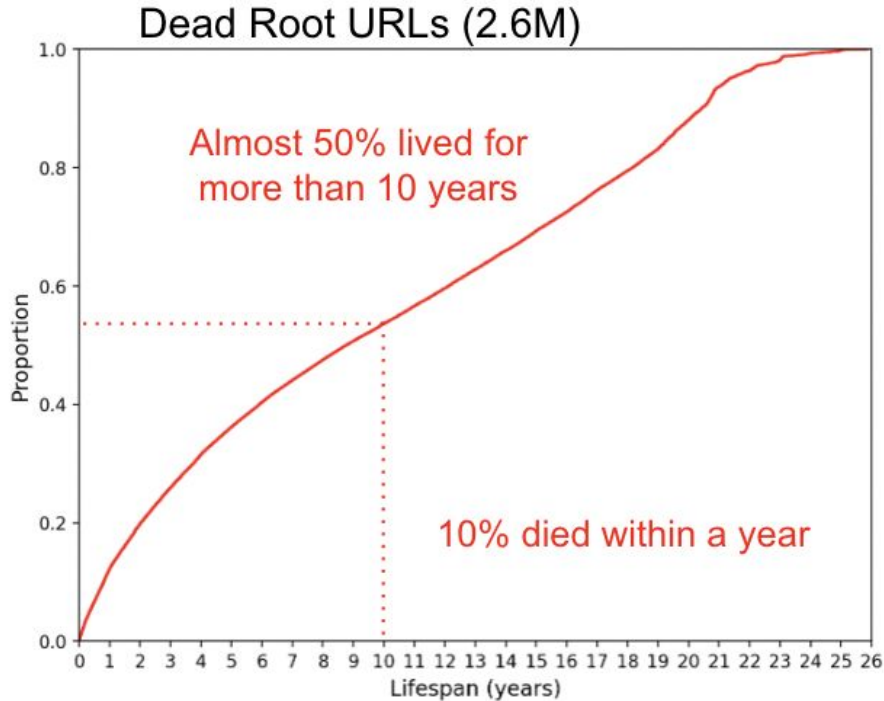
x-axis: time, early (left) → now (right), y-axis: URLs discovered by IA in 1996/2005/2011

Green = 200 OK, Yellow = 30X redirection, Orange = 404 Not Found

Data taken from "[Not your parents' web...](#)" project

See also: my Twitter threads for more citations on [linear decay](#) of web resources.

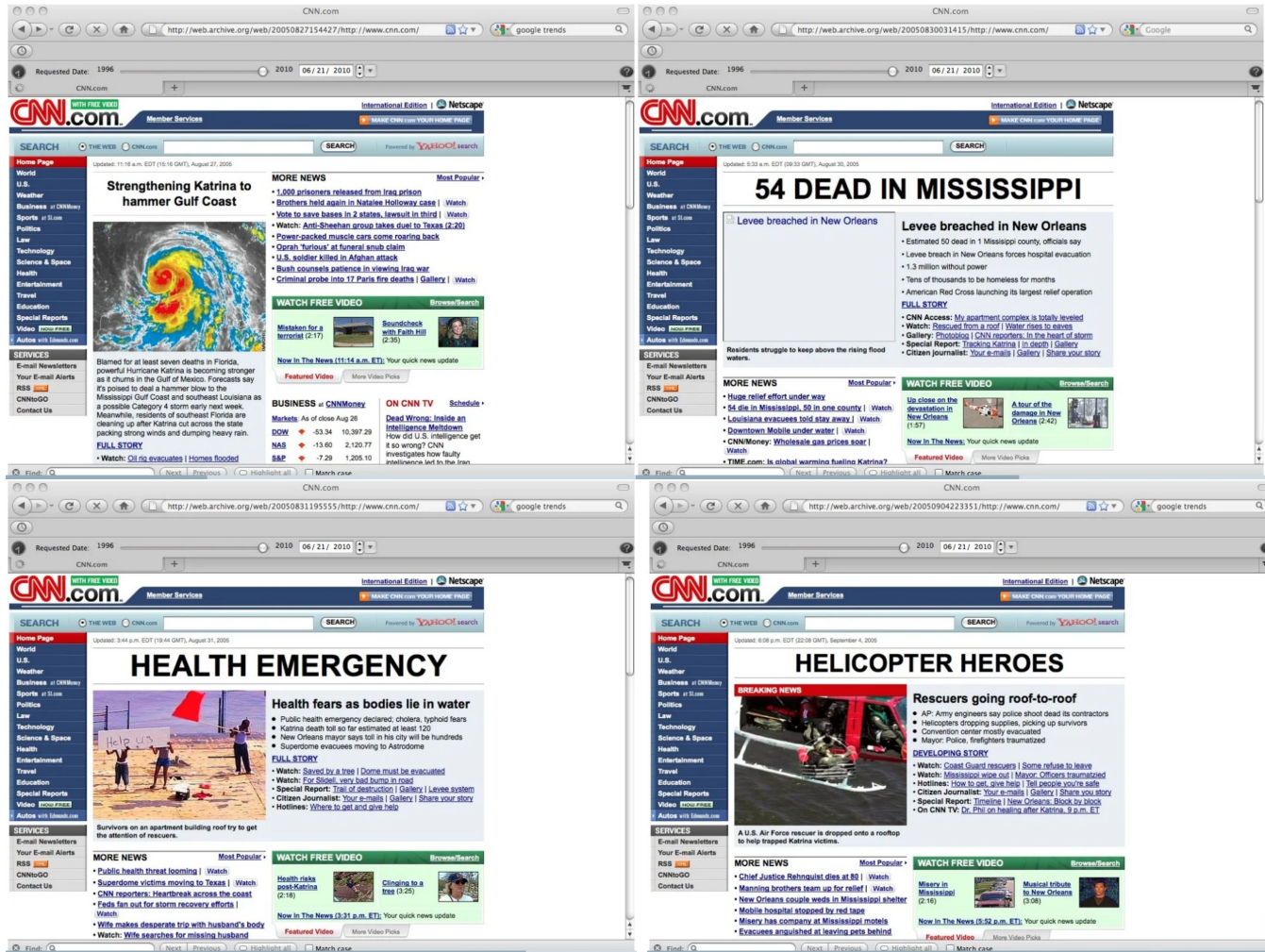
Because the median lifespan of a URL = 2.3 years



example root = www.dpconline.org
example deep link = www.dpconline.org/events/previous-events

Data taken from "[Not your parents' web...](#)" project

Because replay of contemporary pages conveys history better than a summary page



"Why Care About the Past?"

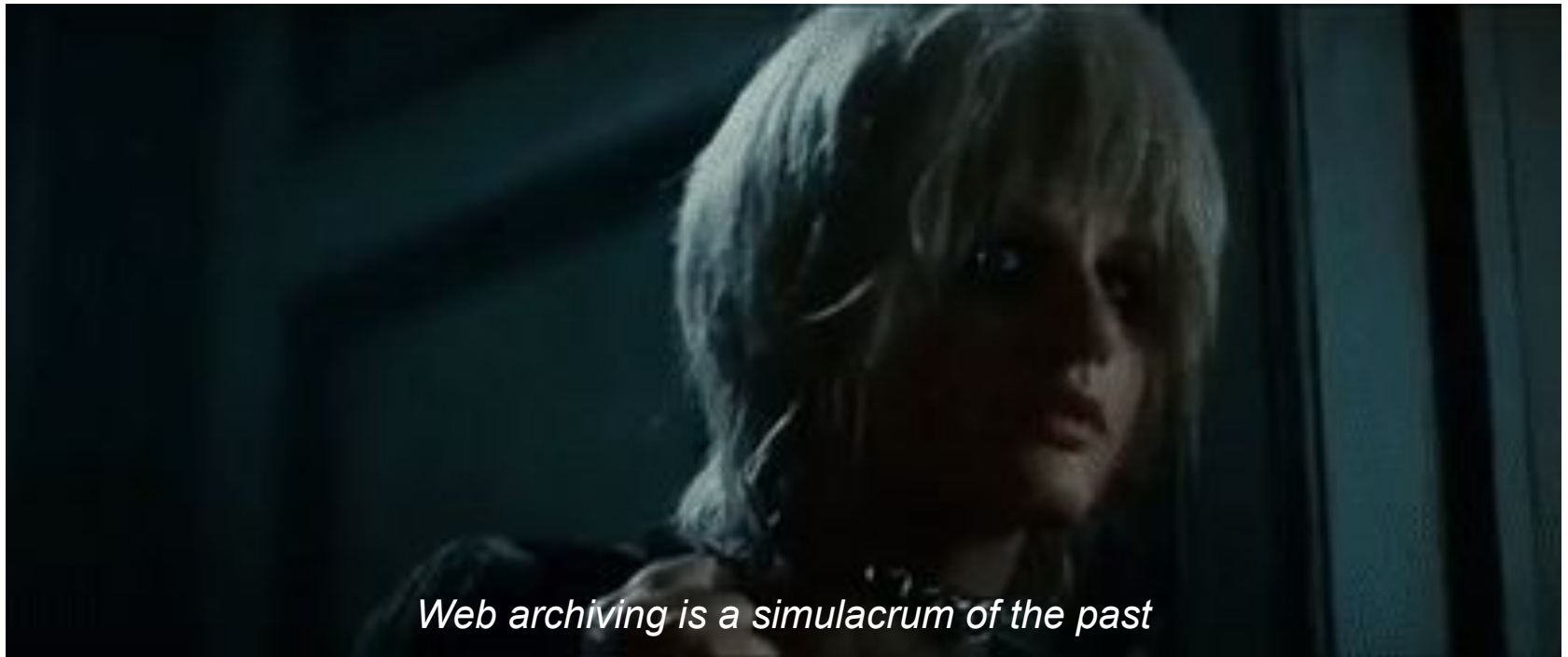
"Examining the News Media's Treatment of COVID-19 Death Milestones"

Now we know a little bit about web archives
and why they're useful.

Web archiving is not file backup

Backup = prevent, detect, repair changes

Web archiving = continuous change to better simulate the past

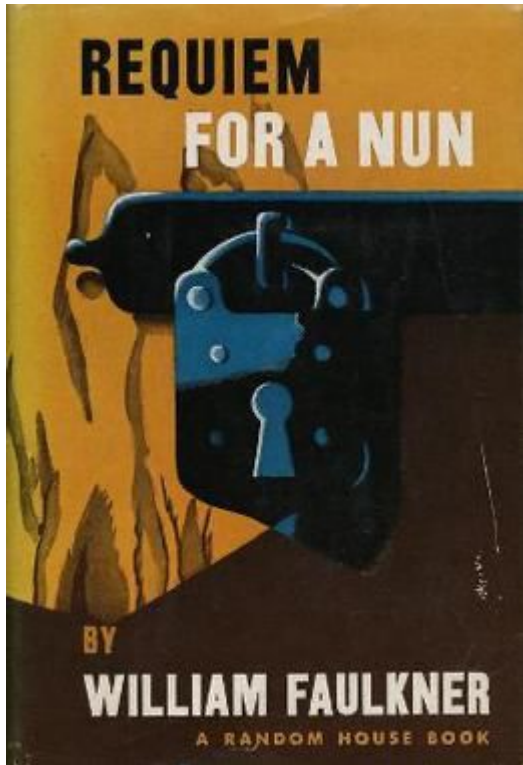


Web archiving is a simulacrum of the past

<https://makeagif.com/gif/blade-runner-iff-sebastians-toys-kaiser-and-bear-AFkWpp>

<https://ws-dl.blogspot.com/2020/03/2020-03-07-at-nexus-of-cni-keynote-and.html>

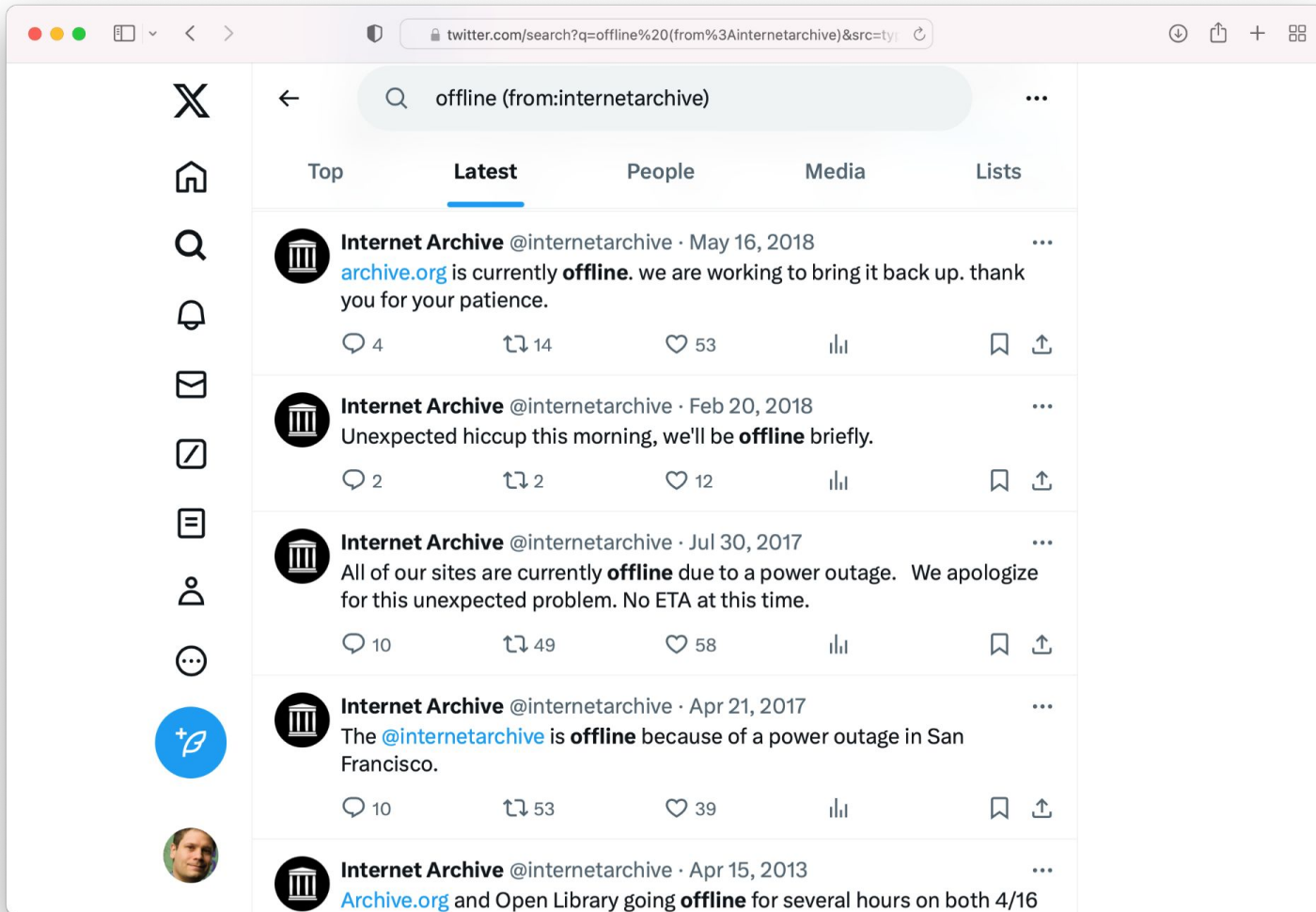
"The Past Web" is a myth – it's all live web



"The past is never dead.
It's not even past."

