Blockchain Can Not Be Used To Verify Replayed Archived Web Pages

Michael L. Nelson

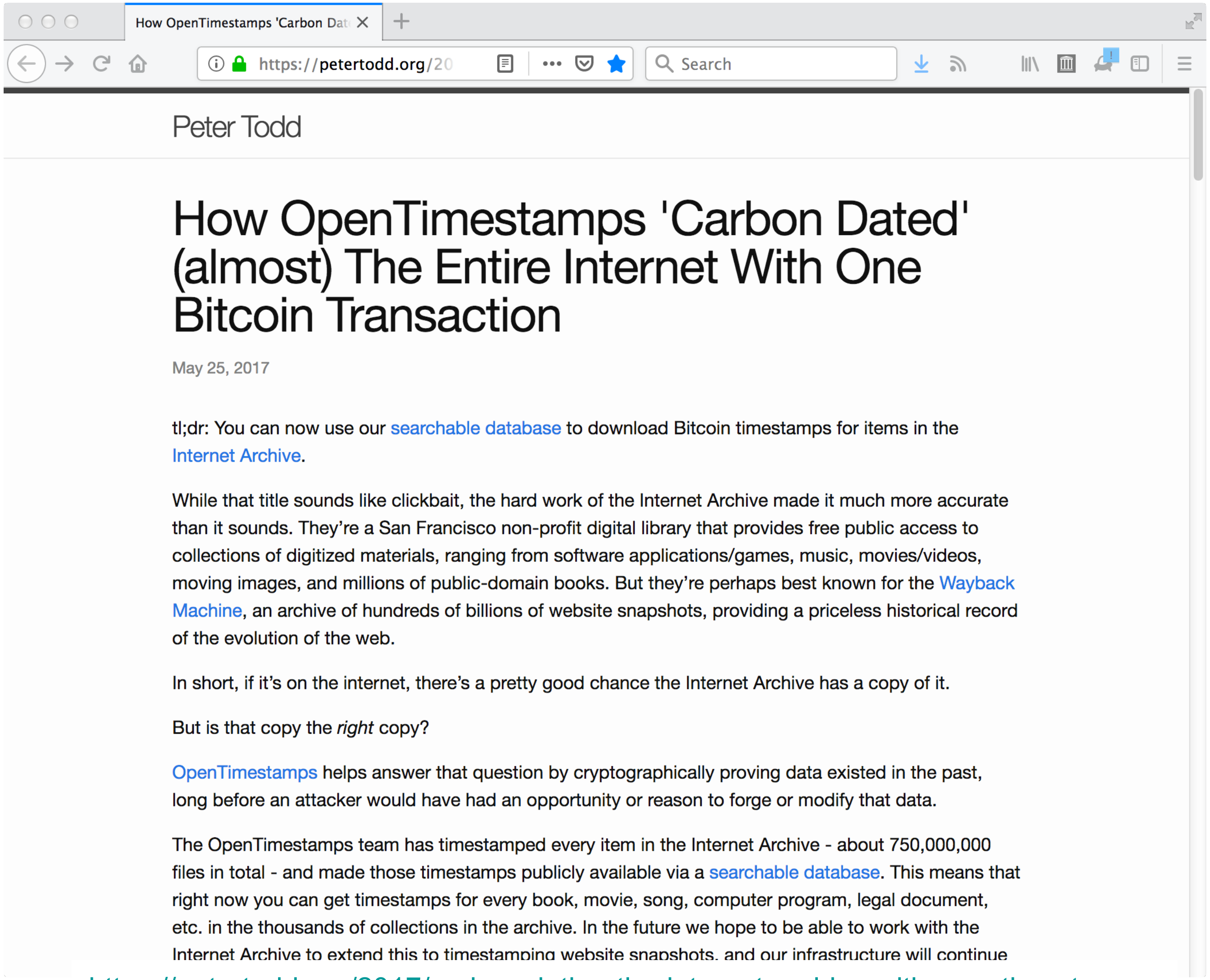
Old Dominion University
Web Science & Digital Libraries Research Group
@WebSciDL, @phonedude_mln

With:
ODU: Michele C. Weigle, Mohamed Aturban
Los Alamos National Laboratory: Herbert Van de Sompel, Martin Klein

Supported in part by The Andrew Mellon Foundation
and the National Science Foundation

CNI Fall 2018 Membership Meeting, 2018-12-11,
@phonedude_mln, @WebSciDL

This is not what you think it is…



Peter Todd

How OpenTimestamps 'Carbon Dated' (almost) The Entire Internet With One Bitcoin Transaction

May 25, 2017

tl;dr: You can now use our searchable database to download Bitcoin timestamps for items in the Internet Archive.

While that title sounds like clickbait, the hard work of the Internet Archive made it much more accurate than it sounds. They're a San Francisco non-profit digital library that provides free public access to collections of digitized materials, ranging from software applications/games, music, movies/videos, moving images, and millions of public-domain books. But they're perhaps best known for the Wayback Machine, an archive of hundreds of billions of website snapshots, providing a priceless historical record of the evolution of the web.

In short, if it's on the internet, there's a pretty good chance the Internet Archive has a copy of it.

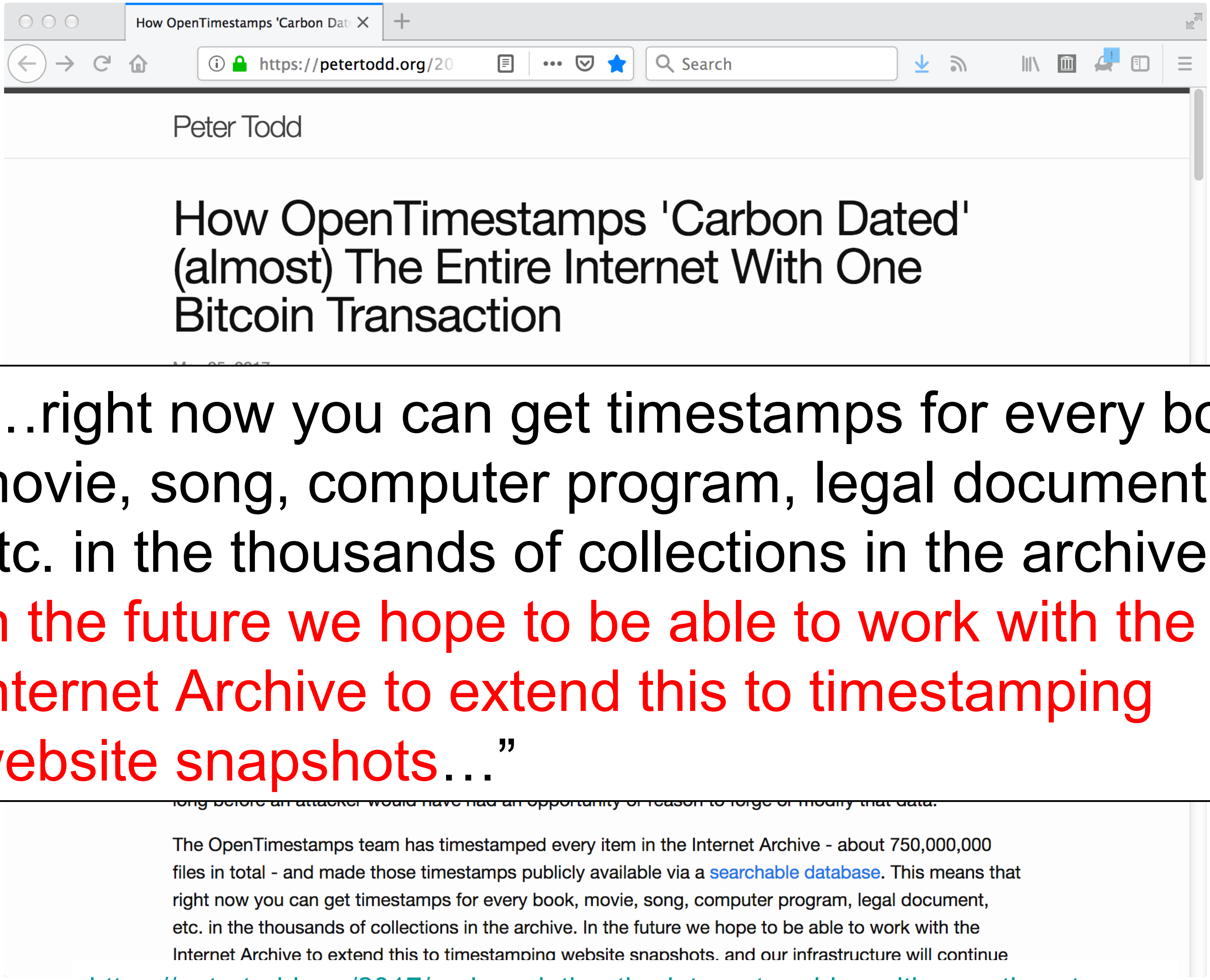
But is that copy the *right* copy?

OpenTimestamps helps answer that question by cryptographically proving data existed in the past, long before an attacker would have had an opportunity or reason to forge or modify that data.

The OpenTimestamps team has timestamped every item in the Internet Archive - about 750,000,000 files in total - and made those timestamps publicly available via a searchable database. This means that right now you can get timestamps for every book, movie, song, computer program, legal document, etc. in the thousands of collections in the archive. In the future we hope to be able to work with the Internet Archive to extend this to timestamping website snapshots, and our infrastructure will continue

https://petertodd.org/2017/carbon-dating-the-internet-archive-with-opentimestamps

This is not what you think it is…



Peter Todd

How OpenTimestamps 'Carbon Dated' (almost) The Entire Internet With One Bitcoin Transaction

long before an attacker would have had an opportunity or reason to forge or modify that data.

The OpenTimestamps team has timestamped every item in the Internet Archive - about 750,000,000 files in total - and made those timestamps publicly available via a searchable database. This means that right now you can get timestamps for every book, movie, song, computer program, legal document, etc. in the thousands of collections in the archive. In the future we hope to be able to work with the Internet Archive to extend this to timestamping website snapshots, and our infrastructure will continue

"…right now you can get timestamps for every book, movie, song, computer program, legal document, etc. in the thousands of collections in the archive. In the future we hope to be able to work with the Internet Archive to extend this to timestamping website snapshots…"

https://petertodd.org/2017/carbon-dating-the-internet-archive-with-opentimestamps

TL;DR

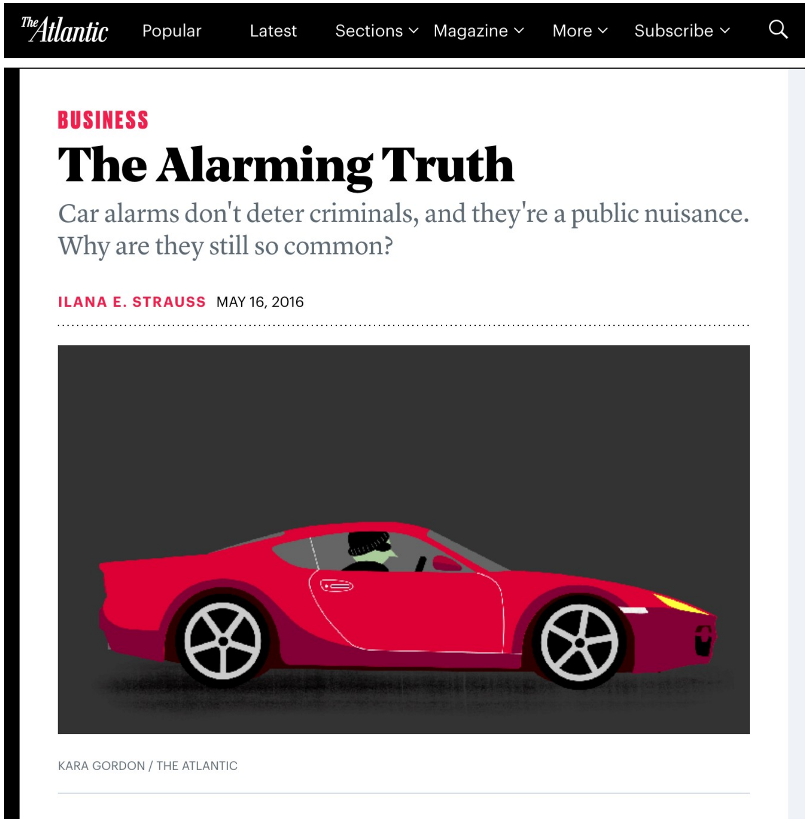
Web archiving is not file backup.

Backup = prevent, detect, repair changes

Web archiving = continuous changes to replicate the past

Naïve fixity techniques are not applicable for web archiving.

Monitoring Fixity To Detect Tampering == Endless False Positives



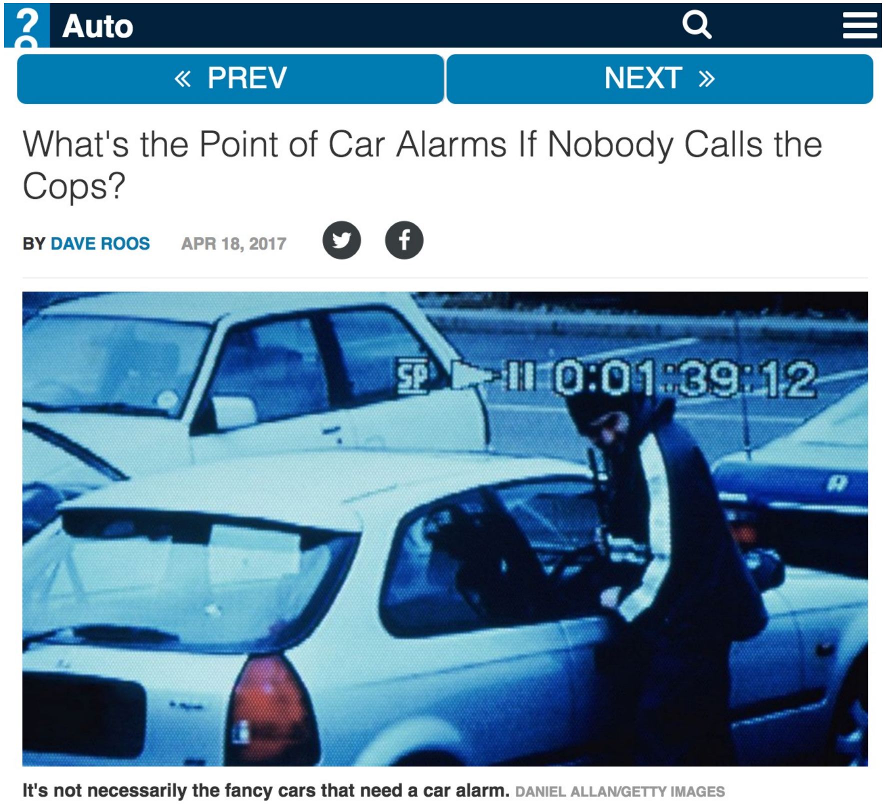
BUSINESS

The Alarming Truth

Car alarms don't deter criminals, and they're a public nuisance. Why are they still so common?