As such, web archives are subject to the same threats – perhaps even more so – as the live web.

IA is not resourced like



[https://twitter.com/search?q=offline%20\(from%3Ainternetarchive\)&src=typed_query&f=live](https://twitter.com/search?q=offline%20(from%3Ainternetarchive)&src=typed_query&f=live)

The crawlable web as collateral damage in pursuit of training data



Helen Hockx-Yu
@hhockx



How long will it take before crawlers can no longer freely collect content off the web? And how long for AI companies to approach those PB-scale digital archives/collections?



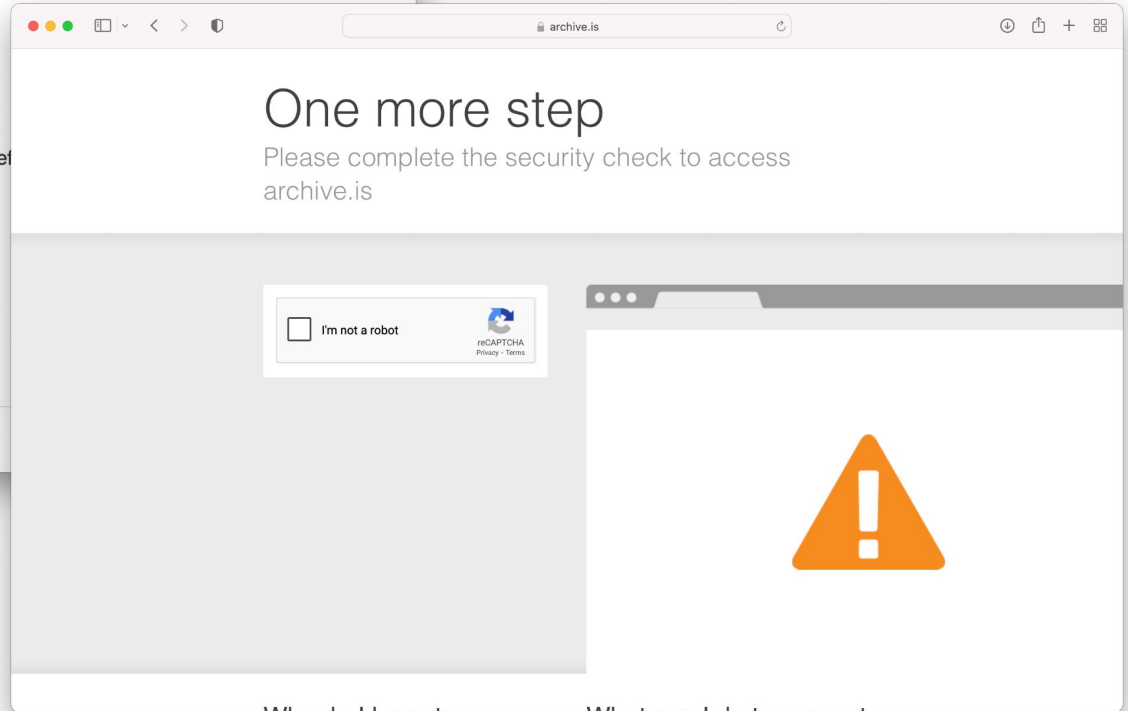
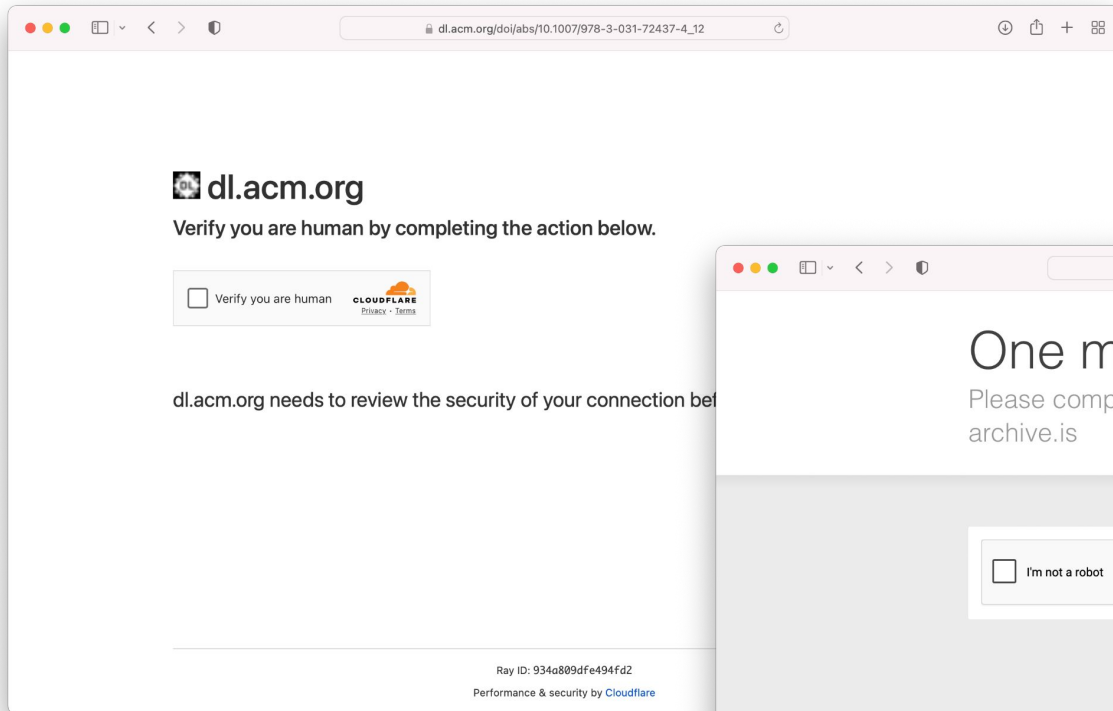
From theverge.com

7:37 AM · Feb 19, 2024 · **227** Views

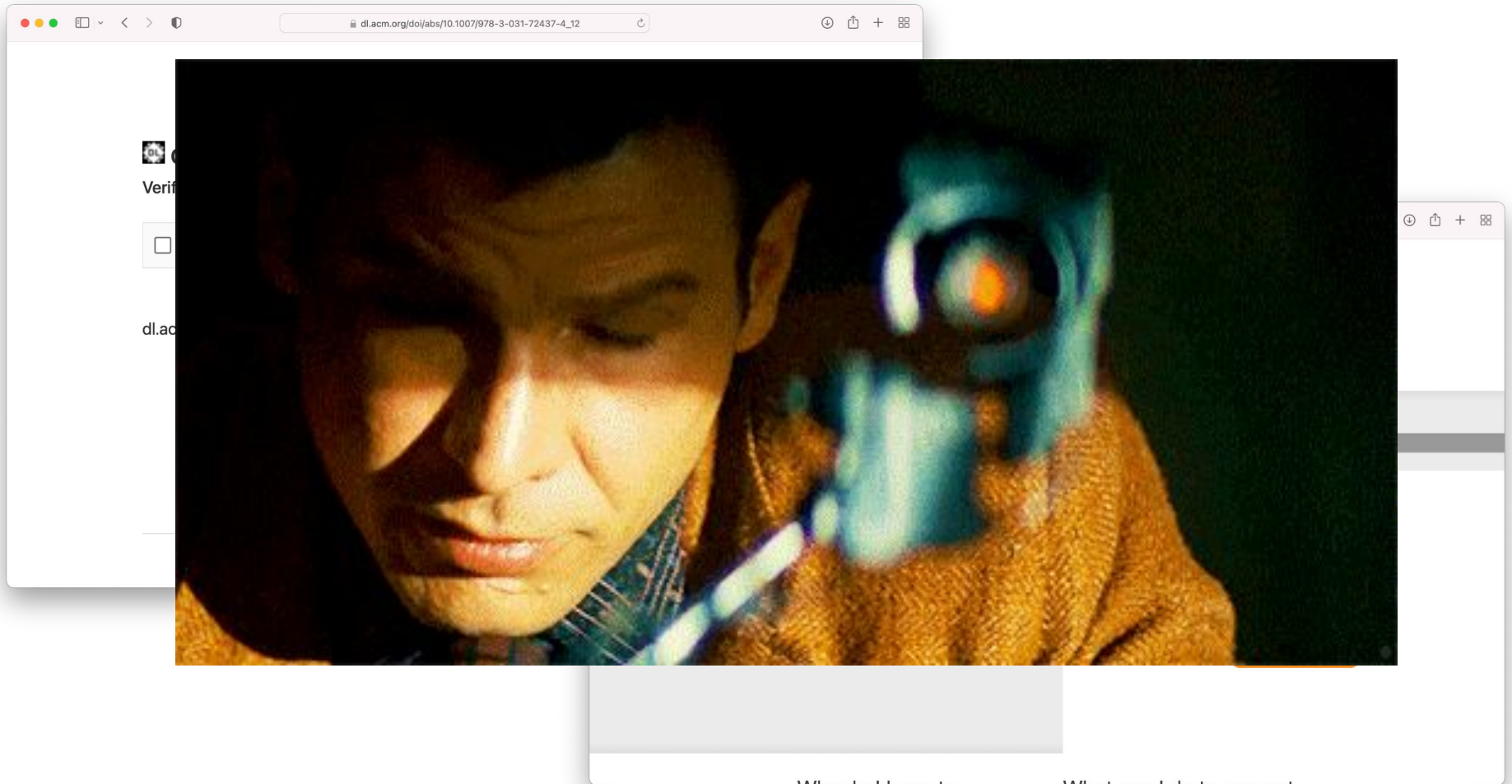
<https://twitter.com/hhockx/status/1759557890388418712>

DPC 2025-04-24 @phonedudemln.bsky.social, @WebSciDL.bsky.social

Robots are second class citizens



Robots are second class citizens



https://bladerunner.fandom.com/wiki/Voight-Kampff_test

Publishers are persistent legal threats

↻ You reposted



Internet Archive
@internetarchive



J Jacobin ✓ @jacobin · 23h

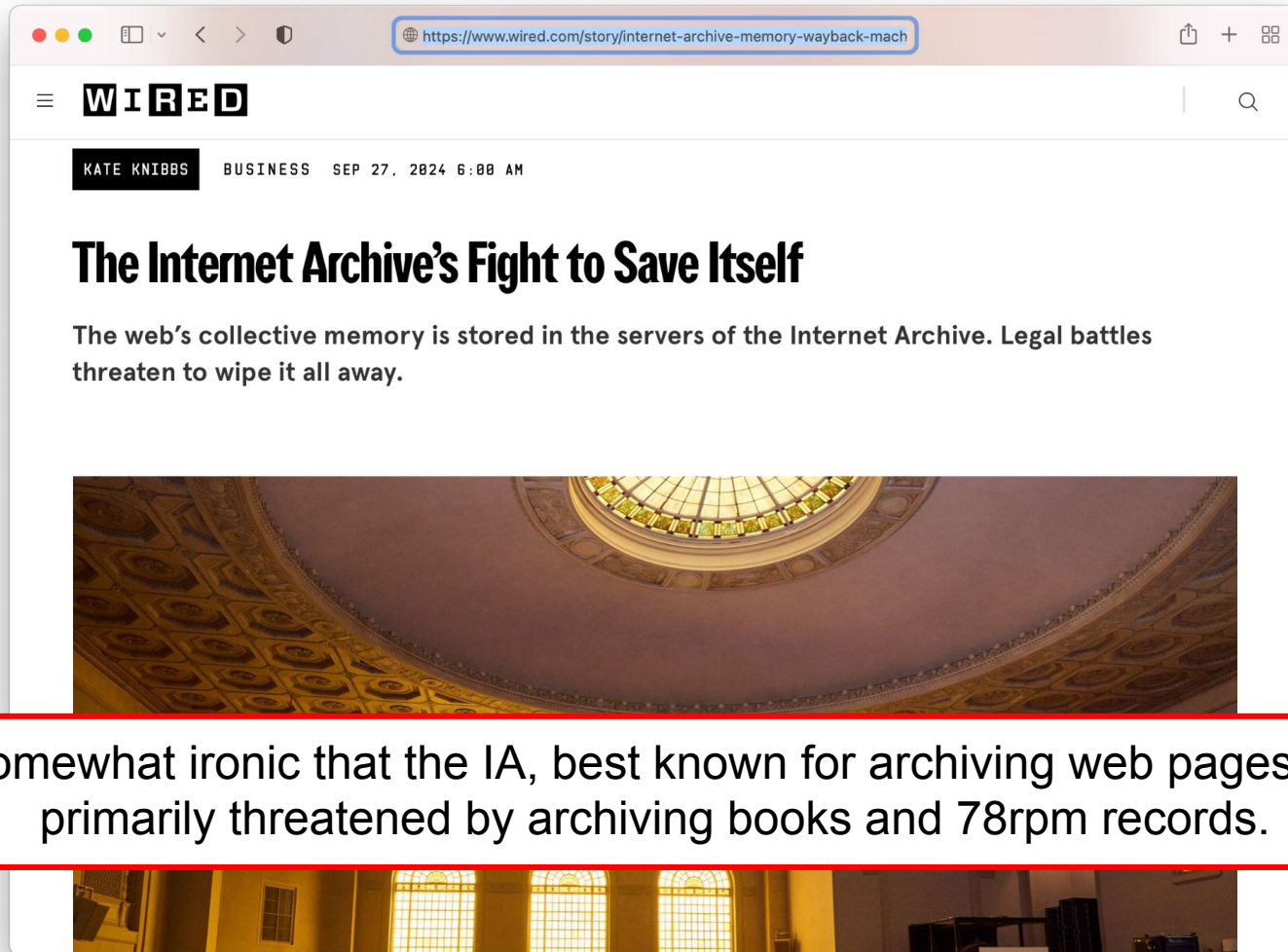
The Internet Archive, a digital library aiming to make published materials accessible to everyone, is in danger of being shut down by four major publishers. If the publishers are successful, free access to knowledge will be in jeopardy. jacobin.com/2024/02/intern...

10:17 AM · Feb 23, 2024 · **94.1K** Views



<https://twitter.com/internetarchive/status/1761047683957948500>

2025 update: the publishers are winning



Somewhat ironic that the IA, best known for archiving web pages, is primarily threatened by archiving books and 78rpm records.

<https://www.wired.com/story/internet-archive-memory-wayback-machine-lawsuits/>

Hacking / ransomware



British Library 
@britishlibrary

...

We are still experiencing technology issues and we expect this to continue for the next few days.

Our website and a number of on site services remain down. Our Reading Rooms will be closed as usual on Sunday and will reopen for personal study on Monday, but collection items and public WiFi may not be available.

We advise researchers who need to view items not to travel to the Library until our collection services are resumed.

Our Fantasy exhibition and public spaces will be open and our public events will be going ahead as planned, but we are only able to take cash payment on site.

We're very sorry for the inconvenience and will update as soon as we can.



<https://twitter.com/britishlibrary/status/1718317423923781849>



UK Web Archive
@UKWebArchive

...

The latest update from the UK Web Archive includes links to alternative web archives and resources while our website is still down:
blogs.bl.uk/webarchive/202...

[#WebArchiveWednesday](#) [#WebArchiving](#) [#WebArchiveResearch](#)
[#UKLegalDeposit](#)



 Bodleian Libraries and 5 others

7:43 AM · Jan 24, 2024 · **744** Views

<https://twitter.com/UKWebArchive/status/1750137173385339211>

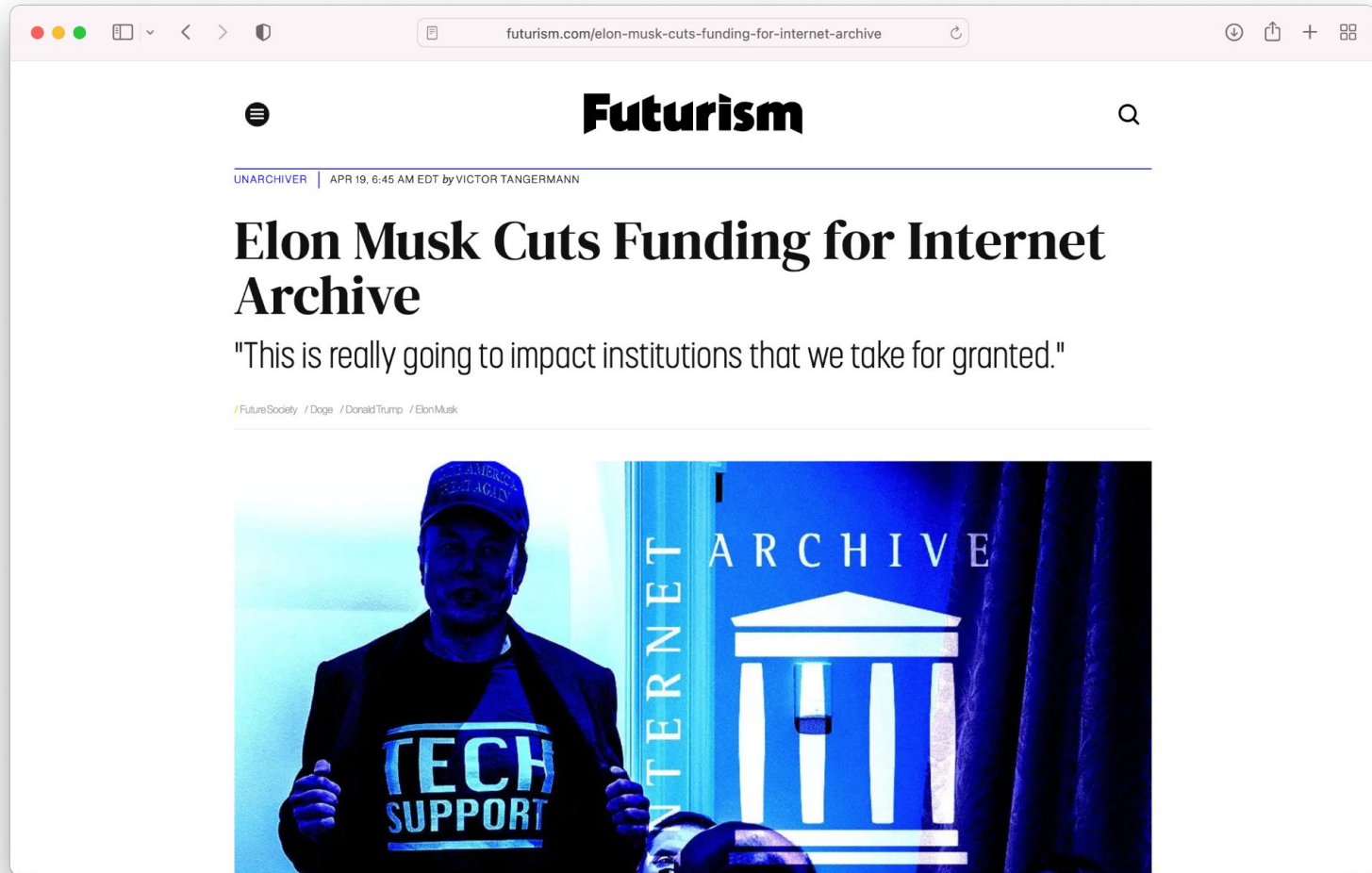
Including attacks with no clear or stated purpose

The image shows a screenshot of a web browser displaying a Forbes article. The browser's address bar shows the URL: forbes.com/sites/larsdaniel/2024/10/20/internet-archive-breached-again-third-cyber-attack-in-october-2024/. The article is titled "Internet Archive Breached Again—Third Cyberattack In October 2024" and is written by Lars Daniel. The article is categorized under "INNOVATION > SCIENCE". Below the title, there is a byline: "By [Lars Daniel](#), Contributor. Lars Daniel covers digital evidence and forensi...". The date of publication is "Oct 20, 2024, 05:13pm EDT". There are social sharing options for "Share", "Save", and "Comment 0". A video player is embedded in the article, with the title "How This AI Company Is Transforming Enterprise Workflows For Billion-Doll...". Below the video player, there is an advertisement for "splunk".

<https://www.forbes.com/sites/larsdaniel/2024/10/20/internet-archive-breached-again-third-cyber-attack-in-october-2024/>

Ideological threats

(even if this article's title is a slight overstatement)



<https://futurism.com/elon-musk-cuts-funding-for-internet-archive>

(It was really only a single, \$350k [NEH grant](#))

DPC 2025-04-24 @phonedudemln.bsky.social, @WebSciDL.bsky.social

Web archives are not just targets of attacks, but can be used as *platforms for attacks*

The screenshot shows a web browser window displaying a blog post. The browser's address bar shows the URL: ws-dl.blogspot.com/2022/09/2022-09-28-using-web-archives-in.html. The page title is "Web Science and Digital Libraries Research Group". Below the title, it says "Research and Teaching Updates from the Web Science and Digital Libraries Research Group (@WebSciDL) at Old Dominion University". The main content of the page is a blog post titled "2022-09-28: Using Web Archives in Disinformation Research" by Michele Weigle, dated September 28, 2022. The blog post features a screenshot of a Medium article titled "Covid-19 had us all fooled, but now we might have finally found its secret." by libertymavenstock, dated Apr 5. The Medium article has a yellow notice at the top stating: "This is an archived web page that Medium.com determined violated their Content Policy. Here is a link to it on the Live Web. In most instances, the archiving of a page is an automated process. The inclusion of a page in the Wayback Machine should not be seen as an endorsement of its content in any way." Below the Medium article, there is a disclaimer: "Anyone can publish on Medium per our Policies, but we don't fact-check every story. For more info about the coronavirus, see cdc.gov." The main text of the Medium article reads: "Covid-19 had us all fooled, but now we might have finally found its secret." and "In the last 3-5 days, a mountain of anecdotal evidence has come out of NYC, Italy, Spain, etc. about COVID-19 and characteristics of patients who get".

<https://ws-dl.blogspot.com/2022/09/2022-09-28-using-web-archives-in.html>
https://twitter.com/phonedude_ml/status/1717022190632100351

My Vision for Trustworthy Web Archiving in 2030

Hundreds of accessible, publicly available,
independent, interoperable, robust,
auditable, cooperating web archives.

[#Disclaimer](#): “...both the live Web and the Wayback Machine [...] are reasonably reliable for everyday use”

My Vision for Trustworthy Web Archiving in 2030

Hundreds of accessible, publicly available, independent, interoperable, robust, auditable, cooperating web archives.

This is doable by 2030. But let's look further at the challenges that could stop us from achieving this goal.

Hundreds of accessible, publicly available,
independent, interoperable, robust,
auditable, cooperating web archives.

Internet Archive: the Walter Cronkite of web archives?

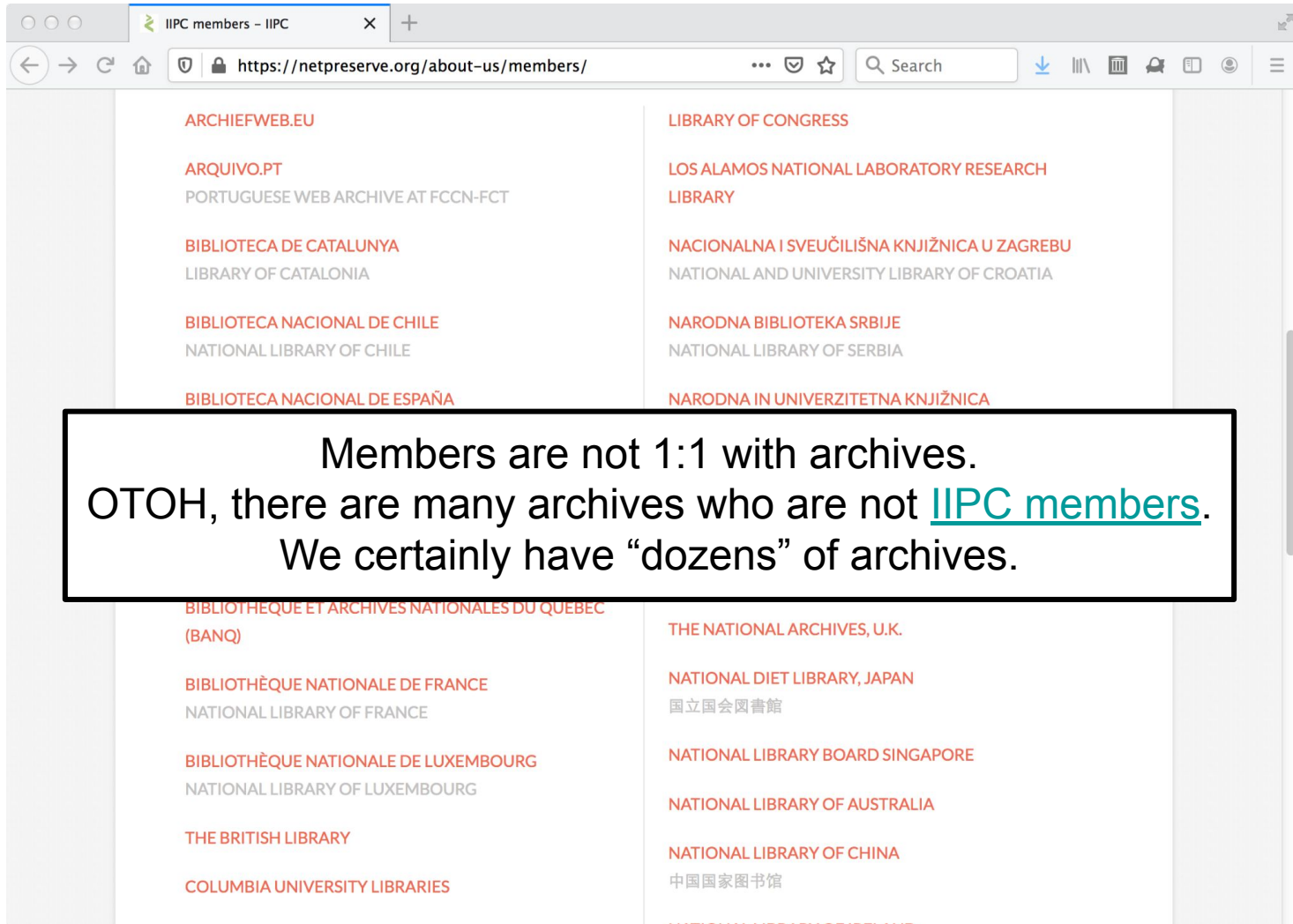


Cable news now offers greater diversity, representation, and POV. However, few anchors offer the gravitas of “Uncle Walter”, “the most trusted man in America”, and some intentionally deceive.

<https://www.britannica.com/biography/Walter-Cronkite>
<https://medium.com/tvnewsanalyzer/visualizing-the-who-and-what-of-cable-tv-news-f51d314b4c2d>

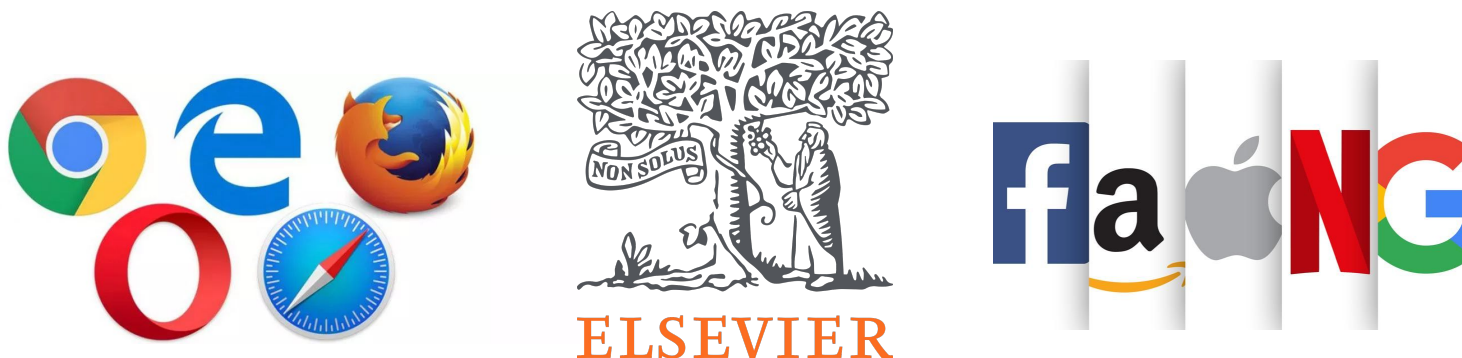
Are we close to 100s of archives?

International Internet Preservation Consortium (IIPC) has 60+ members



Members are not 1:1 with archives.
OTOH, there are many archives who are not [IIPC members](#).
We certainly have “dozens” of archives.

Will the number of archives continue to grow? Maybe not -- innumerable examples point toward centralization / consolidation



IA has admirably supported the Decentralized Web movement.

<https://blog.archive.org/tag/decentralized-web/>

But centralization is about economics, not technologies:

DSHR: “Unless **decentralized technologies** specifically address the issue of how to avoid increasing returns to scale they **will not, of themselves, fix this economic problem**. Their increasing returns to scale will drive layering centralized businesses on top of decentralized infrastructure, replicating the problem we face now, just on different infrastructure.”

<https://blog.dshr.org/2017/08/why-is-web-centralized.html>

<https://www.currentware.com/the-state-of-the-web-browser-in-2020/>

<https://en.wikipedia.org/wiki/Elsevier>

<https://www.forbes.com/sites/sergeiklebnikov/2019/10/15/faang-facebook-amazon-etc-stocks-have-lagged-this-year-heres-why/>

Hundreds of **accessible**, publicly available,
independent, interoperable, robust,
auditable, cooperating web archives.

Web pages are frequently have limited accessibility for the vision impaired, and web archives often make accessibility worse

Accessibility Issues

Usability Issues



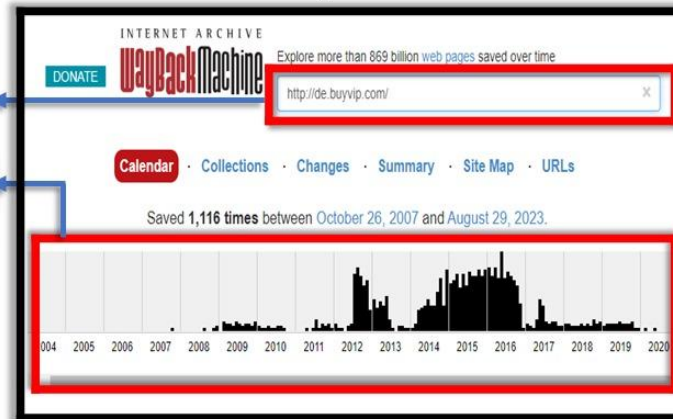
(a) Pandora

Screen readers struggles to find the search bar.