ILANA E. STRAUSS MAY 16, 2016

KARA GORDON / THE ATLANTIC



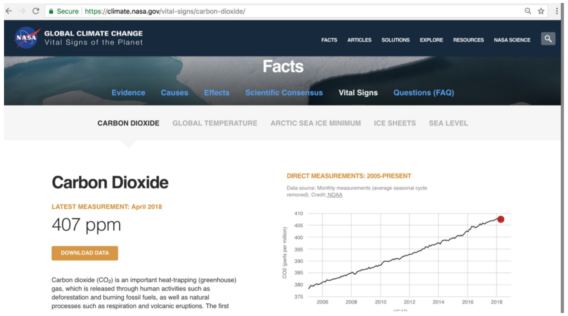
What's the Point of Car Alarms If Nobody Calls the Cops?

BY DAVE ROOS APR 18, 2017

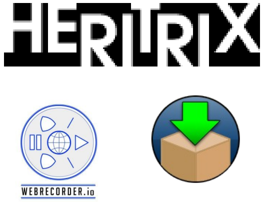
It's not necessarily the fancy cars that need a car alarm. DANIEL ALLAN/GETTY IMAGES

https://www.theatlantic.com/business/archive/2016/05/car-alarms-dont-work-why-so-common/482769/
https://auto.howstuffworks.com/car-driving-safety/safety-regulatory-devices/whats-point-car-alarms-nobody-calls-cops.htm

A simplified workflow of web archiving


https://climate.nasa.gov/vital-signs/carbon-dioxide/
GLOBAL CLIMATE CHANGE
Vital Signs of the Planet
Facts
Carbon Dioxide
LATEST MEASUREMENT: April 2018
407 ppm
DOWNLOAD DATA
DIRECT MEASUREMENTS: 2005-PRESENT

1) live web site
https://climate.nasa.gov/vital-signs/carbon-dioxide/

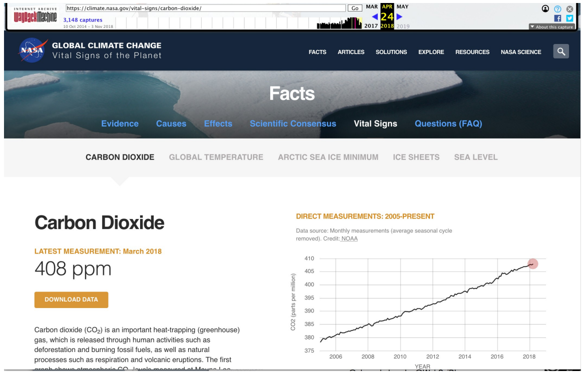


```
$ wget
```

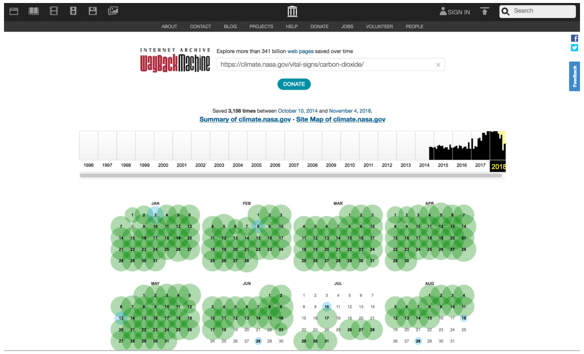
2) Crawled by any of several archival crawlers

```
WARC/1.0
WARC-Type: warcinfo
Content-Type: application/warc-fields
WARC-Date: 2018-11-03T17:20:02Z
WARC-Record-ID: <urn:uuid:6d14bf1d-0ef7-
4f03-9de2-e578d105d3cb>
WARC-Filename: foo.warc.warc.gz
WARC-Block-Digest:
sha1:WWSSYDYY7HTP4JTVOZANSIFPFHUJU64E
Content-Length: 257

software: Wget/1.15 (linux-gnu)
format: WARC File Format 1.0
[much deletia]
```

3) Result stored in a WARC File
(like tar or zip, but for Web archives)


https://climate.nasa.gov/vital-signs/carbon-dioxide/
3,148 captures
MAR APR MAY
24
2017 2018 2019
GLOBAL CLIMATE CHANGE
Vital Signs of the Planet
Facts
Carbon Dioxide
LATEST MEASUREMENT: March 2018
408 ppm
DOWNLOAD DATA
DIRECT MEASUREMENTS: 2005-PRESENT

6) Page replayed with banner, rewritten links, etc.


INTERNET ARCHIVE WaybackMachine
https://climate.nasa.gov/vital-signs/carbon-dioxide/
DONATE
Saved 3,156 times between October 10, 2014 and November 4, 2018.
Summary of climate.nasa.gov · Site Map of climate.nasa.gov
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG

5) User chooses date of capture (Memento-Datetime)

4) WARC files are indexed, served by replay software
(there are several variations of Wayback Machine)

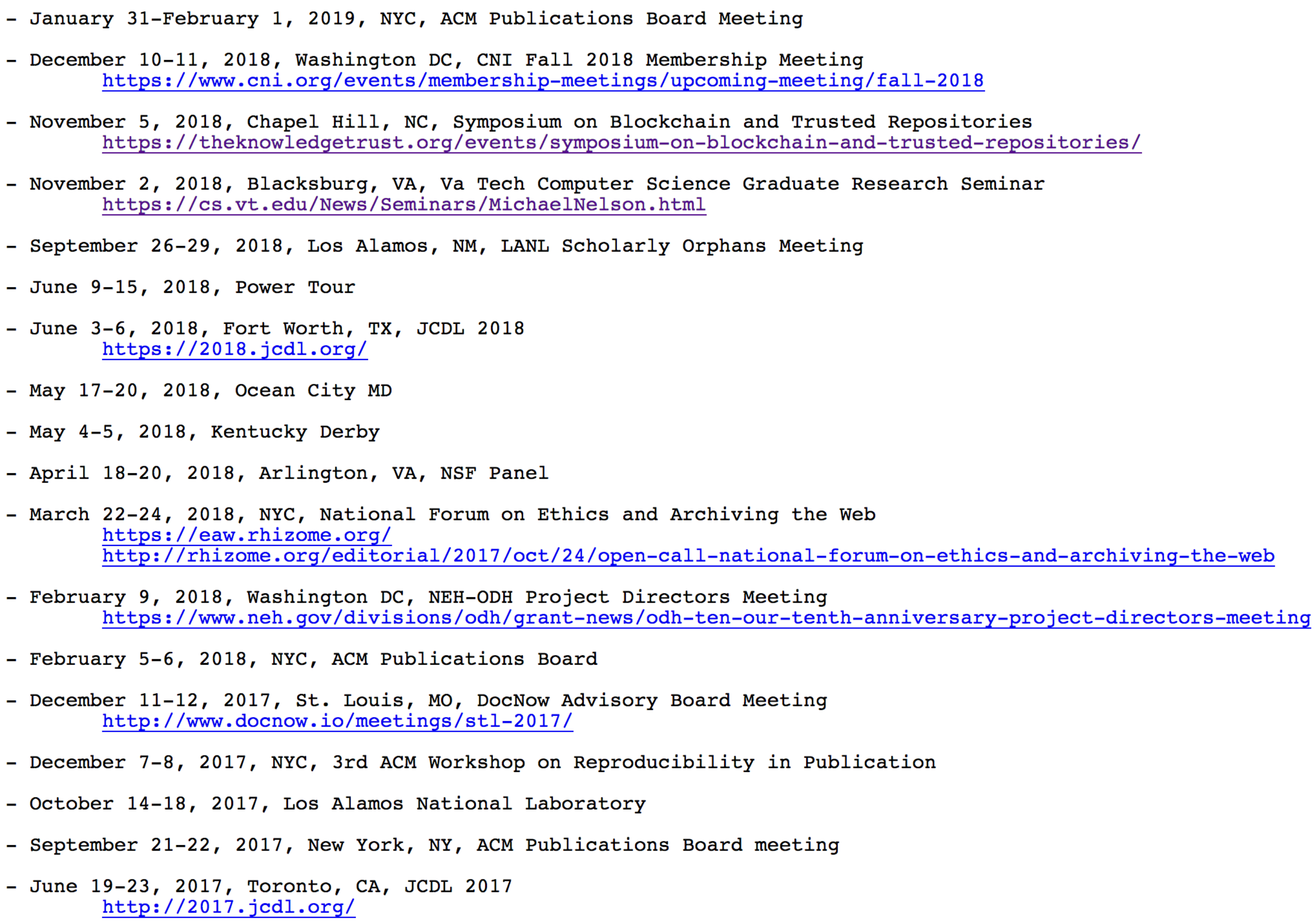
CNI Fall 2018 Membership Meeting, 2018-12-11,
@phonedude_mln, @WebSciDL

(apologies to Peter Arnett)

“In order to save the page, we had to completely change it”

Yes, some archives (including most versions of Wayback) provide “raw” access, but modifications can still happen (how/why is beyond the scope of this presentation).

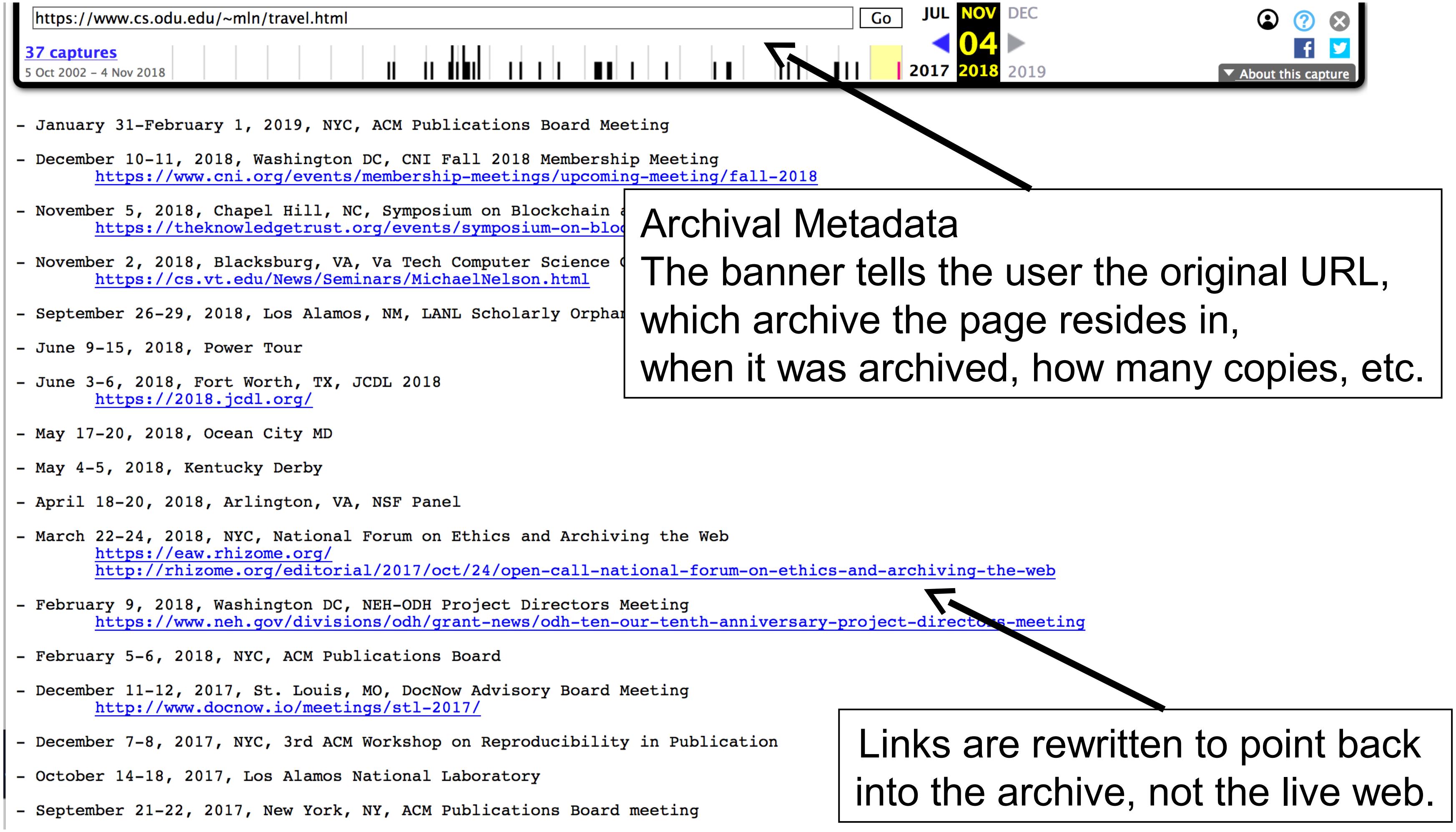
I’ve got mad HTML skillz



- January 31-February 1, 2019, NYC, ACM Publications Board Meeting
- December 10-11, 2018, Washington DC, CNI Fall 2018 Membership Meeting
  https://www.cni.org/events/membership-meetings/upcoming-meeting/fall-2018
- November 5, 2018, Chapel Hill, NC, Symposium on Blockchain and Trusted Repositories
  https://theknowledgetrust.org/events/symposium-on-blockchain-and-trusted-repositories/
- November 2, 2018, Blacksburg, VA, Va Tech Computer Science Graduate Research Seminar
  https://cs.vt.edu/News/Seminars/MichaelNelson.html
- September 26-29, 2018, Los Alamos, NM, LANL Scholarly Orphans Meeting
- June 9-15, 2018, Power Tour
- June 3-6, 2018, Fort Worth, TX, JCDL 2018
  https://2018.jcdl.org/
- May 17-20, 2018, Ocean City MD
- May 4-5, 2018, Kentucky Derby
- April 18-20, 2018, Arlington, VA, NSF Panel
- March 22-24, 2018, NYC, National Forum on Ethics and Archiving the Web
  https://eaw.rhizome.org/
  http://rhizome.org/editorial/2017/oct/24/open-call-national-forum-on-ethics-and-archiving-the-web
- February 9, 2018, Washington DC, NEH-ODH Project Directors Meeting
  https://www.neh.gov/divisions/odh/grant-news/odh-ten-our-tenth-anniversary-project-directors-meeting
- February 5-6, 2018, NYC, ACM Publications Board
- December 11-12, 2017, St. Louis, MO, DocNow Advisory Board Meeting
  http://www.docnow.io/meetings/stl-2017/
- December 7-8, 2017, NYC, 3rd ACM Workshop on Reproducibility in Publication
- October 14-18, 2017, Los Alamos National Laboratory
- September 21-22, 2017, New York, NY, ACM Publications Board meeting
- June 19-23, 2017, Toronto, CA, JCDL 2017
  http://2017.jcdl.org/

https://www.cs.odu.edu/~mln/travel.html

Same page, archived at IA



https://www.cs.odu.edu/~mln/travel.html

- January 31-February 1, 2019, NYC, ACM Publications Board Meeting

- December 10-11, 2018, Washington DC, CNI Fall 2018 Membership Meeting
  https://www.cni.org/events/membership-meetings/upcoming-meeting/fall-2018