Users struggle to pick the date.

Very low contrast.

Missing alternative text.



(b) Wayback Machine

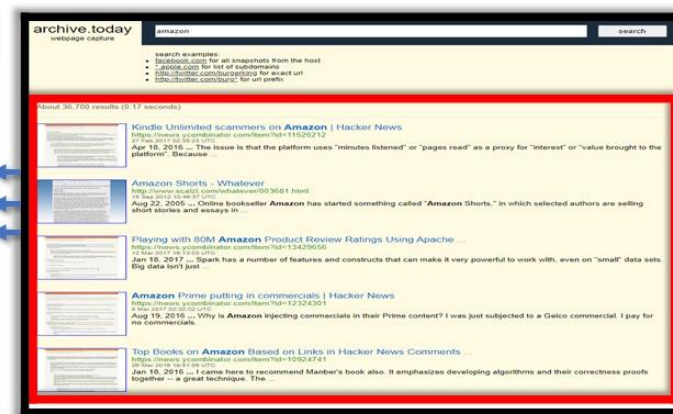


(c) Trove

Broken ARIA reference.

Difficulty in Filtering the Desired Date Snapshot.

Missing Landmark.



(d) Archive.today

https://doi.org/10.1007/978-3-031-72437-4_12

Hundreds of accessible, **publicly available**,
independent, interoperable, robust,
auditable, cooperating web archives.

Two basic kinds of public web archives

Page-at-a-time

"This page is important and I want to preserve it as is for the next 10 years."

Perma.cc 

WebCite

archive.today
webpage capture



Continuous crawl

"I just learned of this page; I wonder what it looked like 10 years ago?"




ARQUIVO.PT

UKWA
UK WEB ARCHIVE



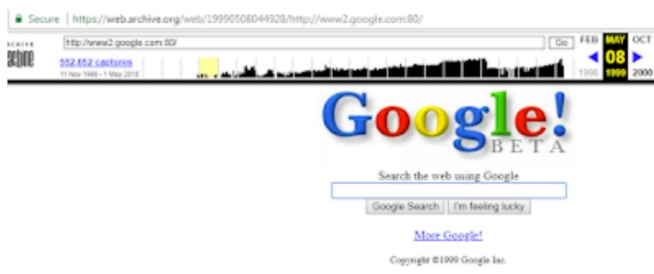
There are subtle but important differences in their operation, capabilities, evaluation, and associated risks. It is confounded by the fact that some archives do both.

We estimated that ~2/3 of web traffic is not publicly archivable



2018-07-18: Why We Need Private Web Archives: Almost Two-Thirds of Web Traffic IS NOT Publicly Archivable

By [Hussam Hallak](#) - July 18, 2018



Google.com mementos from May 8th 1999 on the Internet Archive

In terms of the ability to be archived in public web archives, web pages fall into one of two categories: publicly archivable, or not publicly archivable.

1. Publicly Archivable Web Pages:

These pages are archivable by public archives. The pages can be accessed without login/authentication. In other words, these pages do not reside behind a paywall. [Grant Atkins](#) examined [navwalls in the Internet Archive](#) for news sites and found that web pages behind

<https://ws-dl.blogspot.com/2018/07/2018-07-18-why-we-need-private-web.html>

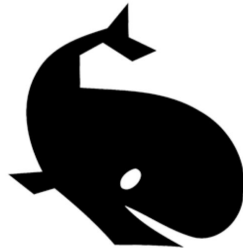
Tools for archiving the private web exist, but the practice, at least as we might think of it, is not yet widespread



WARCreate

Create WARC files from any webpage

[Download for Google Chrome](#)
(source)



Web Archiving Integration Layer (WAIL)

One-Click User Instigated Preservation

Download for [macOS](#), [Windows](#), [Linux](#)
(source)

<https://oduwsdl.github.io/nehdhig2017/>

<https://ws-dl.blogspot.com/2019/09/2019-09-02-so-long-and-thanks-for-all.html>

About ReplayWeb.page



v1.4.5

ReplayWeb.page is a browser-based viewer that loads web archive files provided by the user and renders them for replay in the browser.

Full source code is available at: <https://github.com/webrecorder/replayweb.page>

See the [User Docs](#) or the GitHub README for more info on how it works.

ReplayWeb.page is part of the [Webrecorder Project](#).

Privacy

No data is uploaded anywhere and no information is collected.

All content rendered stays directly in your browser.

When loading an archive from Google Drive, the site may ask for account authorization to download the specified file only.

Disclaimer of Warranties

The application is provided "as is" without any guarantees.

► Legalese:

Close

<https://replayweb.page/>

Commercial private (web) archives largely uninformed by IIPC, Wayback, Heritrix, pywb, Brozzler et al.

PageFreezer Alternatives & Competitors

★★★★☆ (31) 4.5 out of 5

Compare

If you are considering PageFreezer, you may also want to investigate similar alternatives or competitors to find the best solution. Other important factors to consider when researching alternatives to PageFreezer include functionality. We have compiled a list of solutions that reviewers voted as the best overall alternatives and competitors to PageFreezer, including Relativity, Commvault Complete Backup & Recovery, ZDiscovery (formerly Legal Hold Pro & Digital Discovery Pro), and Google Vault.

in | | | | | |

Answer a few questions to help the PageFreezer community

Have you used PageFreezer before?

Yes No

Top Alternatives to PageFreezer

- Relativity
- Commvault Complete Backup & Recovery
- ZDiscovery (formerly Legal Hold Pro & Digital Discovery Pro)
- Google Vault
- MirrorWeb

Show More

Top 20 Alternatives & Competitors to PageFreezer

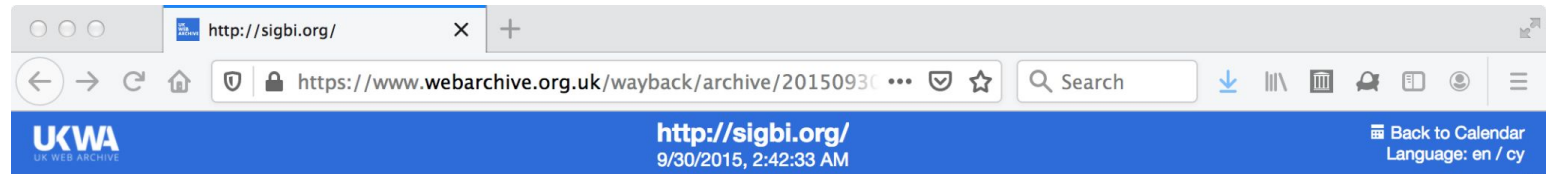
Browse options below. Based on reviewer data you can see how PageFreezer stacks up to the competition, check reviews from current & previous users in industries like Financial Services, Investment Management, and Government Administration, and find the best product for your business.

We use cookies to enhance the functionality of our site and conduct anonymous analytics. With your permission, we may also use cookies to share information about your use of our Site with our social media, advertising, analytics, software, and other third party partners. To learn more, see our [Cookie Policy](#).

I accept I do not accept

<https://www.g2.com/products/pagefreezer/competitors/alternatives>

Dark web archives :-)



Available in Legal Deposit Library Reading Rooms only

This Legal Deposit resource can only be accessed on site in a Legal Deposit Library reading room.

Next steps

If you are in a Legal Deposit Library Reading Room, you should be able to access this resource using a library computer, via the appropriate secure access gateway:

- [via the British Library's Secure Gateway](#)
- [via the National Library of Wales' Secure Gateway](#)
- [via the National Library of Scotland's Secure Gateway](#)
- [via the Cambridge University Library's Secure Gateway](#)
- [via the Trinity College Library's Secure Gateway](#)
- [via the Bodleian Libraries' Secure Gateway](#)

You may also wish to try:

- [Visiting the current, live site](#)
- [Seeing if other web archives hold copies of](#)

```
$ curl -I
https://www.webarchive.org.uk/wayback/archive/20150930064233mp_/http://sigbi.org/
HTTP/1.1 451 Unavailable For Legal Reasons
Server: nginx/1.20.1
Date: Tue, 08 Jun 2021 16:46:14 GMT
Content-Type: text/html
Content-Length: 3947
Connection: keep-alive
$
```

https://en.wikipedia.org/wiki/HTTP_451

Hundreds of accessible, publicly available,
independent, interoperable, robust,
auditable, cooperating web archives.

Three copies archived at exactly the same time -- What are the chances?!

Actually, there are three copies of the *same* observation, not three *independent* observations.

```
$ curl -iLs  
mengator.cs.odu.edu/timemap/link/https://blog.reidreport.com |  
grep 20051213063757  
<https://webarchive.loc.gov/all/20051213063757/http://blog.reidre  
port.com/>; rel="memento"; datetime="Tue, 13 Dec 2005 06:37:57  
GMT",  
<http://archive.md/20051213063757/http://blog.reidreport.com/>;  
rel="memento"; datetime="Tue, 13 Dec 2005 06:37:57 GMT",  
<https://web.archive.org/web/20051213063757/http://blog.reidrepor  
t.com/>; rel="memento"; datetime="Tue, 13 Dec 2005 06:37:57 GMT",
```

It will never be [2005](#) again, so hosting IA's WARC files from 2005 is the best we can do. Going forward, it would be nice to have 3+ independent observations, which could *all be different* because of GeolP, personalization, CDN status, etc.

***Then it's up to the reader to determine if the differences
are semantically meaningful.***

Three copies archived at exactly the same time -- What are the chances?!

Actually, there are three copies of the *same* observation, not three *independent* observations.

```
$ curl -iLs
```

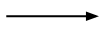
```
mimgator.cs.odu.edu/timemap/link/https://blog.reidreport.com |
```

```
gr  
<h  
po  
GM  
<h  
re  
<h
```

this means we have 3 different copies
of the *same image* ...



...not three *independent images* taken
at (approximately) the same time



```
t.com/>; rel="memento"; datetime="Tue, 13 Dec 2005 06:37:57 GMT",
```

It will never be [2005](#) again, so hosting IA's WARC files from 2005 is the best we can do. Going forward, it would be nice to have 3+ independent observations, which could *all be different* because of GeolP, personalization, CDN status, etc.

Then it's up to the reader to determine if the differences are semantically meaningful.

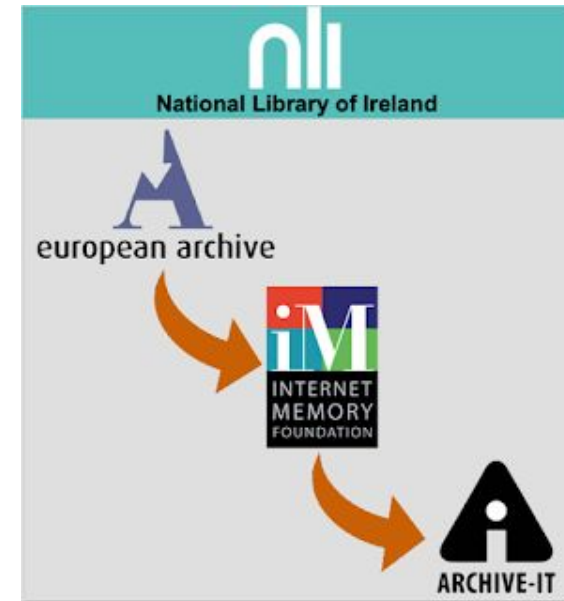
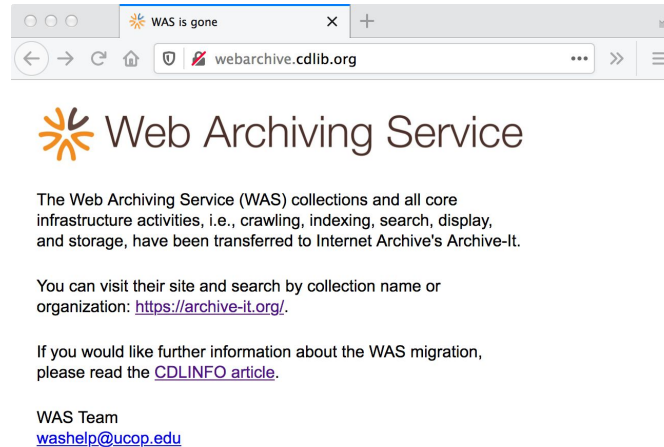
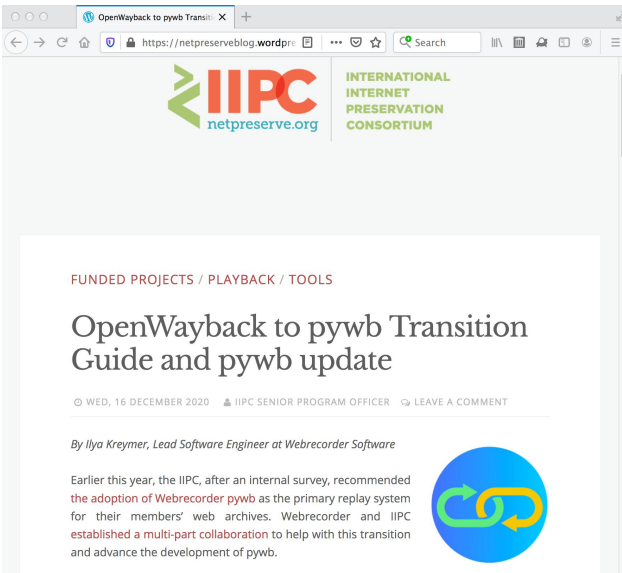
Hundreds of accessible, publicly available,
independent, **interoperable**, robust,
auditable, cooperating web archives.

Homogeneity is not true interoperability

“Nobody ever gets fired for buying IBM”

I don't fault the staff who converge on popular, high-quality tech stacks & services,
but I do lament the loss of heterogeneity.

True interoperability comes through the hard work of protocols and standards.



<https://netpreserveblog.wordpress.com/2020/12/16/openwayback-to-pywb-transition-guide/>

<http://webarchive.cdlib.org/>

<https://ws-dl.blogspot.com/2019/09/2019-09-10-where-did-archive-go-part-2.html>

Hundreds of accessible, publicly available,
independent, interoperable, **robust**,
auditable, cooperating web archives.

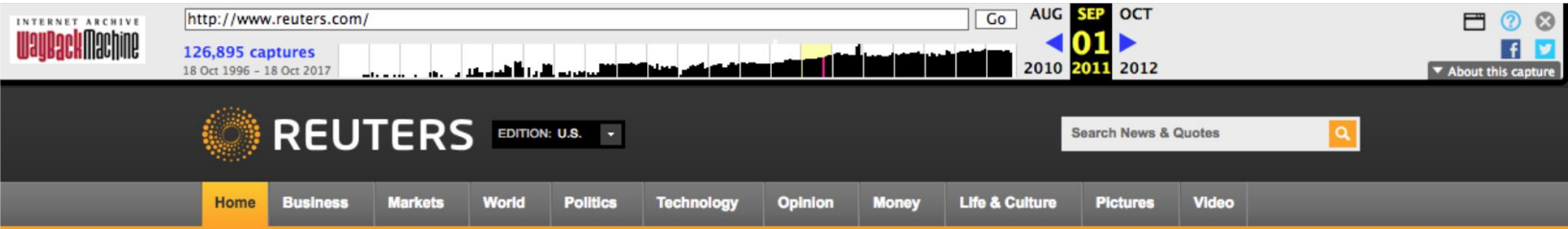
Archives must protect the past web from intrusion by the live web



Content from the live web (ca. 2012) accidentally leaking into this page archived in 2008

<https://ws-dl.blogspot.com/2012/10/2012-10-10-zombies-in-archives.html>

2017: Lerner et al. demonstrate live web leakage as an attack on archived web pages



Recession fears recede as factories show growth

Exclusive: Regulators seek trading secrets

WASHINGTON/NEW YORK (Reuters) - U.S. securities regulators have taken the unprecedented step of asking high-frequency trading firms to hand over the details of their trading strategies, and in some cases, their secret computer codes.

August auto sales defy consumer caution

Fed orders Goldman to review foreclosures

to pull content from as talks fail

rm strikes out with all's New York Mets

is court asked to vacate "Ask" judgment

OPINION

CCS 2017 will be held in Dallas, USA!

By Cliff Young and Chris Jackson

The conventional wisdom now is that the president is beatable in 2012. As a whole, Obama is still a favorite, but the numbers show he is dangerously close to having from a clear favorite to a struggling contender. [Commentary](#)

Easterbrook: The election's first bogeyman

Chen: China is still waiting for inflation to peak

They inserted (2017) content of their choosing into this page from 2011

<https://rewritinghistory.cs.washington.edu>

2017: First published works about robustness vs. malicious .html/.js?

Thinking like a hacker:
Security Considerations for
High-Fidelity Web Archives

Jack Cushman, Perma.cc
Ilya Kreymer, Webrecorder

[http://labs.rhizome.org/presentations/security.html/#/](http://labs.rhizome.org/presentations/security.html#/)
<https://blog.dshr.org/2017/06/wac2017-security-issues-for-web-archives.html>

WEBRECORDER Perma.cc

Prior to these works, our group (@WebSciDL) had observed: [Zombies](#) (live web leakage into the archive), [Temporal Violations](#) (replaying web pages that never existed), [Cookie Violations](#), [Twitter replay problems](#), etc., but we never considered [ingesting malicious .html/.js](#) until these groundbreaking pubs.

Session H3: Web Security
CCS'17, October 30-November 3, 2017, Dallas, TX, USA

Rewriting History: Changing the Archived Web from the Present

Ada Lerner^{*}
Wellesley College
alerner@wellesley.edu

Tadayoshi Kohno
Paul G. Allen School
of Computer Science & Engineering
University of Washington
yoshi@cs.washington.edu

Franziska Roesner
Paul G. Allen School
of Computer Science & Engineering
University of Washington
franzi@cs.washington.edu

ABSTRACT
The Internet Archive's Wayback Machine is the largest modern web archive, preserving web content since 1996. We discover and analyze several vulnerabilities in how the Wayback Machine archives data, and then leverage these vulnerabilities to create what are to our knowledge the first attacks against a user's view of the archived web. Our vulnerabilities are enabled by the unique interaction between the Wayback Machine's archives, other websites, and a user's browser, and attackers do not need to compromise the archives in order to compromise users' views of a stored page. We demonstrate the effectiveness of our attacks through proof-of-concept implementations. Then, we conduct a measurement study to quantify the prevalence of vulnerabilities in the archive. Finally, we explore defenses which might be deployed by archives, website publishers, and the users of archives, and present the prototype of a defense for clients of the Wayback Machine, ArchiveWatcher.

CCS CONCEPTS
• Information systems → Digital libraries and archives; • Security and privacy → Web application security.

KEYWORDS
web archives; web security

1 INTRODUCTION
The Wayback Machine is a publicly browsable web archive which has cataloged and preserved a collection of over 286 billion web pages over the period from 1996 to 2017 [26]. Like other web archives, which use similar techniques and technologies, the Wayback Machine allows clients using ordinary web browsers to access snapshots of past websites through a web interface¹, enabling ordinary citizens as well as technical experts to see how the web has changed and what it once contained. These archival snapshots of websites are rendered in HTML, Javascript, and CSS just like

^{*}This work was performed while Dr. Lerner was a PhD Candidate at the Paul G. Allen School of Computer Science at the University of Washington.
¹<https://web.archive.org>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CCS '17, October 30-November 3, 2017, Dallas, TX, USA

the modern web, preserving not only their content but also their client-side dynamic behaviors, making them a rich cultural and technical preserve.

The Wayback Machine is frequently used in a variety of contexts critical to our free society, including scholarly articles, journalism, and legal proceedings. Scientists may cite archived snapshots in their scientific papers to increase the durability of their references [18, 41], while journalists have used archives to understand how websites such as official government pages have changed [20], and lawyers often use archival snapshots as evidence in legal cases, including civil and criminal ones, administrative proceedings, and patent litigation (e.g., [1, 2, 4, 40]). While other researchers have studied inaccuracies in the Wayback Machine which arise accidentally, we observe that these socially and financially important uses suggest incentives to intentionally manipulate archives after the fact. For example, governments might want to suppress or change historical information, companies might want to manipulate evidence of prior art in a patent case, organizations might want to hide evidence of past wrongdoing, and news sources might want to manipulate source material for their reporting.

To our knowledge, this paper is the first to investigate the technical vulnerabilities and attacks that might be used to perform such intentional manipulation. That is: how might attackers attempt to rewrite history? How might they intentionally cause clients who view the archive to see archived websites with content, appearance, and behavior that are different from the actual website at archival time? We analyze the way that the Wayback Machine functions, finding that in fact, there are several types of vulnerabilities which would allow an attacker today to take full control of clients' views of snapshots. For example, snapshots sometimes cause clients to accidentally miss content from the live web into an archived page, allowing servers on the live web to inject content or code into clients' views of the archive. Our attacks are global – they affect the appearance and behavior of snapshots for all visitors, and they do not involve the direct compromise of archival or publisher servers or databases.

We demonstrate the viability of our attacks with proofs-of-concept. For example, we demonstrate the ability to inject arbitrary Javascript code into client views of archival snapshots, allowing us to modify text, images, styling, and behavior, subtly or completely rewriting the web of the past. Figure 1 shows such an attack, in which we took complete control of a snapshot of minters.com from 2011.²

We then quantify the prevalence of the types of vulnerabilities we discovered, seeking them in the wild through a measurement study of archived websites. We find that vulnerabilities to our attacks are

<https://acmccs.github.io/papers/p1741-lernerAT3.pdf>
<https://blog.dshr.org/2017/09/attacking-users-of-wayback-machine.html>

2018: Web IDL & Client-side rewriting

2020: Analysis of attacks on *rehosting* sites

TO RELIEVE THE WEB: A FRAMEWORK FOR THE
TRANSFORMATION AND ARCHIVAL REPLAY OF
WEB PAGES

by

John Andrew Berlin
B.S. December 2015, Old Dominion University

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
May 2018

I signed off on John's thesis 6 years ago, but
I'm only now *really* understanding it.

Michele C. Weigle (Director)

Michael L. Nelson (Member)

Justin F. Brunelle (Member)

<https://ws-dl.blogspot.com/2018/04/2018-05-01-high-fidelity-ms-thesis-to.html>

<https://www.ndss-symposium.org/ndss-paper/melting-pot-of-origins-compromising-the-intermediary-web-services-that-rehost-websites/>

Melting Pot of Origins: Compromising the Intermediary Web Services that Rehost Websites

Takuya Watanabe*¹, Eitaro Shioji*, Mitsuaki Akiyama* and Tatsuya Mori†^{1,§}

*NTT Secure Platform Laboratories, Tokyo, Japan

¹Waseda University, Tokyo, Japan

[†]NICT, Tokyo, Japan

[§]RIKEN AIP, Tokyo, Japan

Email: watanabe@nsl.cs.waseda.ac.jp, mori@nsl.cs.waseda.ac.jp

Abstract—Intermediary web services such as web proxies, web translators, and web archives have become pervasive as a means to enhance the *openness* of the web. These services aim to remove the intrinsic obstacles to web access; i.e., access blocking, language barriers, and missing web pages. In this study, we refer to these services as *web rehosting* services and make the first exploration of their security flaws. The web rehosting services use a single domain name to rehost several websites that have distinct domain names; this characteristic makes web rehosting services intrinsically vulnerable to violating the same origin policy if not operated carefully. Based on the intrinsic vulnerability of web rehosting services, we demonstrate that an attacker can perform five different types of attacks that target users who make use of web rehosting services: persistent man-in-the-middle attack, abusing privileges to access various resources, stealing credentials, stealing browser history, and session hijacking/injection. Our extensive analysis of 21 popular web rehosting services, which have more than 200 million accesses per day, revealed that these attacks are feasible. In response to this observation, we provide effective countermeasures against each type of attack.

I. INTRODUCTION

While the architecture of the web follows the fundamental Internet design framework — the *End-to-End principle* [53] — several web services that deviate from the principle have been developed to enhance the openness of the web. Typical examples of such services include *web proxy*, *web translator*, and *web archive*, which all aim to enhance the openness of

one of the most popular web proxy services, ProxySite [51] had more than 20 million accesses per day in September 2019. The world's top search engine companies such as Google, Microsoft, Baidu, and Yandex offer web translator services. Google website translator alone supports over 100 languages [22], [55] and serves over 80 million accesses per day from all over the world. Through an analysis of anonymized access logs collected from Wayback Machine servers in February 2012, AlNoamany et al. [3] reported that the service had about 82 million accesses per day. All these web rehosting services have become popular because they enhance the openness of the web. Furthermore, they are easy to use, as a user can simply access such services by using a normal browser and inputting a URL of interest, unlike other alternative solutions such as an HTTP proxy¹, VPN, or Tor, they do not require setting changes and the installation of special applications.

As web rehosting services offer enhanced web access to various websites, an attacker has an incentive to exploit them because users may input privacy-sensitive information while accessing *rehosted websites* via a web rehosting service. For instance, a user who uses a web proxy to access to a webmail service needs to input the credential of the webmail account via the web proxy. Similarly, content rewriting on web rehosting services is useful for inciting and intimidating users. An attacker may want to abuse the web translator to create fake news by completely rewriting an original news

the web access transla in a fe Web a publish several All the call "b

Key contribution: web archives
as subclass of rehosting sites.

With the increase in the number of web users, web rehosting services have become pervasive. For instance, according to the web traffic statistics provided by SimilarWeb [55],

Network and Distributed Systems Security (NDSS) Symposium 2020
23-26 February 2020, San Diego, CA, USA
ISBN 1-891562-41-4
<https://dx.doi.org/10.14722/ndss.2020.24140>
www.ndss-symposium.org

the web the st of ively and s the sting model

The key idea of the attacks is to leverage the fact that a single domain name provided by a web rehosting service is used to access multiple rehosted websites; this *“melting pot of origins”* situation allows an attacker to bypass the filtration of the same origin policy (SOP). A malicious website rehosted by an attacker to the web rehosting has the same

¹In this paper, we call proxies with web-based interfaces (i.e., a type of web rehosting services) web proxies and distinguish them from HTTP proxies.

It is now harder to get live web resources into IA

The screenshot shows the Chrome DevTools Network tab. The main table lists various requests, with the following rows highlighted by red boxes:

Status	Method	Domain	File	Initiator	Type	Transferred	Size
200	GET	web.archive...	styles.css?v=X-Ndhul9	stylesheet	css	cached	3.75 ...
404	GET	web.archive...	favicon.ico	FaviconLoad...	html	cached	153 B
200	GET	web.archive...	/	stillLiveWebV...	html	26.60 KB	98.5...
200	GET	web.archive...	stillLiveWebV2.html	document	html	cached	2 KB
200	GET	polyim.archi...	polyim.min.js?features=fetch,intersectionOB...	script	js	cached	101 B
200	GET	img.wenner...	rs-135838-fc57703c7f9b5a96f4e56cf74f17...	stillLiveWebV...		0 B	0 B
200	GET	i0.kym-cdn....	4b4.jpg	stillLiveWebV...	jpeg	cached	48.6...
304	GET	gregfranko....	JavaScript-logo-small.png	stillLiveWebV...	png	cached	7.09 ...

The Headers panel for the selected request (GET http://i0.kym-cdn.com/photos/images/facebook/000/883/283/4b4.jpg) shows the following details:

- Status: 200 OK
- Version: HTTP/1.1
- Transferred: 48.67 KB (48.67 KB size)
- Referrer Policy: no-referrer-when-downgrade
- Response Headers (523 B):
 - Accept-Ranges: bytes
 - Access-Control-Allow-Origin: http://web.archive.org
 - Age: 155641
 - Cache-Control: max-age=315360000
 - Connection: keep-alive
 - Content-Length: 49837
 - Content-Type: image/jpeg

As a result of the Lerner et al. paper in 2017, the IA added a `Content-Security-Policy` header to prevent live web leakage.

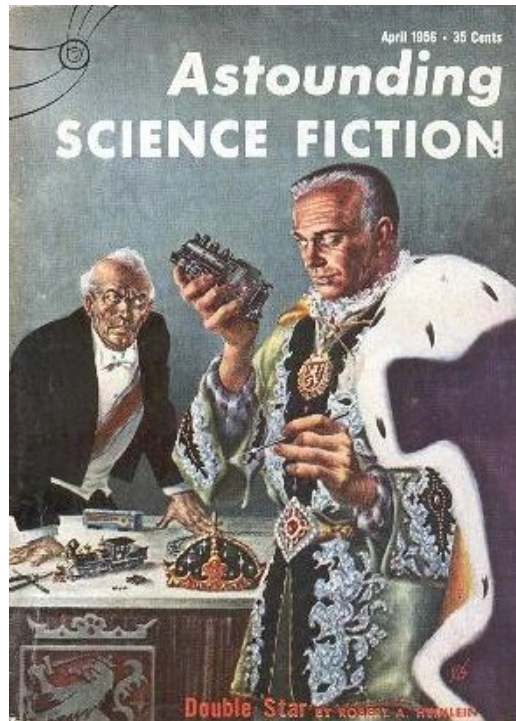
But some web archives still allow live web leakage (AKA "Zombies")

The screenshot shows a browser window displaying a CNN article from the Arquivo.pt archive. The article title is "'Freedom Day' is a huge gamble by Johnson". The browser's developer tools are open to the Network tab, showing a list of requests. A red box highlights the 'Domain' column, which lists various domains including google-analytics.com, stats.g.doubleclick.net, signal-metrics-collector-b..., signal-beacon.s-onetag.com, fonts.googleapis.com, and arquivo.pt. An arrow points from the 'arquivo.pt' domain entry to a console message showing a script source: `<script nomodule="" src="https://arquivo.pt/wayback/static/@ionic/core/dist/ionic/ionic.js"></script>`.