- November 5, 2018, Chapel Hill, NC, Symposium on Blockchain a
  https://theknowledgetrust.org/events/symposium-on-bloc

- November 2, 2018, Blacksburg, VA, Va Tech Computer Science C
  https://cs.vt.edu/News/Seminars/MichaelNelson.html

- September 26-29, 2018, Los Alamos, NM, LANL Scholarly Orphan

- June 9-15, 2018, Power Tour

- June 3-6, 2018, Fort Worth, TX, JCDL 2018
  https://2018.jcdl.org/

- May 17-20, 2018, Ocean City MD

- May 4-5, 2018, Kentucky Derby

- April 18-20, 2018, Arlington, VA, NSF Panel

- March 22-24, 2018, NYC, National Forum on Ethics and Archiving the Web
  https://eaw.rhizome.org/
  http://rhizome.org/editorial/2017/oct/24/open-call-national-forum-on-ethics-and-archiving-the-web

- February 9, 2018, Washington DC, NEH-ODH Project Directors Meeting
  https://www.neh.gov/divisions/odh/grant-news/odh-ten-our-tenth-anniversary-project-directors-meeting

- February 5-6, 2018, NYC, ACM Publications Board

- December 11-12, 2017, St. Louis, MO, DocNow Advisory Board Meeting
  http://www.docnow.io/meetings/stl-2017/

- December 7-8, 2017, NYC, 3rd ACM Workshop on Reproducibility in Publication

- October 14-18, 2017, Los Alamos National Laboratory

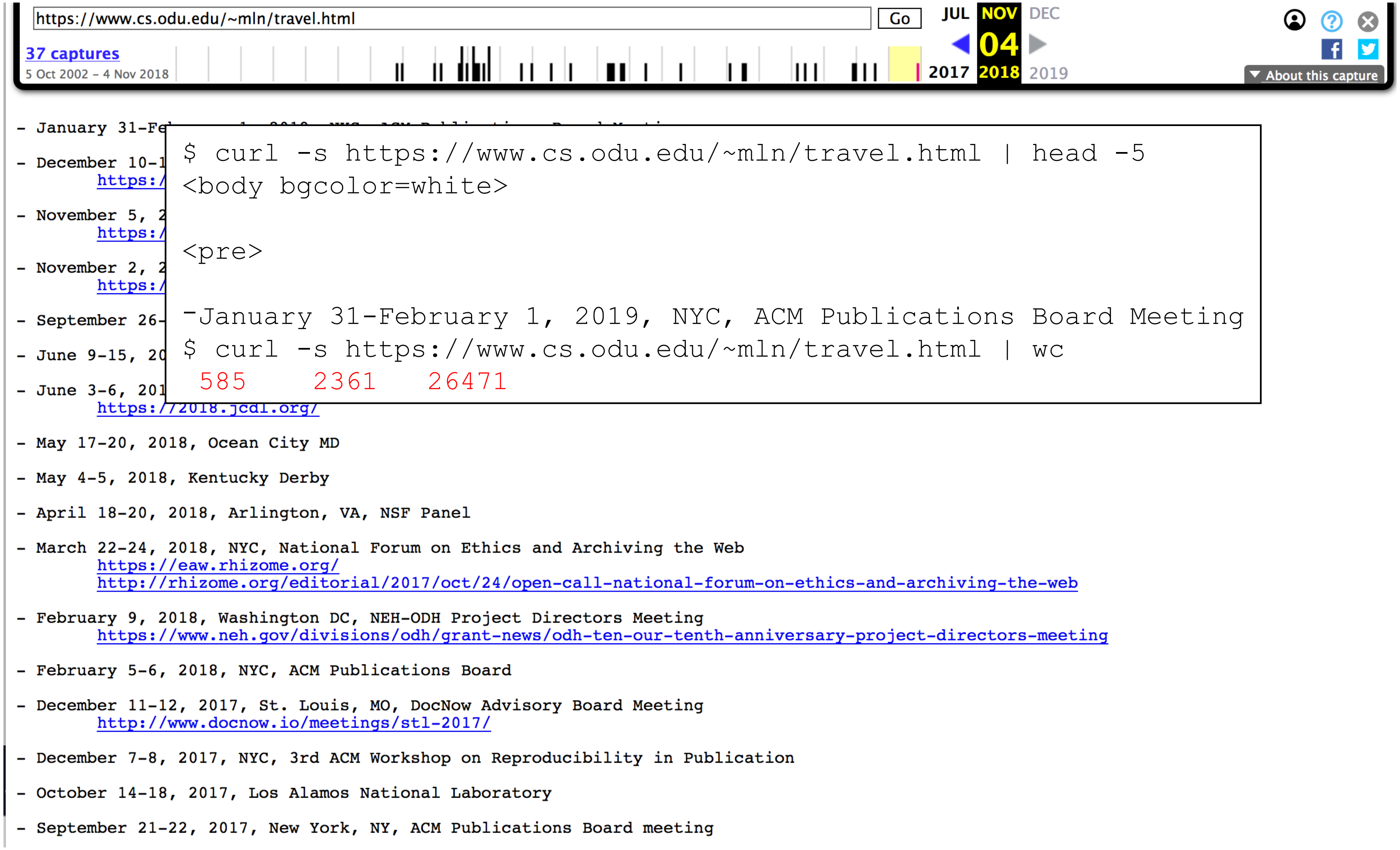
- September 21-22, 2017, New York, NY, ACM Publications Board meeting

Archival Metadata
The banner tells the user the original URL,
which archive the page resides in,
when it was archived, how many copies, etc.

Links are rewritten to point back
into the archive, not the live web.

https://web.archive.org/web/20181104174441/https://www.cs.odu.edu/~mln/travel.html

Same page, archived at IA



- January 31-Fe

- December 10-1
  https:/

- November 5, 2
  https:/

- November 2, 2
  https:/

- September 26-

- June 9-15, 20

- June 3-6, 201
  https://2018.jcdl.org/

- May 17-20, 2018, Ocean City MD

- May 4-5, 2018, Kentucky Derby

- April 18-20, 2018, Arlington, VA, NSF Panel

- March 22-24, 2018, NYC, National Forum on Ethics and Archiving the Web
  https://eaw.rhizome.org/
  http://rhizome.org/editorial/2017/oct/24/open-call-national-forum-on-ethics-and-archiving-the-web

- February 9, 2018, Washington DC, NEH-ODH Project Directors Meeting
  https://www.neh.gov/divisions/odh/grant-news/odh-ten-our-tenth-anniversary-project-directors-meeting

- February 5-6, 2018, NYC, ACM Publications Board

- December 11-12, 2017, St. Louis, MO, DocNow Advisory Board Meeting
  http://www.docnow.io/meetings/stl-2017/

- December 7-8, 2017, NYC, 3rd ACM Workshop on Reproducibility in Publication

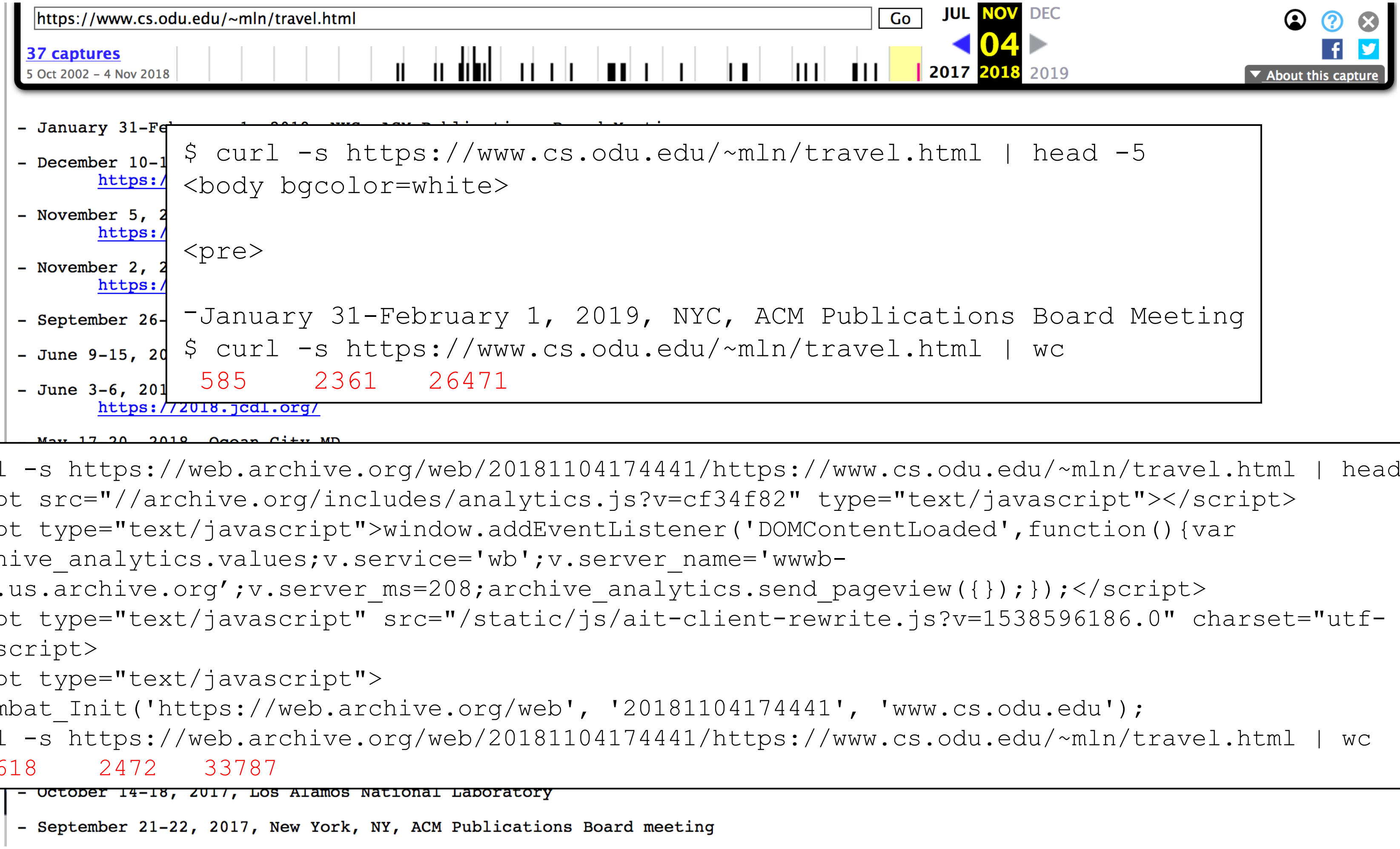
- October 14-18, 2017, Los Alamos National Laboratory

- September 21-22, 2017, New York, NY, ACM Publications Board meeting

```
$ curl -s https://www.cs.odu.edu/~mln/travel.html | head -5
<body bgcolor=white>

<pre>

-January 31-February 1, 2019, NYC, ACM Publications Board Meeting
$ curl -s https://www.cs.odu.edu/~mln/travel.html | wc
 585    2361   26471
```

https://web.archive.org/web/20181104174441/https://www.cs.odu.edu/~mln/travel.html