Status	Method	Domain	File	Initiator	Type	Transferred	Size	Time
200	GET	www.google-analytics.com	collect?v=1&_v=j96&a=1693312744&t=pageview&s=3&dl=https://arquivo.pt/...	analytics.js:31 (img)	gif	620 B	35 B	15 ms
200	GET	www.google-analytics.com	collect?v=1&_v=j96&a=1693312744&t=pageview&s=2&dl=https://arquivo.pt/...	analytics.js:31 (img)	gif	620 B	35 B	16 ms
200	POST	www.google-analytics.com	collect?v=1&_v=j96&a=1693312744&t=pageview&s=1&dl=https://arquivo.pt/...	analytics.js:44 (xhr)	plain	634 B	4 B	16 ms
304	GET	www.google-analytics.com	analytics.js	/wayback/202109300828...	js	cached	49.03 KB	181 ms
200	POST	stats.g.doubleclick.net	collect?t=dc&aip=1&_r=3&v=1&_v=j96&tid=UA-21825027-1&cid=222534983...	analytics.js:44 (xhr)	plain	704 B	1 B	234 ms
200	POST	signal-metrics-collector-b...	metrics	cnn-beacon.min.js:37 (be...	plain	967 B	0 B	
304	GET	signal-beacon.s-onetag.com	n-beacon.min.js?affiliateId=276894	tag.min.js line 84 > Funct...	js	cached	29.70 KB	1 ms
200	GET	fonts.googleapis.com	cs?family=Roboto&display=swap	stylesheet	css	1.25 KB (raced)	2.13 KB	96 ms
200	GET	arquivo.pt	beacon.js	wombat.js:21 (script)				

these URLs should all be from arquivo.pt

If you can't get files from the live web...
...get them from elsewhere in the archive



“...the chronoscope is inherently limited to recent times—but what if, instead of focusing it upon the past of a generation earlier, it were tuned to the past of one-hundredth of a second ago? The dead past, Araman says, is only a synonym for “the living present”. ”

https://en.wikipedia.org/wiki/The_Dead_Past

Temporal spread, temporal violations:

Tactical: improve the UI; Strategic: better idioms that convey temporality

ReidBlog

web.archive.org/web/20060111221738/http://blog.reidreport.com/

431 captures
11 Mar 2005 - 10 Aug 2022

COLLECTED BY
Collection: web_sup
Crawl performed by Internet Archive. This data is currently not publicly accessible.

TIMESTAMPS

- http://embed.technorati.com/embed/9i7maurw2w.js +2 years 5 months
- http://static.technorati.com/pix/tn-tiny.gif +2 years 5 months
- http://rpc.blogrolling.com/display.php?r=b974d0ed8af3a2544556827467da9780 +1 year 1 month
- http://truthlaidbear.com/MyDetails.php?url=http://blog.reidreport.com/&style=javascript +1 year 1 month
- http://www.blogrankers.com/tracker.php?do=in&id=132/ +1 year 1 month
- http://www.yceml.net/0675/10385059-5.gif -2 months 17 days
- http://www.yceml.net/0379/5346683-12.gif -2 months 17 days
- http://www.haloscan.com/load/reidreport +26 days 6 hours
- http://www.haloscan.com/load/reidreport +26 days 6 hours
- http://metapager.com/giz/G05/debt.html?zWidth=140&zHeight=18&zFontSpeed=8&zFontSize=14&zFontColor=red&zBorderColor=%23CCC&zBackColor=& -22 days 22 hours
- http://mediamatters.org/static/images/mmfa_logo150.gif +2 days 2 hours
- http://mediamatters.org/tools/syndication/headlinenews.html +9 hours 17 minutes
- http://pagead2.googleadsyndication.com/pagead/show_ads.js -9 hours 15 minutes
- http://www.haloscan.com/halolink.gif -9 hours 15 minutes
- http://www.zfacts.com/giz/G05/debt.js -9 hours 13 minutes
- http://static.flickr.com/23/25822676_789bf55448_t.jpg -8 hours 54 minutes
- http://technorati.com/pix/tech-logo-embed.gif -8 hours 51 minutes

A large temporal spread, but not necessarily a temporal violation.

This is the semantically meaningful temporal violation

https://twitter.com/phonedude_mln/status/990054945457147904
<https://ws-dl.blogspot.com/2015/12/2015-12-08-evaluating-temporal.html>

Youtube video of Brian Williams rapping "Gin & Juice" archived in the Wayback Machine in 1992

Everything in the above statement is obviously false, but the page gives the *illusion* of being archived in 1992.



Due to improvements in the IA playback, this illusion no longer works. Similar illusions are still possible, but are harder to construct.

John Berlin, MS Thesis, 2018

<https://www.youtube.com/watch?v=k3QTcJZdFfs>

(actual URI-R & URI-M have also been obscured in the video to hide the technique)

The tradespace of functionality & security

- My hypothesis: no archive can be immune to illusions if it supports:
 - JavaScript replay
 - cf. archive.today's approach of removing all JavaScript
 - unauthenticated write operations
 - e.g., Save Page Now
 - "raw" or "unwritten" mode
 - typically with "id_" in the URL

Hundreds of accessible, publicly available,
independent, interoperable, robust,
auditable, cooperating web archives.

Load the archived page, get an eagle

You are viewing a Web site, archived on **6:06:24 Jan 19, 2013** It is now a Federal record managed by the National Archives and Records Administration.
External links, forms, and search boxes may not function within this collection. [[hide](#)]

Conserving the Nature of America

News Endangered Species Offices Wildlife Refuges Social Media Jobs Kids, Educators, Parents Permits Multimedia

the Open Spaces blog
A **talk** on the **WILD** side [Click to Read](#)

Public Questions

Email the USFWS Customer Service Center, or
Call 1-800-344-WILD
1-800-344-9453

Programs/Resources

[Budget](#)
[Business Opportunities](#)
[Climate Change](#)
[Coastal](#)
[Congressional/Legislative Affairs](#)
[Diversity and Equal](#)

News & Feature Stories

Connecting People With Nature: Ensuring the Future of Conservation New Publication Marks 75th Anniversary of Conservation Effort
January 17, 2012

The Fish and Wildlife Service today released a landmark publication celebrating the 75th Anniversary of the Wildlife and Sport Fish Restoration Program, the cornerstone of fish and wildlife conservation in North America. Each year this unique program provides more than \$700 million to support habitat conservation and outdoor recreation projects across the nation through the sale of outdoor equipment.

The full-color 75-page anniversary publication – **Celebrating the Wildlife and Sport Fish Restoration Program, 75 years of Conservation and Partnership Success** – highlights a year-long celebration involving a host of partners including state fish and wildlife agencies, non-governmental conservation organizations, and industry partners. To accompany release of this new

U.S. Fish and Wildlife Service
Like 37,959

Did You Know ?

About Our Migratory Birds Program
The Fish and Wildlife Service's **Migratory Bird Program** works to manage bird populations and their habitats, increase awareness of and appreciation for the value of migratory birds, and improve migratory bird hunting, birdwatching, and other outdoor bird-related experiences.
[Learn More](#)
[Video](#)

Director's Corner

<https://www.webharvest.gov/congress112th/20130119060624/http://www.fws.gov/>

Hit “reload”, get a tiger

You are viewing a Web site, archived on **6:06:24 Jan 19, 2013** It is now a Federal record managed by the National Archives and Records Administration.
External links, forms, and search boxes may not function within this collection. [[hide](#)]

Conserving the Nature of America

[News](#) [Endangered Species](#) [Offices](#) [Wildlife Refuges](#) [Social Media](#) [Jobs](#) [Kids, Educators, Parents](#) [Permits](#) [Multimedia](#)

the Open Spaces blog
A *talk* on the **WILD** side [Click to Read](#)

Public Questions

Email the USFWS Customer Service Center, or
Call 1-800-344-WILD
1-800-344-9453

Programs/Resources

[Budget](#)
[Business Opportunities](#)
[Climate Change](#)
[Coastal](#)
[Congressional/Legislative Affairs](#)
[Diversity and Equal](#)

News & Feature Stories

Connecting People With Nature: Ensuring the Future of Conservation
New Publication Marks 75th Anniversary of Conservation Effort
January 17, 2012

The Fish and Wildlife Service today released a landmark publication celebrating the 75th Anniversary of the Wildlife and Sport Fish Restoration Program, the cornerstone of fish and wildlife conservation in North America. Each year this unique program provides more than \$700 million to support habitat conservation and outdoor recreation projects across the nation through the sale of outdoor equipment.

The full-color 75-page anniversary publication – **Celebrating the Wildlife and Sport Fish Restoration Program, 75 years of Conservation and Partnership Success** – highlights a year-long celebration involving a host of partners including state fish and wildlife agencies, non-governmental conservation organizations, and industry partners. To accompany release of this new

U.S. Fish and Wildlife Service
Like 37,959

Did You Know ?

About Our Migratory Birds Program
The Fish and Wildlife Service's **Migratory Bird Program** works to manage bird populations and their habitats, increase awareness of and appreciation for the value of migratory birds, and improve migratory bird hunting, birdwatching, and other outdoor bird-related experiences.
[Learn More](#)
[Video](#)

Director's Corner

<https://www.webharvest.gov/congress112th/20130119060624/http://www.fws.gov/>

Hit “reload” again, get a mountain

You are viewing a Web site, archived on **6:06:24 Jan 19, 2013** It is now a Federal record managed by the National Archives and Records Administration.
External links, forms, and search boxes may not function within this collection. [[hide](#)]

Conserving the Nature of America

News Endangered Species Offices Wildlife Refuges Social Media Jobs Kids, Educators, Parents Permits Multimedia

the Open Spaces
blog

A **talk** on the **WILD** side [Click to Read](#)

Public Questions

Email the USFWS Customer Service Center, or
Call 1-800-344-WILD
1-800-344-9453

Programs/Resources

[Budget](#)
[Business Opportunities](#)
[Climate Change](#)
[Coastal](#)
[Congressional/Legislative Affairs](#)
[Diversity and Equal](#)

News & Feature Stories

Connecting People With Nature: Ensuring the Future of Conservation
New Publication Marks 75th Anniversary of Conservation Effort
January 17, 2012

The Fish and Wildlife Service today released a landmark publication celebrating the 75th Anniversary of the Wildlife and Sport Fish Restoration Program, the cornerstone of fish and wildlife conservation in North America. Each year this unique program provides more than \$700 million to support habitat conservation and outdoor recreation projects across the nation through the sale of outdoor equipment.

The full-color 75-page anniversary publication – **Celebrating the Wildlife and Sport Fish Restoration Program, 75 years of Conservation and Partnership Success** – highlights a year-long celebration involving a host of partners including state fish and wildlife agencies, non-governmental conservation organizations, and industry partners. To accompany release of this new

U.S. Fish and Wildlife Service
Like 37,959

Did You Know ?

About Our Migratory Birds Program
The Fish and Wildlife Service's **Migratory Bird Program** works to manage bird populations and their habitats, increase awareness of and appreciation for the value of migratory birds, and improve migratory bird hunting, birdwatching, and other outdoor bird-related experiences.
[Learn More](#)
[Video](#)

Director's Corner

<https://www.webharvest.gov/congress112th/20130119060624/http://www.fws.gov/>

Randomness is caused by JavaScript

As long as all three images are archived, all three valid states for this page can be replayed

```
function random_imglink() {
  myimages[1]="/congress112th/20130119060624/http://www.fws.gov/home/feature/home-banner/open-spaces/bannerbluemnt.jpg";
  myimages[2]="/congress112th/20130119060624/http://www.fws.gov/home/feature/home-banner/open-spaces/bannereagle.jpg";
  myimages[3]="/congress112th/20130119060624/http://www.fws.gov/home/feature/home-banner/open-spaces/bannertiger.jpg";

  var ry=Math.floor(Math.random(1)*myimages.length)

  if (ry==0)
    ry=1

  document.write('<a href='+''+imagelinks[ry]+'''+'><img src='' +myimages[ry]+''' border="0" alt="The Open Spaces Blog. A Talk on the Wild Side. Click to Read"></a>')
}
```

The HTML remains unchanged, but the composite page changes randomly.
<https://ws-dl.blogspot.com/2017/12/2017-12-11-difficulties-in-timestamping.html>

The fws.gov example is easy to diagnose; most changes are much harder to explain

The screenshot shows a Chrome browser window displaying a CNN.com page from the Wayback Machine. The browser's address bar shows the URL `https://web.archive.org/web/20130530221910/http://www.cnn.com/`. The Wayback Machine search bar is also visible, containing the URL `http://www.cnn.com/` and showing 154,854 captures from 20 Jun 2000 to 21 Sep 2017. The page content includes a 'Must Watch TV' section with program listings, a weather widget for Atlanta, GA, and a navigation menu at the bottom.

Must Watch TV

Program	Time
The Lead	4pm ET / 1pm PT on CNN
The Situation Room	5pm ET / 2pm PT on CNN
Erin Burnett: OutFront	7pm ET / 4pm PT on CNN
AC 360	8pm ET / 5pm PT on CNN
Piers Morgan Live	9pm ET/PT on CNN

Weather: 83° HI 87° LO 71° Atlanta, GA Weather forecast

Navigation: Home | Video | CNN Trends | U.S. | World | Politics | Justice | Entertainment | Tech | Health | Living | Travel | Opinion | iReport | Money | Sports | Tools & widgets | RSS | Podcasts | Blogs | CNN mobile | My profile | E-mail alerts | CNN shop | Site map | Contact us

<http://web.archive.org/web/20130530221910/http://www.cnn.com/>

2024 replay: less is more?

The screenshot shows the CNN website interface from a web archive capture. The browser address bar displays the URL: <http://www.cnn.com/>. The page header includes the CNN logo and navigation links. The main content area is titled "TV & VIDEO" and features a "Must Watch TV" section with five program thumbnails: "The Lead", "The Situation Room", "Erin Burnett: OutFront", "AC 360", and "Piers Morgan Live". Below this is a "Trending Video" section. A red box highlights an empty area on the left side of the page. Another red box contains the text: "The inaccurate weather is now missing. Is that better? Probably?".