Same page, archived at IA



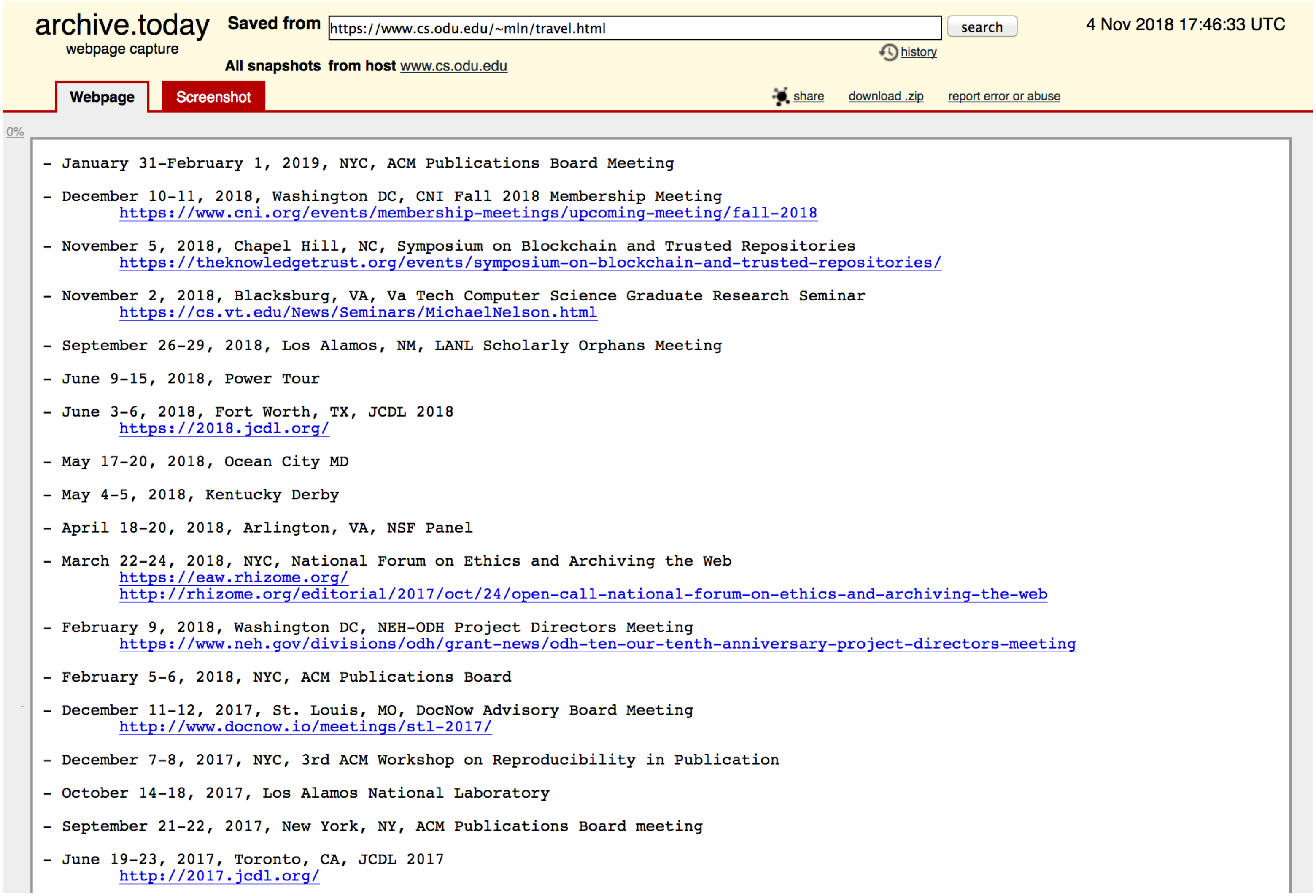
- January 31-Fe
- December 10-1
  https:/
- November 5, 2
  https:/
- November 2, 2
  https:/
- September 26-
- June 9-15, 20
- June 3-6, 201
  https://2018.jcdl.org/
- May 17-20, 2018, Ocean City, MD
- October 14-18, 2017, Los Alamos National Laboratory
- September 21-22, 2017, New York, NY, ACM Publications Board meeting

```
$ curl -s https://www.cs.odu.edu/~mln/travel.html | head -5
<body bgcolor=white>

<pre>

-January 31-February 1, 2019, NYC, ACM Publications Board Meeting
$ curl -s https://www.cs.odu.edu/~mln/travel.html | wc
 585     2361    26471
```

```
$ curl -s https://web.archive.org/web/20181104174441/https://www.cs.odu.edu/~mln/travel.html | head -5
<script src="//archive.org/includes/analytics.js?v=cf34f82" type="text/javascript"></script>
<script type="text/javascript">window.addEventListener('DOMContentLoaded',function(){var
v=archive_analytics.values;v.service='wb';v.server_name='wwwb-
app40.us.archive.org';v.server_ms=208;archive_analytics.send_pageview({});});</script>
<script type="text/javascript" src="/static/js/ait-client-rewrite.js?v=1538596186.0" charset="utf-
8"></script>
<script type="text/javascript">
WB_wombat_Init('https://web.archive.org/web', '20181104174441', 'www.cs.odu.edu');
$ curl -s https://web.archive.org/web/20181104174441/https://www.cs.odu.edu/~mln/travel.html | wc
     618     2472    33787
```

https://web.archive.org/web/20181104174441/https://www.cs.odu.edu/~mln/travel.html

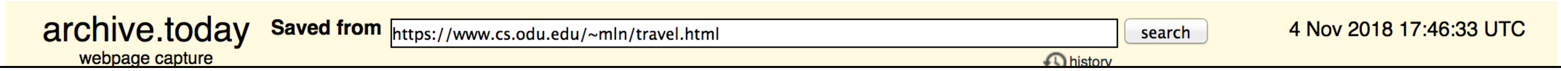
Same page, archived at archive.today



```
- January 31-February 1, 2019, NYC, ACM Publications Board Meeting

- December 10-11, 2018, Washington DC, CNI Fall 2018 Membership Meeting
        https://www.cni.org/events/membership-meetings/upcoming-meeting/fall-2018

- November 5, 2018, Chapel Hill, NC, Symposium on Blockchain and Trusted Repositories
        https://theknowledgetrust.org/events/symposium-on-blockchain-and-trusted-repositories/

- November 2, 2018, Blacksburg, VA, Va Tech Computer Science Graduate Research Seminar
        https://cs.vt.edu/News/Seminars/MichaelNelson.html

- September 26-29, 2018, Los Alamos, NM, LANL Scholarly Orphans Meeting

- June 9-15, 2018, Power Tour

- June 3-6, 2018, Fort Worth, TX, JCDL 2018
        https://2018.jcdl.org/

- May 17-20, 2018, Ocean City MD

- May 4-5, 2018, Kentucky Derby

- April 18-20, 2018, Arlington, VA, NSF Panel

- March 22-24, 2018, NYC, National Forum on Ethics and Archiving the Web
        https://eaw.rhizome.org/
        http://rhizome.org/editorial/2017/oct/24/open-call-national-forum-on-ethics-and-archiving-the-web

- February 9, 2018, Washington DC, NEH-ODH Project Directors Meeting
        https://www.neh.gov/divisions/odh/grant-news/odh-ten-our-tenth-anniversary-project-directors-meeting

- February 5-6, 2018, NYC, ACM Publications Board

- December 11-12, 2017, St. Louis, MO, DocNow Advisory Board Meeting
        http://www.docnow.io/meetings/stl-2017/

- December 7-8, 2017, NYC, 3rd ACM Workshop on Reproducibility in Publication

- October 14-18, 2017, Los Alamos National Laboratory

- September 21-22, 2017, New York, NY, ACM Publications Board meeting

- June 19-23, 2017, Toronto, CA, JCDL 2017
        http://2017.jcdl.org/
```

http://archive.is/20181104174633/https://www.cs.odu.edu/~mln/travel.html

Same page, archived at archive.today



archive.today webpage capture

Saved from https://www.cs.odu.edu/~mln/travel.html search

4 Nov 2018 17:46:33 UTC

```
$ curl -s http://archive.is/20181104174633/https://www.cs.odu.edu/~mln/travel.html | head -5
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"><html style="background-
color:#EEEEEE" prefix="og: http://ogp.me/ns# article: http://ogp.me/ns/article#" itemscope
itemtype="http://schema.org/Article"><!--174.109.72.208--><!--curl/7.30.0--><head><meta http-
equiv="Content-Type" content="text/html;charset=utf-8"/><meta name="robots"
content="index,noarchive"/><meta name="viewport" content="device-width=300, initial-scale=1"/><meta
property="twitter:card" content="summary"/><meta property="twitter:site" content="@archiveis"/><meta
property="og:type" content="article"/><meta property="og:site_name" content="archive.is"/><meta
property="og:url" content="http://archive.is/l6QdV" itemprop="url"/><meta property="og:title"
content="https://www.cs.odu.edu/~mln/travel.html"/><meta property="twitter:title"
content="https://www.cs.odu.edu/~mln/travel.html"/><meta property="twitter:description"
content="archived 4 Nov 2018 17:46:33 UTC" itemprop="description"/><meta
property="article:published_time" content="2018-11-04T17:46:33Z" itemprop="dateCreated"/><meta
property="article:modified_time" content="2018-11-04T17:46:33Z" itemprop="dateModified"/><link
rel="image_src"
href="https://archive.is/l6QdV/d7e3acef18a0433590880dfcc26f8e1f5f18f91e/scr.png"/><meta
property="og:image"
content="https://archive.is/l6QdV/d7e3acef18a0433590880dfcc26f8e1f5f18f91e/scr.png"
itemprop="image"/><meta property="twitter:image"
content="https://archive.is/l6QdV/d7e3acef18a0433590880dfcc26f8e1f5f18f91e/scr.png"/><meta
property="twitter:image:src"
content="https://archive.is/l6QdV/d7e3acef18a0433590880dfcc26f8e1f5f18f91e/scr.png"/><meta
property="twitter:image:width" content="1024"/><meta property="twitter:image:height"
content="768"/><link rel="icon" href="//www.google.com/s2/favicons?domain=www.cs.odu.edu"/><link
rel="canonical" href="https://archive.is/l6QdV"/><link rel="bookmark"
href="http://archive.today/20181104174633/https://www.cs.odu.edu/~mln/travel.html"/><title></title><
/head><body style="margin:0;background-color:#EEEEEE"><center><div id="HEADER" style="font-
family:sans-serif;background
[much deletia - you get the point]
$ curl -s http://archive.is/20181104174633/https://www.cs.odu.edu/~mln/travel.html | wc
     730    3640   62392
```

If we just had isolated, static pages
(e.g., individual jpegs, pdfs, mp3s)
then there'd be no problem.

But HTML has:
1) links,
2) embedded resources (including iframes), and
3) Javascript, which can modify the HTML.

And HTTP has no "bulk download",
so you can't grab an entire site instantaneously.

We could hash the WARC file

```
WARC/1.0
WARC-Type: warcinfo
Content-Type: application/warc-fields
WARC-Date: 2018-11-03T17:20:02Z
WARC-Record-ID: <urn:uuid:6d14bf1d-0ef7-4f03-9de2-e578d105d3cb>
WARC-Filename: climate.nasa.gov.warc.gz
WARC-Block-Digest: sha1:WWSSYDYY7HTP4JTVOZANSIFPFHUJU64E
Content-Length: 257

software: Wget/1.15 (linux-gnu)
format: WARC File Format 1.0
conformsTo: http://bibnum.bnf.fr/WARC/WARC_ISO_28500_version1_latestdraft.pdf
robots: classic
wget-arguments: "--warc-file=climate.nasa.gov" "https://climate.nasa.gov/vital-signs/carbon-dioxide/"
```

```
$ md5sum climate.nasa.gov.warc.gz
652853fe1bc8cb273cdf73aad8a489ca  climate.nasa.gov.warc.gz
```

```
WARC/1.0
WARC-Type: request
WARC-Target-URI: https://climate.nasa.gov/vital-signs/carbon-dioxide/
Content-Type: application/http;msgtype=request
WARC-Date: 2018-11-03T17:20:02Z
WARC-Record-ID: <urn:uuid:e44bc1ea-61a1-4200-b94f-60042456f638>
WARC-IP-Address: 54.230.195.16
WARC-Warcinfo-ID: <urn:uuid:6d14bf1d-0ef7-4f03-9de2-e578d105d3cb>
WARC-Block-Digest: sha1:CLODKYDXCHPVOJMJWHJVT3EJJDKI2RTQ
Content-Length: 141

GET /vital-signs/carbon-dioxide/ HTTP/1.1
User-Agent: Wget/1.15 (linux-gnu)
Accept: */*
Host: climate.nasa.gov
Connection: Keep-Alive


WARC/1.0
WARC-Type: response
WARC-Record-ID: <urn:uuid:5d8861ef-93c5-4d9c-87b8-4f427f963f7c>
WARC-Warcinfo-ID: <urn:uuid:6d14bf1d-0ef7-4f03-9de2-e578d105d3cb>
WARC-Concurrent-To: <urn:uuid:e44bc1ea-61a1-4200-b94f-60042456f638>
WARC-Target-URI: https://climate.nasa.gov/vital-signs/carbon-dioxide/
WARC-Date: 2018-11-03T17:20:02Z
```

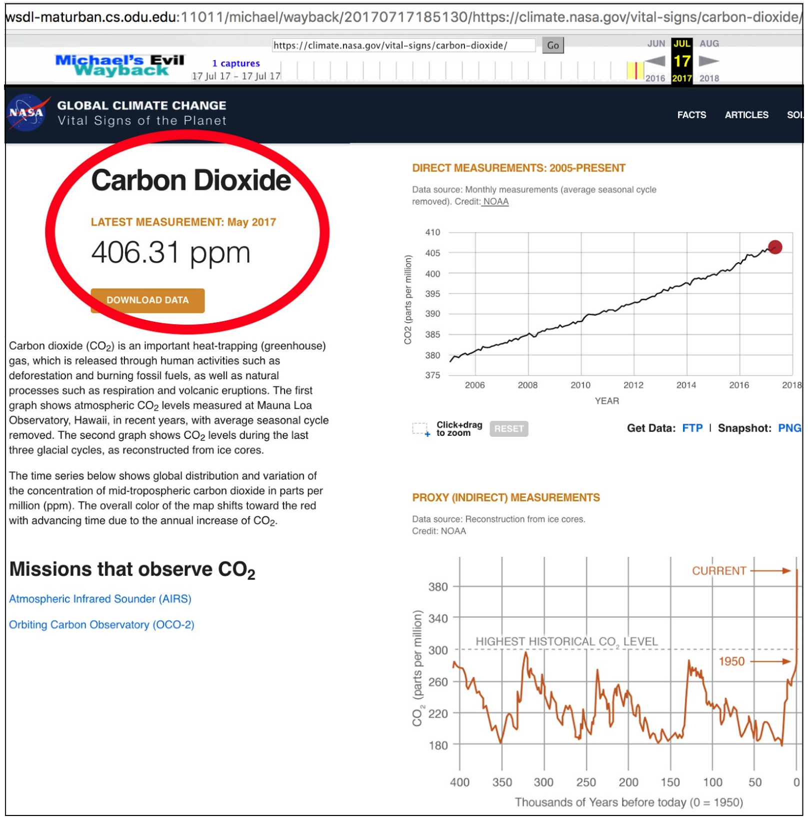
But this nasa.gov page contains:

- 201 images
- 19 Javascript files
- 3 CSS files

At a large archive like IA they could be in multiple WARC files; worst case is 224 WARC files.

In general, the WARC file(s) corresponding to the replayed page will be unavailable to the user replaying the page.

We can detect changes in the root HTML



wsdl-maturban.cs.odu.edu:11011/michael/wayback/20170717185130/https://climate.nasa.gov/vital-signs/carbon-dioxide/

Carbon Dioxide

LATEST MEASUREMENT: May 2017

406.31 ppm

DOWNLOAD DATA

Carbon dioxide (CO_2) is an important heat-trapping (greenhouse) gas, which is released through human activities such as deforestation and burning fossil fuels, as well as natural processes such as respiration and volcanic eruptions. The first graph shows atmospheric CO_2 levels measured at Mauna Loa Observatory, Hawaii, in recent years, with average seasonal cycle removed. The second graph shows CO_2 levels during the last three glacial cycles, as reconstructed from ice cores.

The time series below shows global distribution and variation of the concentration of mid-tropospheric carbon dioxide in parts per million (ppm). The overall color of the map shifts toward the red with advancing time due to the annual increase of CO_2.

Missions that observe CO_2

Atmospheric Infrared Sounder (AIRS)

Orbiting Carbon Observatory (OCO-2)

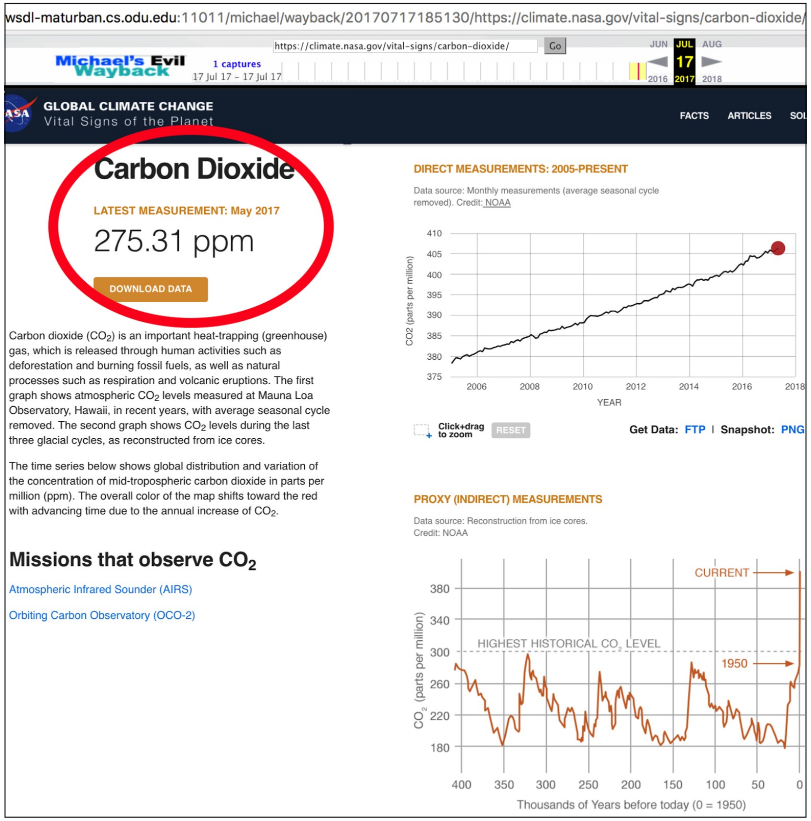
DIRECT MEASUREMENTS: 2005-PRESENT

Data source: Monthly measurements (average seasonal cycle removed). Credit: NOAA

Get Data: FTP | Snapshot: PNG

PROXY (INDIRECT) MEASUREMENTS

Data source: Reconstruction from ice cores. Credit: NOAA



wsdl-maturban.cs.odu.edu:11011/michael/wayback/20170717185130/https://climate.nasa.gov/vital-signs/carbon-dioxide/

Carbon Dioxide

LATEST MEASUREMENT: May 2017

275.31 ppm

DOWNLOAD DATA

Carbon dioxide (CO_2) is an important heat-trapping (greenhouse) gas, which is released through human activities such as deforestation and burning fossil fuels, as well as natural processes such as respiration and volcanic eruptions. The first graph shows atmospheric CO_2 levels measured at Mauna Loa Observatory, Hawaii, in recent years, with average seasonal cycle removed. The second graph shows CO_2 levels during the last three glacial cycles, as reconstructed from ice cores.

The time series below shows global distribution and variation of the concentration of mid-tropospheric carbon dioxide in parts per million (ppm). The overall color of the map shifts toward the red with advancing time due to the annual increase of CO_2.

Missions that observe CO_2

Atmospheric Infrared Sounder (AIRS)

Orbiting Carbon Observatory (OCO-2)

DIRECT MEASUREMENTS: 2005-PRESENT

Data source: Monthly measurements (average seasonal cycle removed). Credit: NOAA

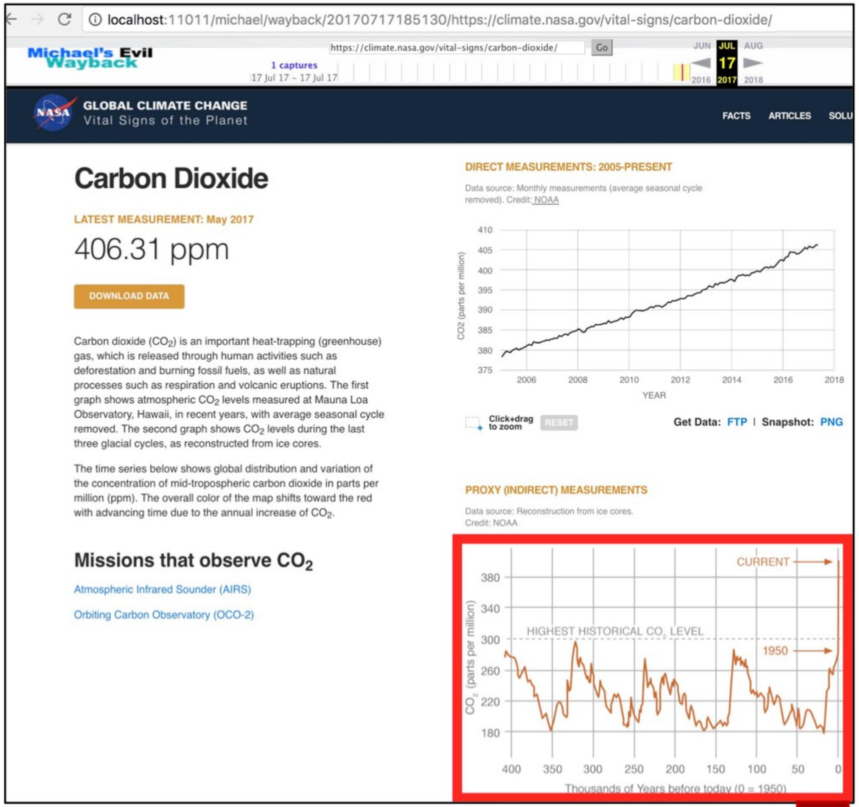
Get Data: FTP | Snapshot: PNG

PROXY (INDIRECT) MEASUREMENTS

Data source: Reconstruction from ice cores. Credit: NOAA

https://ws-dl.blogspot.com/2017/12/2017-12-11-difficulties-in-timestamping.html

But what if the change is in an embedded resource?



localhost:11011/michael/wayback/20170717185130/https://climate.nasa.gov/vital-signs/carbon-dioxide/

Carbon Dioxide

LATEST MEASUREMENT: May 2017

406.31 ppm

DOWNLOAD DATA

Carbon dioxide (CO_2) is an important heat-trapping (greenhouse) gas, which is released through human activities such as deforestation and burning fossil fuels, as well as natural processes such as respiration and volcanic eruptions. The first graph shows atmospheric CO_2 levels measured at Mauna Loa Observatory, Hawaii, in recent years, with average seasonal cycle removed. The second graph shows CO_2 levels during the last three glacial cycles, as reconstructed from ice cores.

The time series below shows global distribution and variation of the concentration of mid-tropospheric carbon dioxide in parts per million (ppm). The overall color of the map shifts toward the red with advancing time due to the annual increase of CO_2.

Missions that observe CO_2

Atmospheric Infrared Sounder (AIRS)

Orbiting Carbon Observatory (OCO-2)

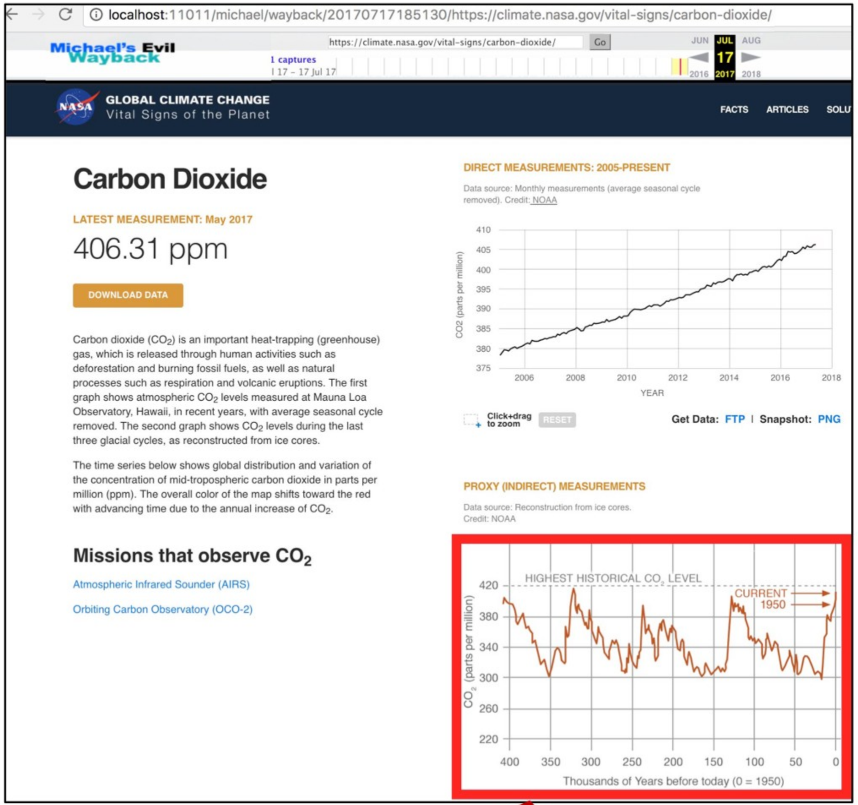
DIRECT MEASUREMENTS: 2005-PRESENT

Data source: Monthly measurements (average seasonal cycle removed). Credit: NOAA

Get Data: FTP | Snapshot: PNG

PROXY (INDIRECT) MEASUREMENTS

Data source: Reconstruction from ice cores.
Credit: NOAA



localhost:11011/michael/wayback/20170717185130/https://climate.nasa.gov/vital-signs/carbon-dioxide/

Carbon Dioxide

LATEST MEASUREMENT: May 2017

406.31 ppm

DOWNLOAD DATA

Carbon dioxide (CO_2) is an important heat-trapping (greenhouse) gas, which is released through human activities such as deforestation and burning fossil fuels, as well as natural processes such as respiration and volcanic eruptions. The first graph shows atmospheric CO_2 levels measured at Mauna Loa Observatory, Hawaii, in recent years, with average seasonal cycle removed. The second graph shows CO_2 levels during the last three glacial cycles, as reconstructed from ice cores.

The time series below shows global distribution and variation of the concentration of mid-tropospheric carbon dioxide in parts per million (ppm). The overall color of the map shifts toward the red with advancing time due to the annual increase of CO_2.

Missions that observe CO_2

Atmospheric Infrared Sounder (AIRS)

Orbiting Carbon Observatory (OCO-2)

DIRECT MEASUREMENTS: 2005-PRESENT

Data source: Monthly measurements (average seasonal cycle removed). Credit: NOAA

Get Data: FTP | Snapshot: PNG

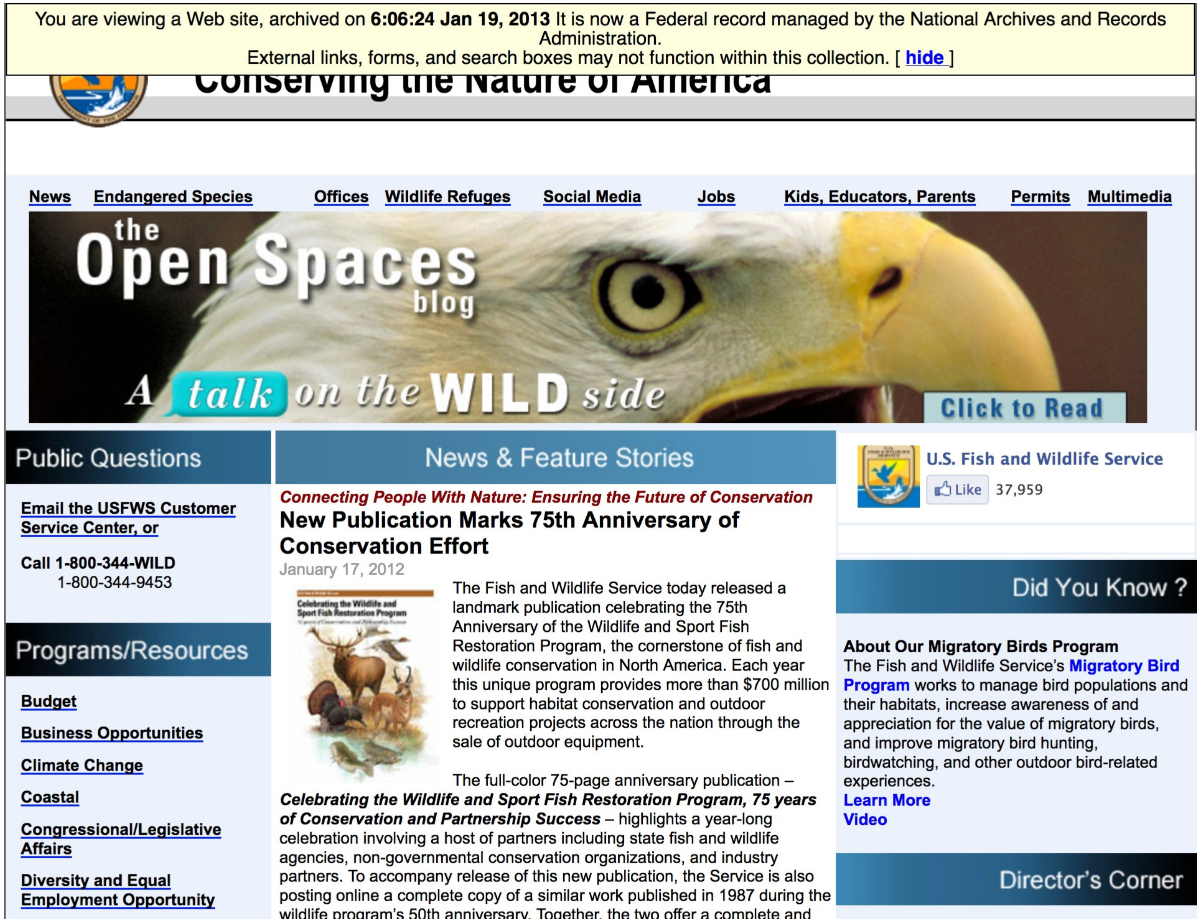
PROXY (INDIRECT) MEASUREMENTS

Data source: Reconstruction from ice cores.
Credit: NOAA

Clearly we need to render the entire page, then compute the hash.