<http://web.archive.org/web/20130530221910/http://www.cnn.com/>

2025 replay: return of the cold front

INTERNET ARCHIVE
Wayback Machine
942,099 captures
20 Jun 2009 – 20 Apr 2025

web.archive.org/web/20130530221910/http://www.cnn.com/

Go APR MAY JUN 30 2012 2013 2014 About this capture

TV & VIDEO TV Programs

ON NOW

Must Watch TV CNN Radio · HLN · Full Schedule

The Lead 4pm ET / 1pm PT on CNN
The Situation Room 5pm ET / 2pm PT on CNN
Erin Burnett: OutFront 7pm ET / 4pm PT on CNN
AC 360 8pm ET / 5pm PT on CNN
Piers Morgan Live 9pm ET/PT on CNN

Trending Video View Collections

Watch TV

ADVERTISEMENT

ADVERTISEMENT

ADVERTISEMENT

Hi 66° LO 44°
Atlanta, GA Weather forecast

SEARCH
POWERED BY Google

Home | Video | CNN Trends | U.S. | World | Politics | Justice | Entertainment | Tech | Health | Living | Travel | Opinion | iReport | Money | Sports
Tools & widgets | RSS | Podcasts | Blogs | CNN mobile | My profile | E-mail alerts | CNN shop | Site map | Contact us

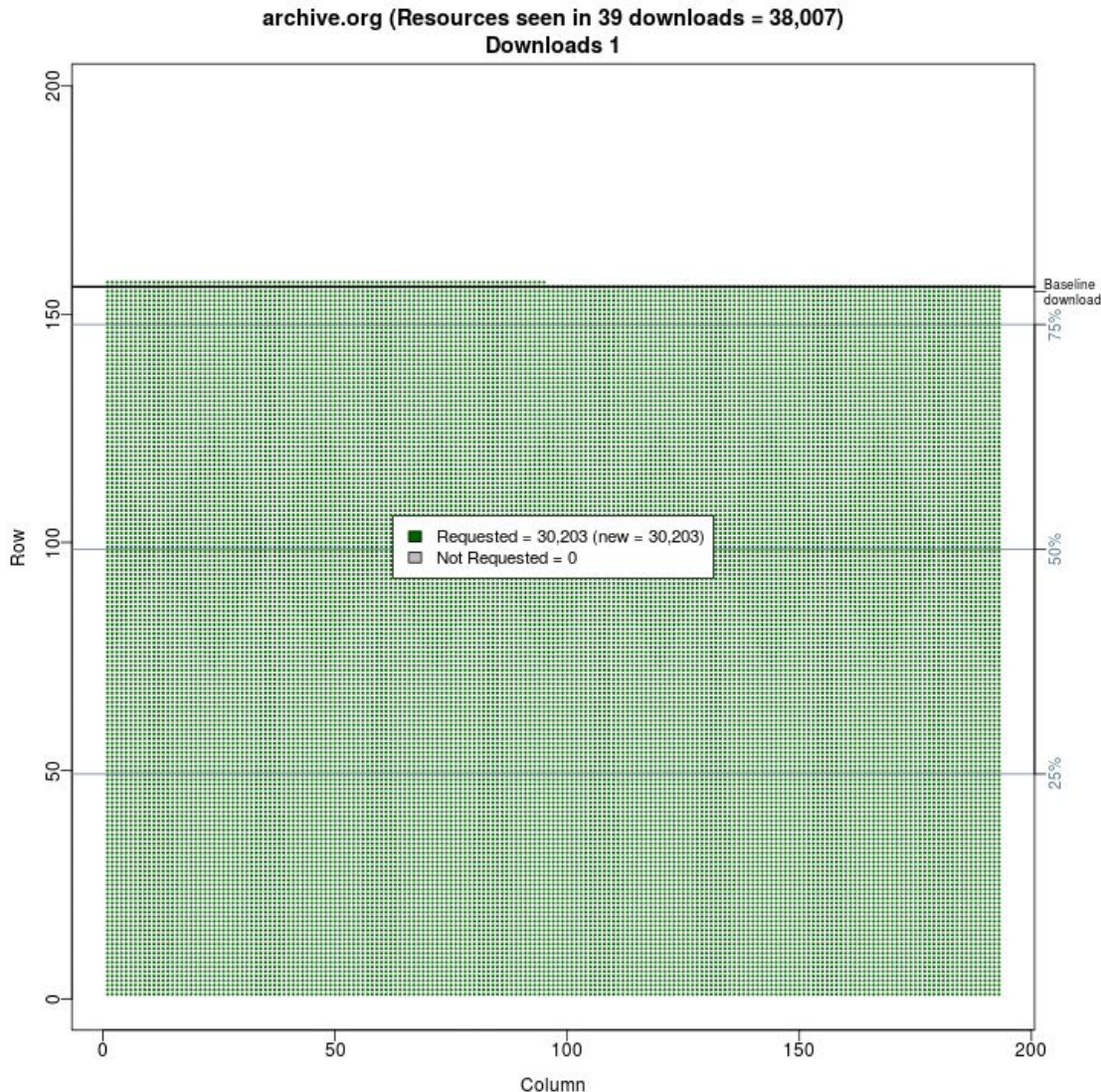
CNN © 2013 Cable News Network. Turner Broadcasting System, Inc. All Rights Reserved.
Terms of service | Privacy guidelines | Ad choices | Advertise with us | About us | Work for us | Help

CNN en ESPAÑOL | CNN México | CNN Chile | CNN Expansión
العربية | 日本語 | Türkçe
CNN TV | HLN | Transcripts |

<http://web.archive.org/web/20130530221910/http://www.cnn.com/>

Answer: Hi 84 Lo 66 <https://www.timeanddate.com/weather/usa/atlanta/historic?month=5&year=2013>

“No man ever steps in the same river twice, for it's not the same river and he's not the same man”



For third party playback, we are unable to do meaningful audits: replaying the same archived page over and over produces different results.

Left: Reload 1566 archived pages 39 times over 1 year.

Green=resource loaded,
Gray = resource not loaded,
Black line = baseline download.

<https://github.com/oduwsdl/mementos-fixity>

“Hashes are not suitable to verify fixity of the public archived web”

[DOI: 10.1371/journal.pone.0286879](https://doi.org/10.1371/journal.pone.0286879)

We can't depend on the archive for fixity; archives change and/or die. Cf. “Where did the archive go?” (parts [1](#), [2](#), [3](#), [4](#)) &

Some pages change on *every single replay*

```
$ curl -ks
"https://wr.perma-archives.org/public/3phb-8lm6/20190818173517mhttp://web.archive.org/web/20161107234222/https://twitter.com/GenFlynn/status/794000841518776320" | md5
af17fbd9c3b724d5b2e194d02cd1b2d0
$ curl -ks
"https://wr.perma-archives.org/public/3phb-8lm6/20190818173517mp_/http://web.archive.org/web/20161107234222/https://twitter.com/GenFlynn/status/794000841518776320" | md5
af17fbd9c3b724d5b2e194d02cd1b2d0
$ curl -ks
"https://wr.perma-archives.org/public/3phb-8lm6/20190818173517mp_/http://web.archive.org/web/20161107234222/https://twitter.com/GenFlynn/status/794000841518776320" | md5
af17fbd9c3b724d5b2e194d02cd1b2d0
$ curl -ks
"https://wr.perma-archives.org/public/3phb-8lm6/20190818173517mp_/http://web.archive.org/web/20161107234222/https://twitter.com/GenFlynn/status/794000841518776320" | md5
af17fbd9c3b724d5b2e194d02cd1b2d0
$ curl -ks
"https://wr.perma-archives.org/public/3phb-8lm6/20190818173517mp_/http://web.archive.org/web/20161107234222/https://twitter.com/GenFlynn/status/794000841518776320" | md5
af17fbd9c3b724d5b2e194d02cd1b2d0
$ curl -ks "https://wr.perma-archives.org/public/r84b-ljwj/20190613150656mp_/http://csrc.nist.gov/" | md5
4e39ba244179d4d83ca5ffb3e4ec57d2
$ curl -ks "https://wr.perma-archives.org/public/r84b-ljwj/20190613150656mp_/http://csrc.nist.gov/" | md5
d7d0cc09dc94899171604274f9b95a80
$ curl -ks "https://wr.perma-archives.org/public/r84b-ljwj/20190613150656mp_/http://csrc.nist.gov/" | md5
b7445e36e4886b6bfaa98038f87f2808
$ curl -ks "https://wr.perma-archives.org/public/r84b-ljwj/20190613150656mp_/http://csrc.nist.gov/" | md5
2f45763a7b9ef28dc440ebff469703d0
$ curl -ks "https://wr.perma-archives.org/public/r84b-ljwj/20190613150656mp_/http://csrc.nist.gov/" | md5
1fc2381b76795d31aa7c28cea7236dbe
```

This is expected behavior: grab the same HTML, get the same hash

This is unexpected behavior: a different hash for the same HTML *every single time*

Abstracted web archive stack

or: verifying fixity is difficult because archives are *active simulations of the past*



replayed page on your machine

What you experience at this level depends on everything below, which means it will change!

Network Infrastructure

Cloudflare, Akamai

HTTP server

Apache, nginx

Replay Engine

Open Wayback, pywb, replayweb.page

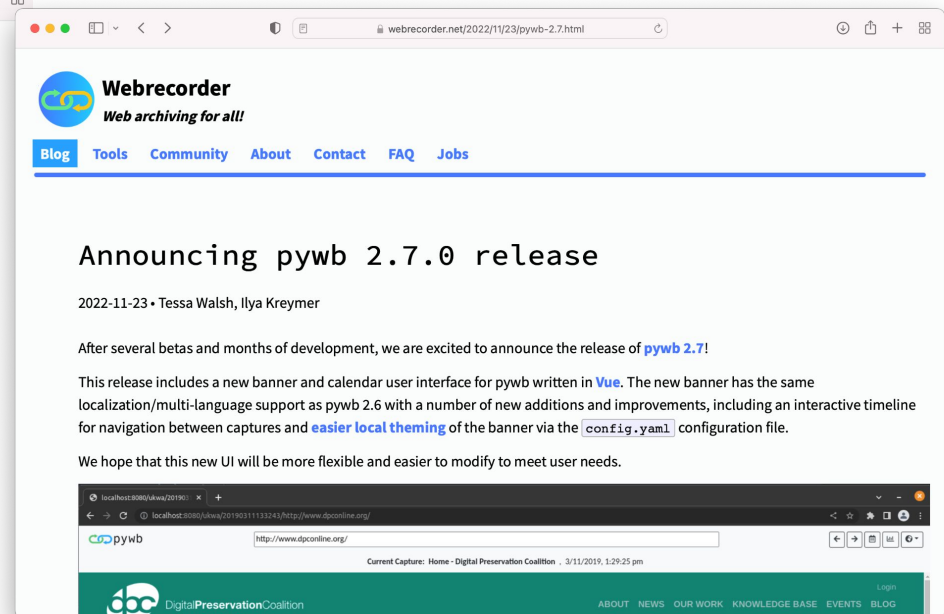
Archived Resources

WARC, WACZ, HAR

The services at these levels *change all the time.*

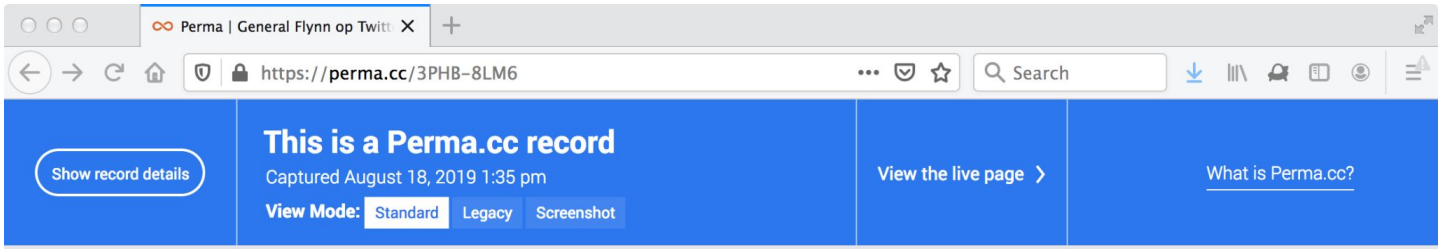
The resources at this level should not change!

The simulation of the past is constantly being improved



Irony: my laptop is no longer new enough to replay this 3 year old page

The latest playback engine does not run on my ancient laptop



Status	Method	Domain	File	Initiator	Type	Transferred	Size	Time
200	GET	perma.cc	3PHB-8LM6	document	html	5.64 KB	12.81 KB	118 ms
200	GET	perma.cc	single-link.css	stylesheet	css	cached	117.37 KB	0 ms
200	GET	perma.cc	single-link.js	script	js	cached	0 B	0 ms
200	GET	replay.perma.cc	/?guid=3PHB-8LM6&embed=replayonly	3PHB-8LM6:207 (su...	html	3.80 KB	9.33 KB	319 ms
200	GET	perma.cc	favicon.ico	FaviconLoader.jsm:1...	x-icon	cached	16.56 KB	0 ms
200	GET	replay.perma.cc	ui.js	script	js	92.57 KB	425.26 KB	81 ms

6 requests | 581.33 KB / 102 KB transferred | Finish: 1.78 s | DOMContentLoaded: 1.11 s | load: 2.22 s

Fortunately "legacy" mode is available

The screenshot shows a web browser window displaying a Perma.cc record. The address bar shows the URL <https://perma.cc/3PHB-8LM6?view-mode=server-side>. The page header is blue and contains the text "This is a Perma.cc record" and "Captured August 18, 2019 1:35 pm". Below this, there are buttons for "View Mode: Standard Legacy Screenshot" and "View the live page >". A "Wayback Machine" logo is visible on the left. The main content is a tweet from General Flynn (@GenFlynn) dated 19:18 - 2 nov. 2016. The tweet text is "U decide - NYPD Blows Whistle on New Hillary Emails: Money Laundering, Sex Crimes w Children, etc...MUST READ! [truepundit.com /breaking-bombs](http://truepundit.com/breaking-bombs) ...". The tweet has 6,959 retweets and 6,385 likes. Below the tweet, there is a reply from Jim (@Jim65783) dated 3 nov. The reply text is "@GenFlynn @DesireeAaron if they had sex with children I say execute the scum.". The page also shows a calendar for November 2016 and a "Go" button.

Transient errors still happen



504 Gateway Time-out

nginx/1.19.5

This is a Perma.cc record
Captured August 18, 2019 1:35 pm
View Mode: Standard Legacy Screenshot

The page isn't redirecting properly
An error occurred during a connection to wr.perma-archives.org.

- This problem can sometimes be caused by disabling or refusing to accept cookies.

[Try Again](#)

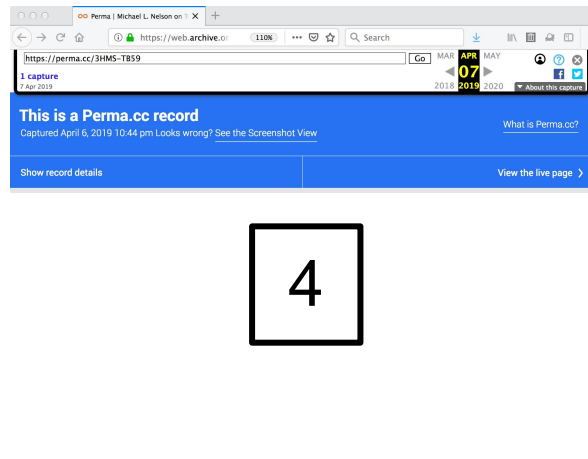
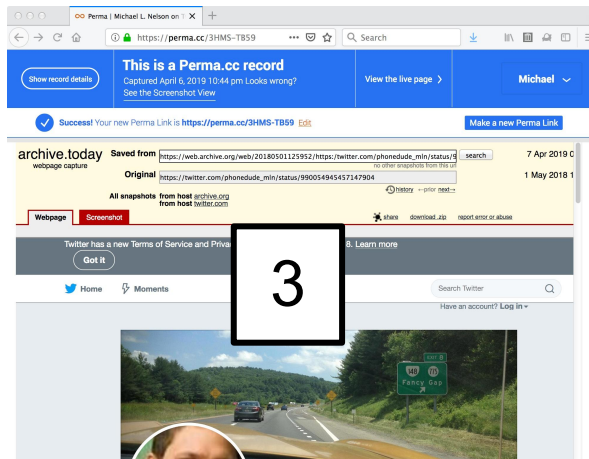
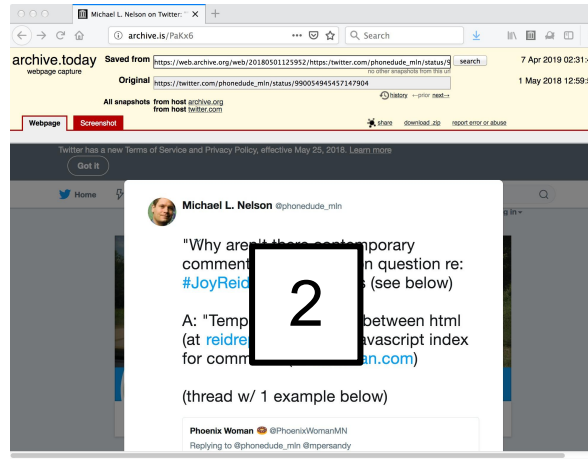
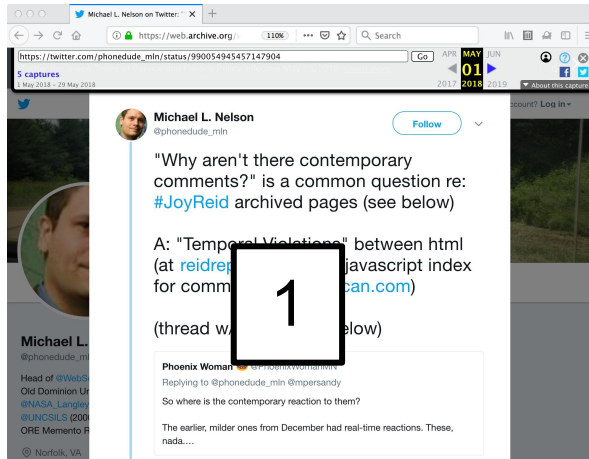
Status	Method	Domain	File	Initiator	Type	Transferred	Size	ms
303	GET	wr.perma-archives...	_set_session?path=/public/3phb-8lm6/20190818173517mp/...	subdocument	html	628 B	0 B	39
303	GET	wr.perma-archives...	794000841518776320	subdocument	html	608 B	0 B	24
302	GET	perma.cc	_set_session?path=/public/3phb-8lm6/20190818173517mp/...	subdocument	html	590 B	0 B	57
303	GET	wr.perma-archives...	_set_session?path=/public/3phb-8lm6/20190818173517mp/...	subdocument	html	492 B	0 B	40
303	GET	wr.perma-archives...	794000841518776320	subdocument	html	608 B	0 B	23
302	GET	perma.cc	_set_session?path=/public/3phb-8lm6/20190818173517mp/...	subdocument	html	590 B	0 B	57
303	GET	wr.perma-archives...	_set_session?path=/public/3phb-8lm6/20190818173517mp/...	subdocument	html	628 B	0 B	41

27 requests | 145.81 KB / 17.95 KB transferred | Finish: 3.13 s | DOMContentLoaded: 707 ms | load: 3.86 s

Hundreds of accessible, publicly available,
independent, interoperable, robust,
auditable, **cooperating** web archives.