Unfortunately, that's not easy.

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

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 Employment Opportunity

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 publication, the Service is also posting online a complete copy of a similar work published in 1987 during the wildlife program's 50th anniversary. Together, the two offer a complete and

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

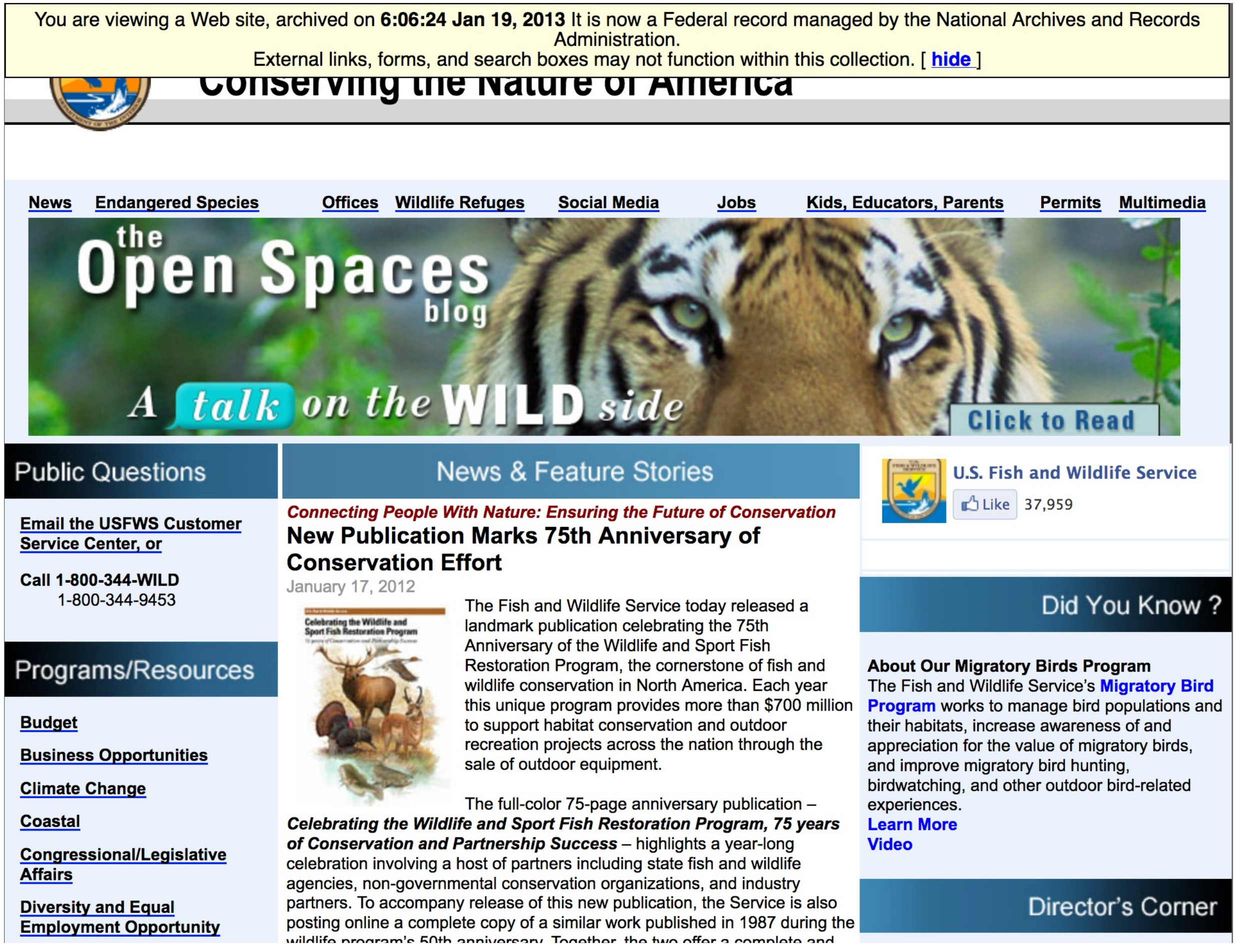
Did You Know ?

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

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 Employment Opportunity

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 publication, the Service is also posting online a complete copy of a similar work published in 1987 during the wildlife program's 50th anniversary. Together, the two offer a complete and

U.S. Fish and Wildlife Service

Like 37,959

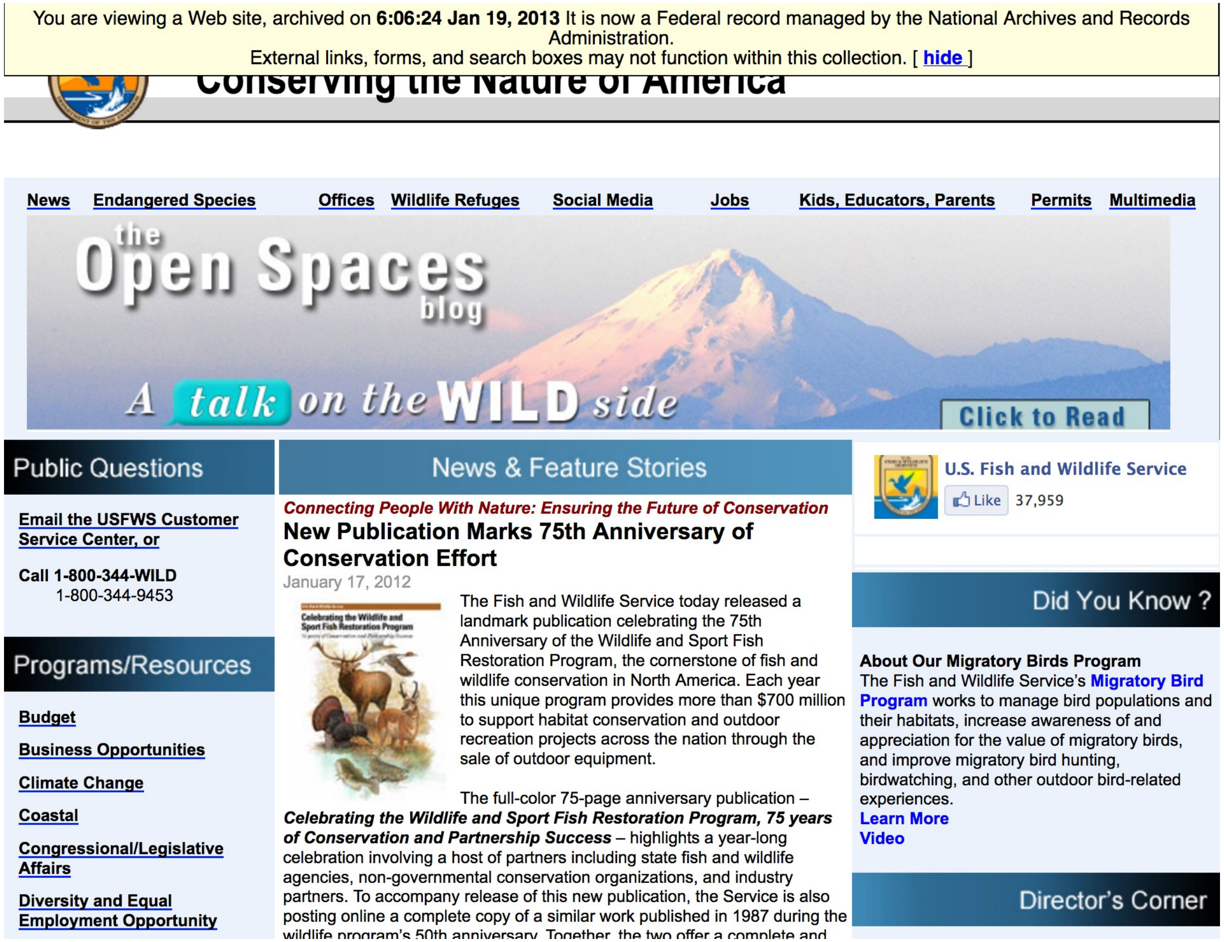
Did You Know ?

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

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 Employment Opportunity

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 publication, the Service is also posting online a complete copy of a similar work published in 1987 during the wildlife program's 50th anniversary. Together, the two offer a complete and

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

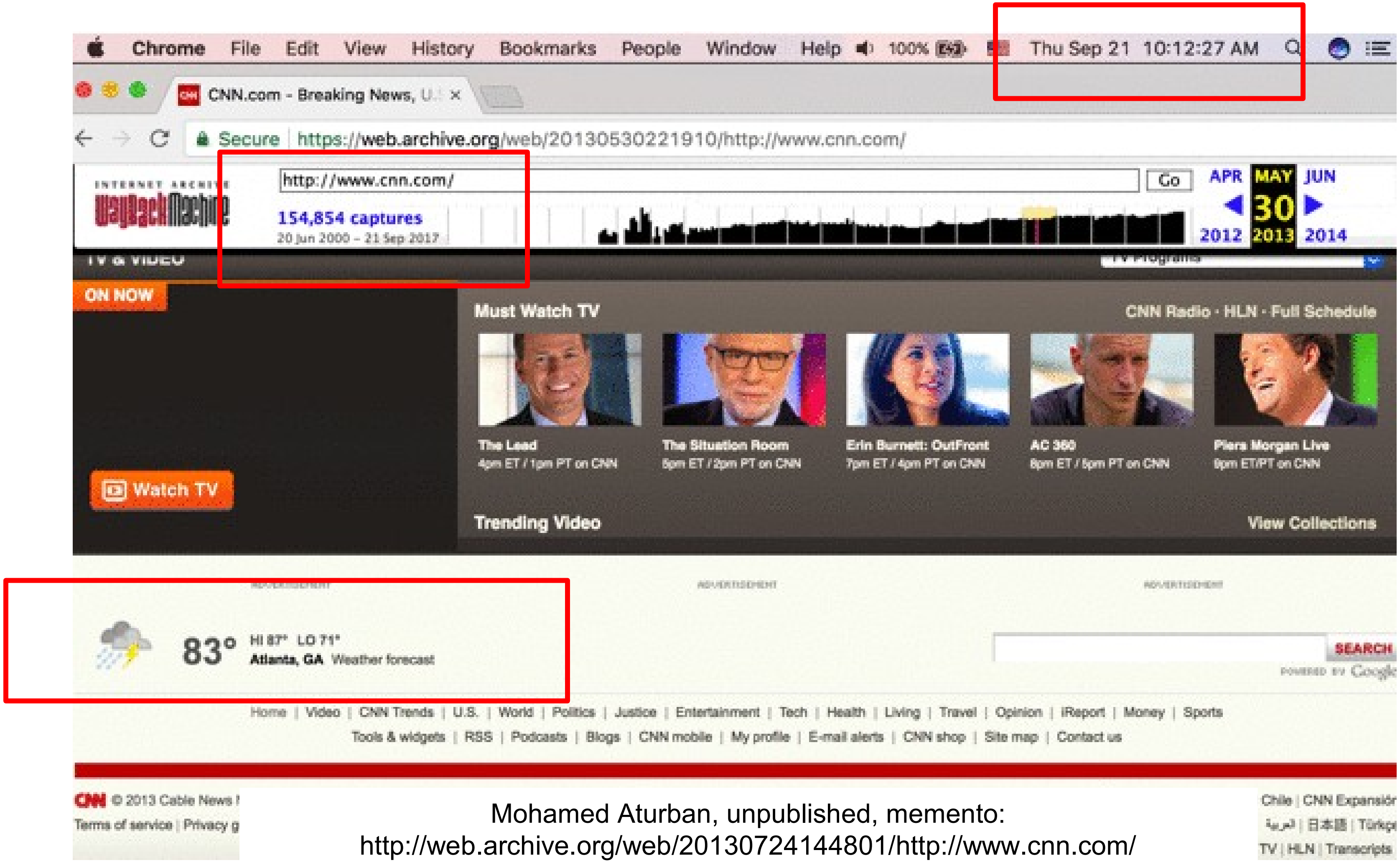
Director's Corner

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

“Look on my Javascript, ye Mighty, and despair!”

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

Actually, the fws.gov example was super easy; most changes are much harder to trace.



Chrome File Edit View History Bookmarks People Window Help 100%
Thu Sep 21 10:12:27 AM
CNN.com - Breaking News, U.
Secure
INTERNET ARCHIVE
WaybackMachine
Go
APR MAY JUN
30
2012 2013 2014
TV & VIDEO
TV Programs
ON NOW
Must Watch TV
CNN Radio · HLN · Full Schedule
Watch TV
Trending Video
View Collections
ADVERTISEMENT
ADVERTISEMENT
ADVERTISEMENT
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
Chile | CNN Expansión
عربي | 日本語 | Türkçe
TV | HLN | Transcripts

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

http://www.cnn.com/

154,854 captures

20 Jun 2000 – 21 Sep 2017

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

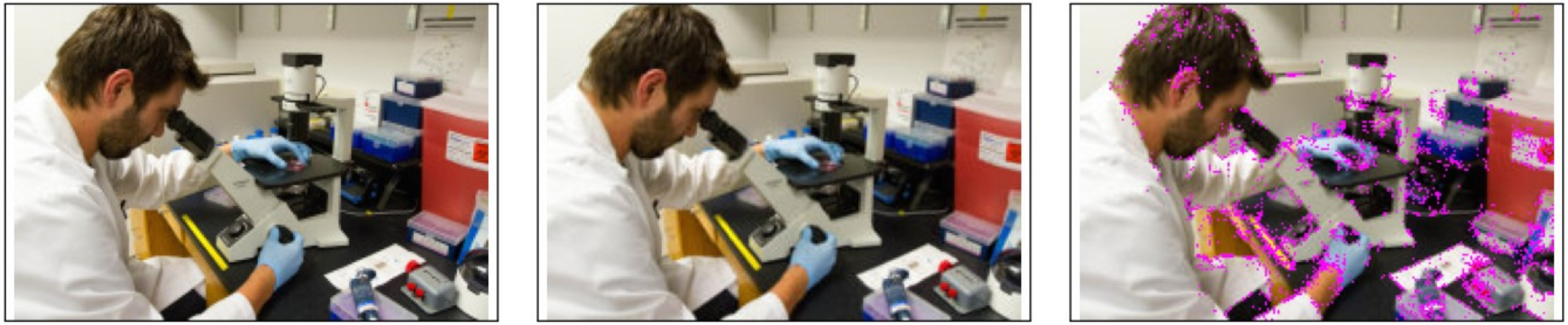
83° HI 87° LO 71°
Atlanta, GA Weather forecast

CNN © 2013 Cable News
Terms of service | Privacy g

Mohamed Aturban, unpublished, memento:
http://web.archive.org/web/20130724144801/http://www.cnn.com/
Animated GIF: https://blog.dshr.org/2017/11/keynote-at-pacific-neighborhood.html

Temporal violations: reconstructing pages that never existed on the live web

(examples below are transient; sometimes you get the 1st image, sometimes the 2nd image)



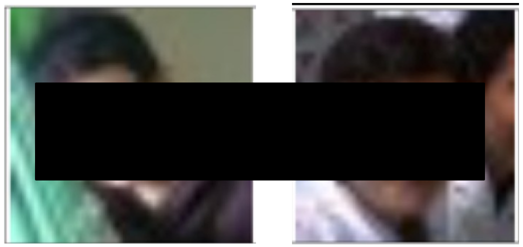
(a) Downloaded on November 16, 2017. Its hash ends in “4465eb88c7”.