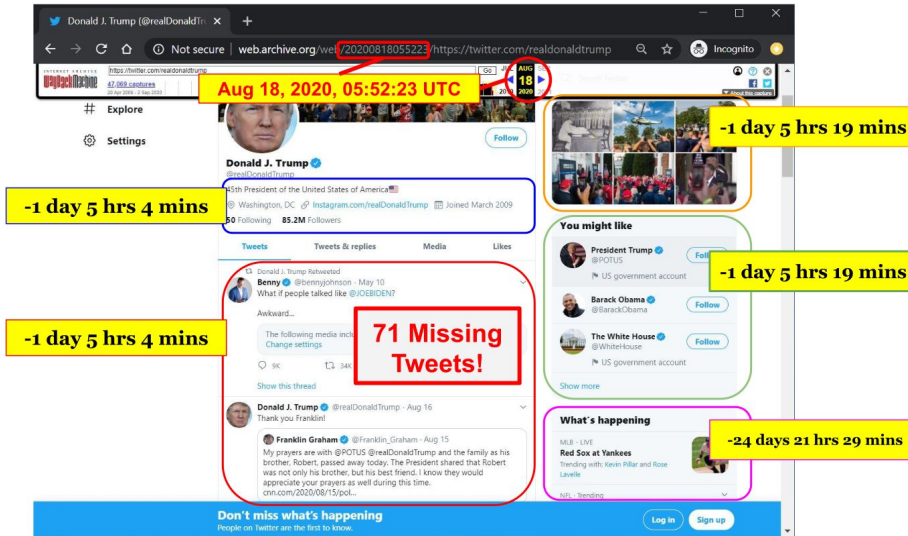
That archives don't ingest the output of other archives is a *lack of interoperability*.

That we're not more concerned about this is a *lack of cooperation*.



<https://www.slideshare.net/phonedude/web-archives-at-the-nexus-of-good-fakes-and-flawed-originals/87>

Archiving the archive, via the replay interface, is critical *because that's what users see*



replay of page, archived on [August 18, 2020](https://web.archive.org/web/20200818055223/https://twitter.com/realdonaldtrump),
<https://web.archive.org/web/20200818055223/https://twitter.com/realdonaldtrump>
in November, 2020 (left) and August, 2022 (right)

The tweets and "What's happening" sections are different,
the media and "You might like" sections are the same.

What will this archived page look like when you click on it?

Kudos to archive.today for preserving machine-readable source metadata and including it in the UI

archive.today
webpage capture

Saved from search 25 Apr 2018 01:58:03 UTC
Original no other snapshots from this url 13 Dec 2005 06:37:57 UTC
All snapshots from host blog.reidreport.com
from host webarchive.loc.gov history ←prior next→

Linked from ja.wikipedia.org » 利用者:Bestlyriccollection
ja.wikipedia.org » 利用者:Bestlyriccollection~jawiki
6 more

Webpage Screenshot share download_zip report bug or abuse donate

LIBRARY OF CONGRESS
Library of Congress Web Archives Collection
This is an archived Web site from the Library of Congress
http://blog.reidreport.com
Archived: 12/13/2005 at 11:37:57
Ask a Librarian

Note: External links, streaming audio/video, forms and search boxes may not function within this web site archive
« Back to Calendar View

« First (12/13/2005) « Previous #1 of 101 Next ▶ Last (03/09/2013) ▶

Blogger SEARCH THIS BLOG SEARCH ALL BLOGS BlogThis! GET YOUR OWN BLOG NEXT BLOG ▶

the reid report reidblog

565 THINK AT YOUR OWN RISK.

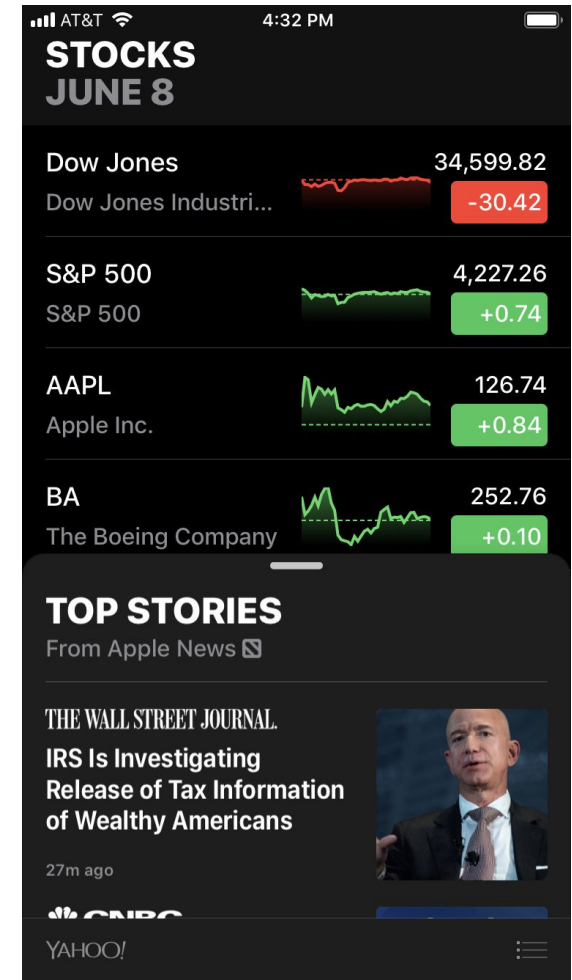
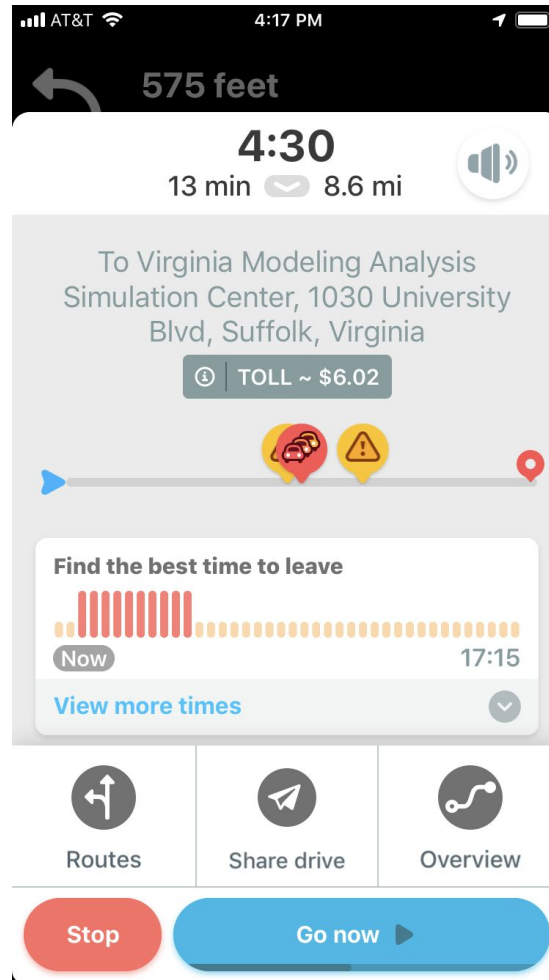
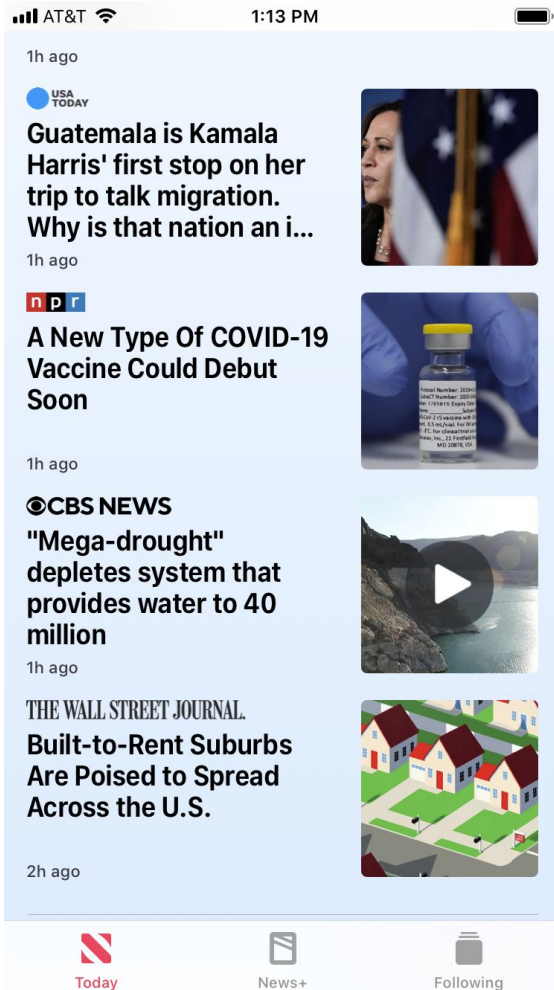
No mercy in Cali. I wonder if anyone will ask the families of the victims if they feel better. Any money says they won't. What a shame.

WEBLOG AWARDS

n.b. tracking source is built-in to [NNTP](#), [SMTP](#), [Atom](#), etc.

Hundreds of accessible, publicly available,
independent, interoperable, robust,
auditable, cooperating **web** archives.

These apps *probably** use HTTP, json, etc.,
but what's their URL? Are they even still *web*?



* I really don't know (WebRTC?). And if they don't, that further proves my point.

Hundreds of accessible, publicly available,
independent, interoperable, robust,
auditable, cooperating web **archives**.

More than just Wayback Machines: we must accommodate any system that supports rehosting and/or revisions

draft-yasskin-wpack-use-case X +
https://datatracker.ietf.org/

[Search] [txt] [html] [xml] [pdf] [bibtex] [Tracker] [Email] [Diff1] [Diff2] [Nits]

Versions: [draft-yasskin-webpackage-use-cases] 00 01 02

Network Working Group J. Yasskin
Internet-Draft Google
Intended status: Informational 13 April 2021
Expires: 15 October 2021

Use Cases and Requirements for Web Packages

draft-yasskin-wpack-use-cases-02

Abstract

This document lists use cases for signing and/or bundling collections of web pages.

Discussion

This note

Discuss mailing list: <https://n>

Source file: <https://c>

Status of This Note

This Internet-Draft

Internet-Draft For Working Group Drafts

Internet-Draft and may be material

This Internet-Draft

Copyright

Wikipedia: National Archives and Records Administration: Revision history - Wikipedia

National Archives and Records Administration: Revision history

View logs for this page (view filter log)

Filter revisions

External tools: Find addition/removal (Alternate) · Find edits by user (Alternate) · Page statistics · Pageviews · Fix dead links

For any version listed below, click on its date to view it.
For more help, see [Help:Page history](#) and [Help:Edit summary](#). (cur) = difference from current version, (prev) = difference from preceding version,
m = minor edit, → = section edit, ← = automatic edit summary
(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) 09:35, 19 August 2022 WikiCleanerBot (talk | contribs) m .. (48,166 bytes) (−1) .. (v2.04b - Bot T20 CWF61 - Fix errors for CW project (Reference before punctuation)) (undo) (Tag: WPCleaner)
- (cur | prev) 02:35, 17 August 2022 DividedFrame (talk | contribs) .. (48,167 bytes) (−27) .. (←Record Keeping Under the Trump Administration: Sub-section title style) (undo)
- (cur | prev) 04:06, 15 August 2022 Nythar (talk | contribs) .. (48,194 bytes) (+3,890) .. (Reverting edit(s) by 24.101.213.206 (talk) to rev. 1104252872 by Misterniceguli: Unexplained

https://github.com/iipc/openwayback

Search or jump to...

Pull requests Issues Marketplace Explore

iipc / openwayback

Code Issues 101 Pull requests 5 Actions Projects Wiki Security Insights

master 12 branches 12 tags

Go to file Add file Code

About

The OpenWayback Development

ldko Merge pull request #435 from iipc/dep... ✓ c7fac11 on Oct 13, 2020 3,399 commits

dist src/site wayback wayback wayback wayback .dockerignore .gitignore .travis.yml CONTRIBUTING

This is historical material "frozen in time". The website is no longer updated and links to external websites and some internal pages may not work.

obamawhitehouse.archives.gov

The White House | whitehouse.gov

the WHITE HOUSE PRESIDENT BARACK OBAMA BRIEFING ROOM ISSUES THE ADMINISTRATION 1600 PENN

Yes, we did. Yes, we can.

"You were the change. You answered people's hopes, and because of you, by almost every measure, America is a better, stronger place than it was when we started."
- President Barack Obama

WATCH THE ADDRESS

AN IN-DEPTH LOOK AT PRESIDENT OBAMA'S RECORD IN FIVE AREAS

HEALTH CARE

see also: <https://www.slideshare.net/ibnesayed/readying-web-archives-to-consume-and-leverage-web-bundles>

Web Archiving in the Year 2030

Hundreds of accessible, publicly available, independent, interoperable, robust, auditable, cooperating web archives.

Can we achieve this by 2030? Yes.
Will we achieve this by 2030? Maybe.
Will we “solve” trust? No.

Technical definitions (e.g., [ISO 16363](#)) notwithstanding, “trust” in web archives might be better understood as analogous to “[relevance](#)” in info retrieval: defined by a user’s information need.