(b) Downloaded on December 25, 2017. Its hash ends in “021e7b224b”.

(c) Comparing images (a) and (b) using [9] (mismatched pixels in pink).

embedded in umich.edu memento, archived in perma.cc

2nd image is compressed (12209 vs. 19448 bytes); 2nd image modified in 2017-03, but replayed in a 2017-01 page



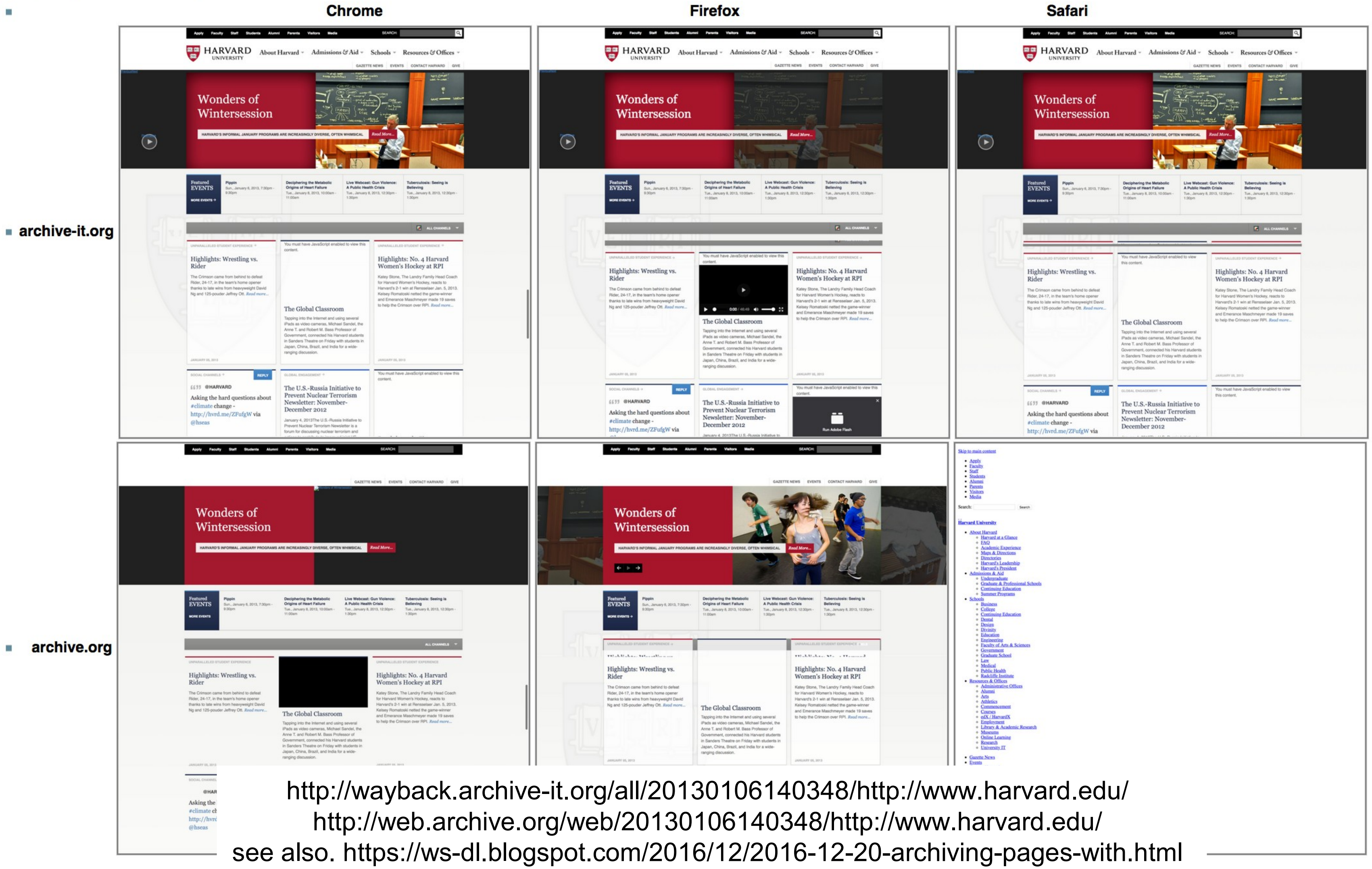
[redacted]

embedded in copybogger.com memento, archived in archive.org

2nd image modified in 2017-12, but replayed in a 2017-11 page; blackout for privacy

Temporal violations: https://ws-dl.blogspot.com/2015/12/2015-12-08-evaluating-temporal.html

1 WARC file, 2 Wayback Machines, 3 Browsers = 6 different replays



http://wayback.archive-it.org/all/20130106140348/http://www.harvard.edu/
http://web.archive.org/web/20130106140348/http://www.harvard.edu/
see also. https://ws-dl.blogspot.com/2016/12/2016-12-20-archiving-pages-with.html

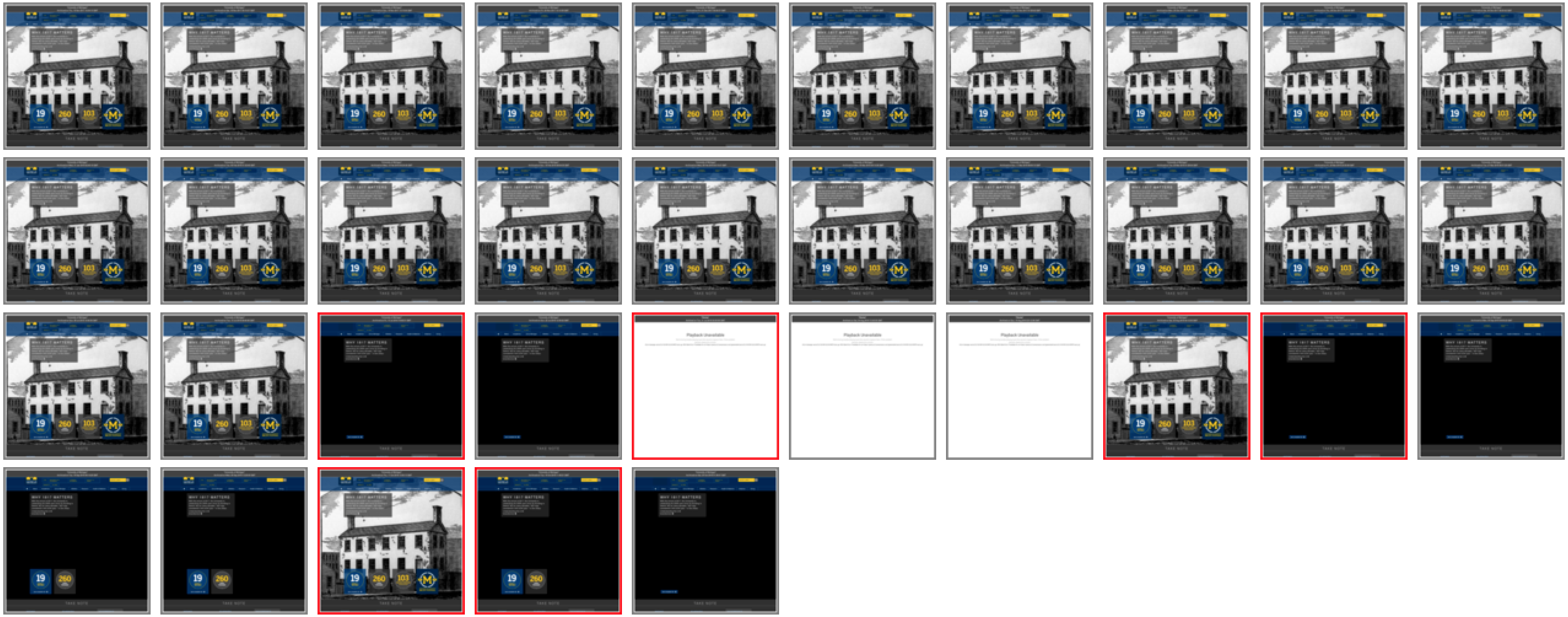
Experiment Design

Sample 16k+ Mementos from 17 Web Archives

Archive	URI-Ms
perma-archives.org	182
bibalex.org	199
webarchive.org.uk	349
bac-lac.gc.ca	351
proni.gov.uk	469
digar.ee	488
webharvest.gov	712
internetmemory.org	979
nationalarchives.gov.uk	994
stanford.edu	1222
archive-it.org	1383
archive.is	1396
web.archive.org	1566
arquivo.pt	1569
webcitation.org	1585
vefsafn.is	1589
loc.gov	1594
Total	16627

Periodically Replay Each *Archived* Page

35 times, from Nov. 2017 – Oct. 2018

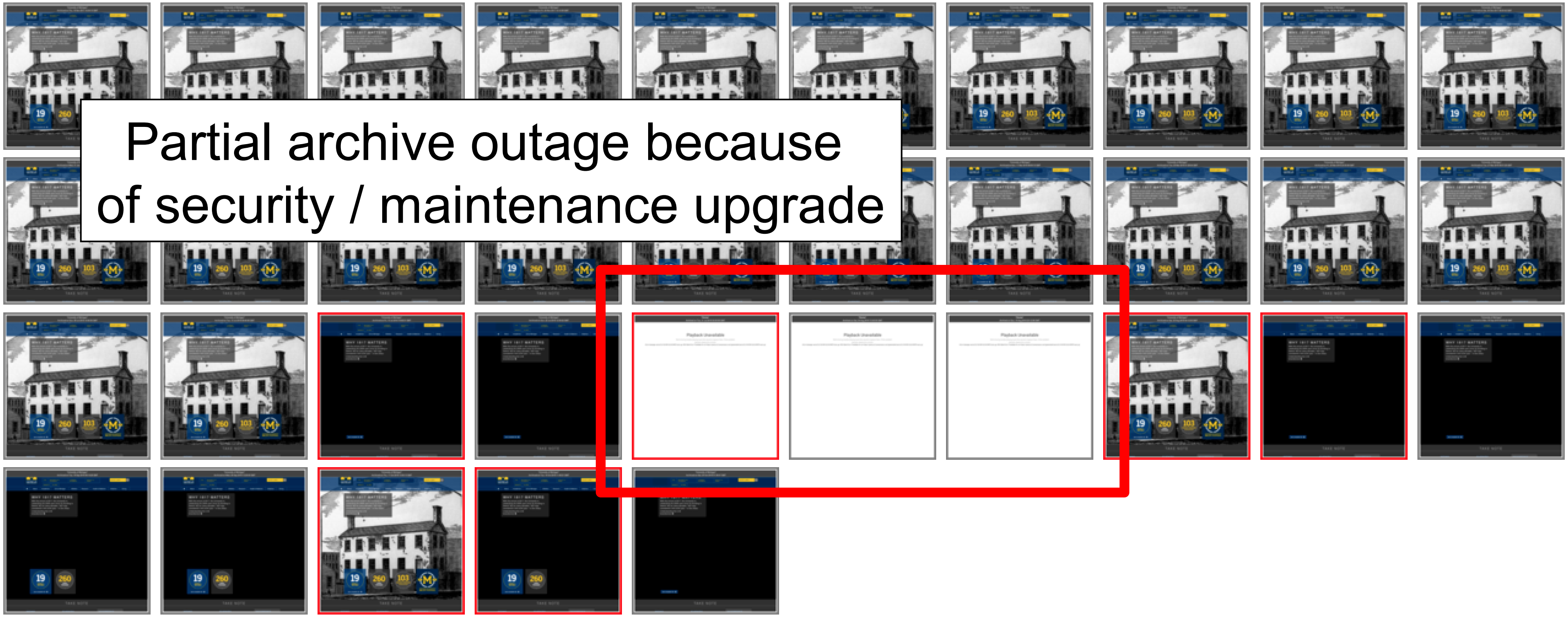


For each replay, we download both the rewritten version and the “raw” version (where possible).

Above example: http://perma-archives.org/warc/20170101182813/http://umich.edu/

Periodically Replay Each *Archived* Page

35 times, from Nov. 2017 – Oct. 2018



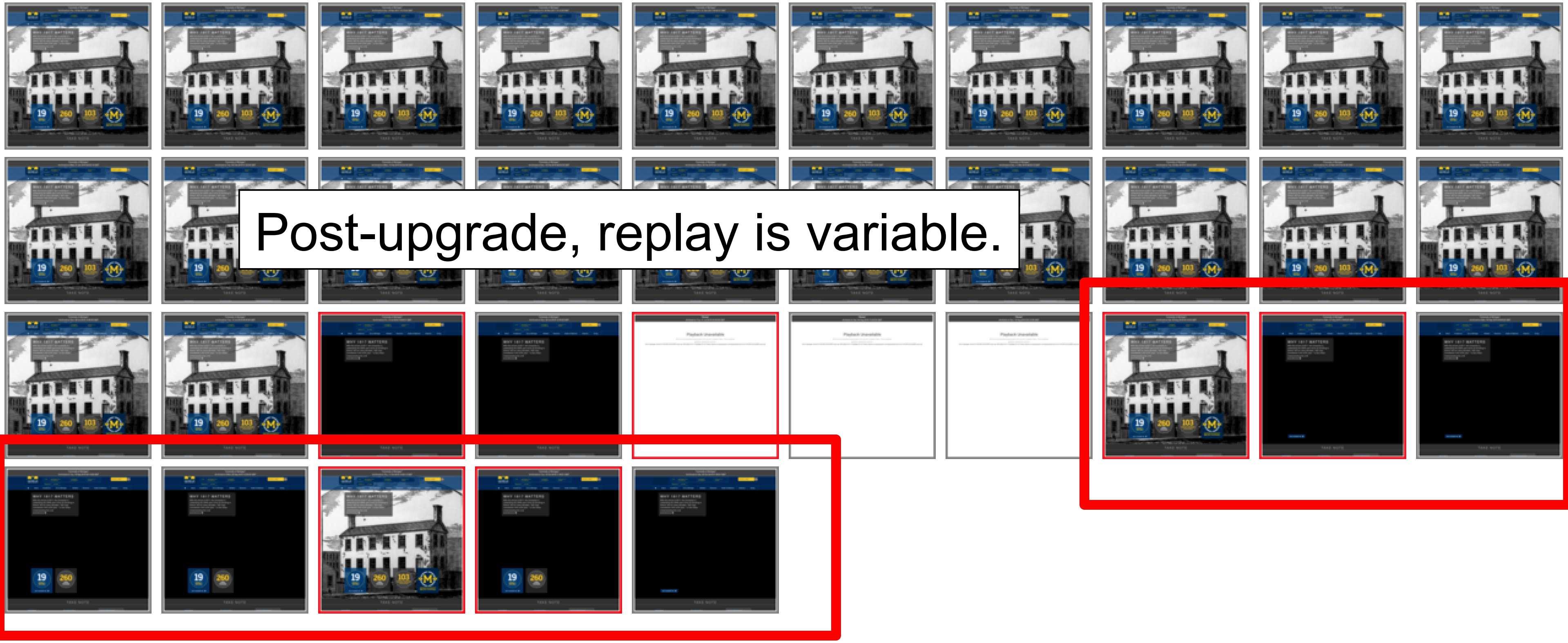
Partial archive outage because of security / maintenance upgrade

For each replay, we download both the rewritten version and the “raw” version (where possible).

Above example: http://perma-archives.org/warc/20170101182813/http://umich.edu/

Periodically Replay Each *Archived* Page

35 times, from Nov. 2017 – Oct. 2018



Post-upgrade, replay is variable.

For each replay, we download both the rewritten version and the “raw” version (where possible).

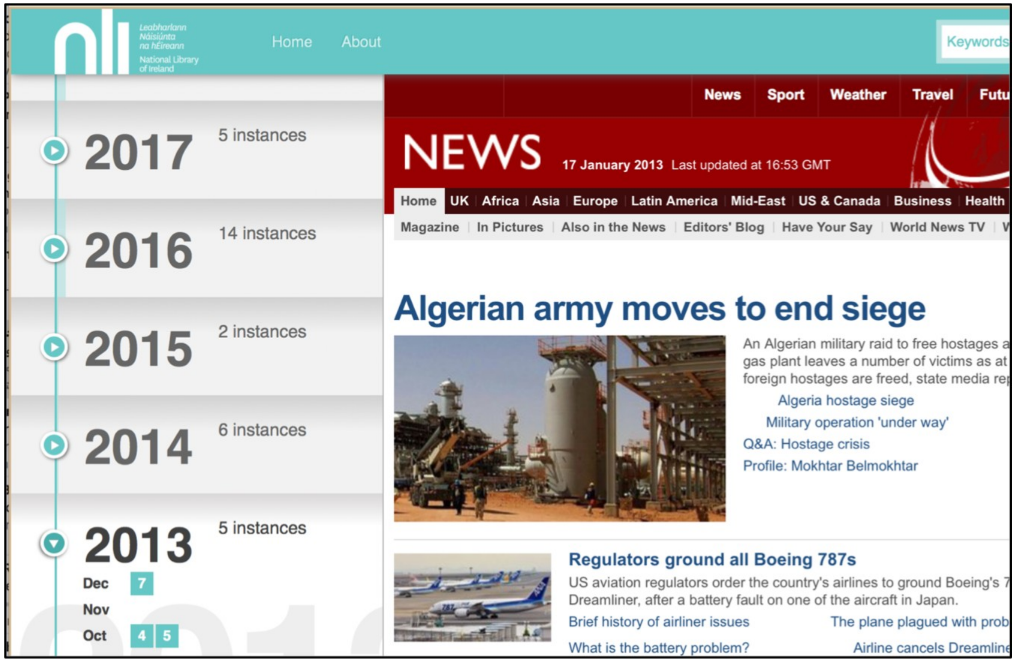
Above example: http://perma-archives.org/warc/20170101182813/http://umich.edu/

OLD DOMINION
UNIVERSITY

In 11 months,
11% of the URLs Disappeared or Changed

820 were renamed & required manual rediscovery
979 disappeared & have not yet been rediscovered

europarchive.org became internetmemory.org



17 January 2013 Last updated at 16:53 GMT

Algerian army moves to end siege

An Algerian military raid to free hostages a gas plant leaves a number of victims as at foreign hostages are freed, state media re

Algeria hostage siege

Military operation 'under way'

Q&A: Hostage crisis

Profile: Mokhtar Belmokhtar

Regulators ground all Boeing 787s

US aviation regulators order the country's airlines to ground Boeing's 7 Dreamliner, after a battery fault on one of the aircraft in Japan.

Brief history of airliner issues

The plane plagued with prob

What is the battery problem?

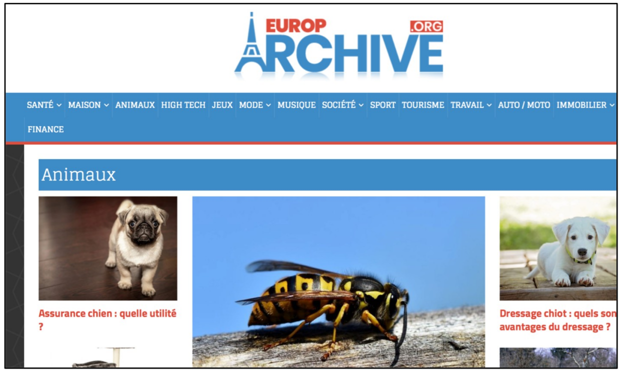
Airline cancels Dreamline

URI-Ms like this:
collection.europarchive.org/nli/20130117165443/http://bbc.co.uk/news/

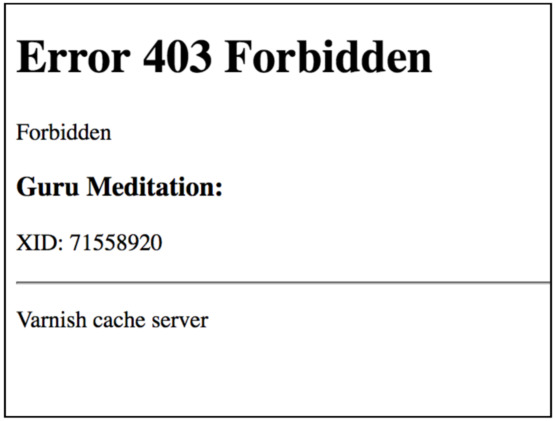
changed domains and became like this:
collections.internetmemory.org/nli/20130117165443/http://bbc.co.uk/news/

europarchive.org

europarchive.org is now spam
internetmemory.org is now down
979 pages lost


EUROP ARCHIVE .ORG
SANTÉ MAISON ANIMAUX HIGH TECH JEUX MODE MUSIQUE SOCIÉTÉ SPORT TOURISME TRAVAIL AUTO / MOTO IMMOBILIER
FINANCE
Animaux
Assurance chien : quelle utilité ?
Dressage chiot : quels son avantages du dressage ?

```
curl -I collection.europarchive.org/nli/20130117165443/http:/bbc.co.uk/news/
HTTP/1.1 301 Moved Permanently
Date: Mon, 10 Dec 2018 04:30:50 GMT
Server: Apache
Expires: Mon, 10 Dec 2018 05:30:50 GMT
Cache-Control: max-age=3600
Location: http://europarchive.org
Connection: close
Content-Type: text/html; charset=UTF-8
```



Error 403 Forbidden

Forbidden

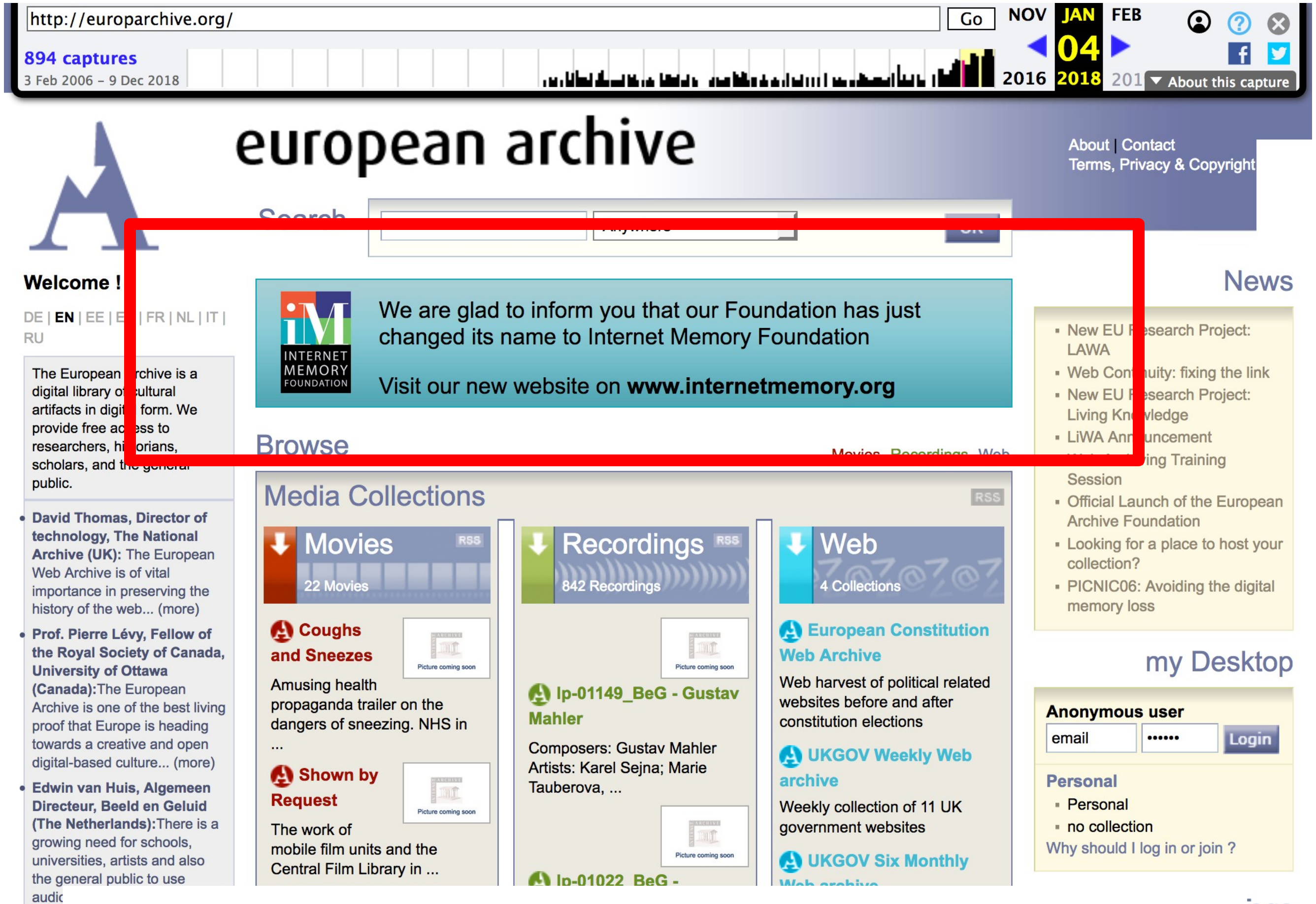
Guru Meditation:

XID: 71558920

Varnish cache server

```
curl -I collections.internetmemory.org/nli/20130117165443/http://bbc.co.uk/news/
HTTP/1.1 403 Forbidden
Date: Mon, 10 Dec 2018 04:31:51 GMT
Server: Varnish
X-Varnish: 71167297
Content-Type: text/html; charset=utf-8
Retry-After: 5
Content-Length: 252
Connection: keep-alive
```

Telling humans your domain is about to change is nice, but please tell robots too…



http://europarchive.org/ Go NOV JAN FEB
894 captures
3 Feb 2006 – 9 Dec 2018
04
2016 2018 201
About this capture

european archive

About | Contact
Terms, Privacy & Copyright

Search

Welcome !

DE | EN | EE | E | FR | NL | IT | RU

The European Archive is a digital library of cultural artifacts in digital form. We provide free access to researchers, historians, scholars, and the general public.

- David Thomas, Director of technology, The National Archive (UK): The European Web Archive is of vital importance in preserving the history of the web... (more)
- Prof. Pierre Lévy, Fellow of the Royal Society of Canada, University of Ottawa (Canada):The European Archive is one of the best living proof that Europe is heading towards a creative and open digital-based culture... (more)
- Edwin van Huis, Algemeen Directeur, Beeld en Geluid (The Netherlands):There is a growing need for schools, universities, artists and also the general public to use audi
- Prof.

INTERNET MEMORY FOUNDATION

We are glad to inform you that our Foundation has just changed its name to Internet Memory Foundation

Visit our new website on **www.internetmemory.org**

Browse

Movies Recordings Web

Media Collections

RSS

Movies RSS
22 Movies

Recordings RSS
842 Recordings

Web
4 Collections

Coughs and Sneezes
Amusing health propaganda trailer on the dangers of sneezing. NHS in ...

Shown by Request
The work of mobile film units and the Central Film Library in ...

Picture coming soon

Ip-01149_BeG - Gustav Mahler
Composers: Gustav Mahler Artists: Karel Sejna; Marie Tauberova, ...

Ip-01022_BeG -

European Constitution Web Archive
Web harvest of political related websites before and after constitution elections

UKGOV Weekly Web archive
Weekly collection of 11 UK government websites

UKGOV Six Monthly Web archive

News

- New EU Research Project: LAWA
- Web Continuity: fixing the link
- New EU Research Project: Living Knowledge
- LiWA Announcement
- ing Training Session
- Official Launch of the European Archive Foundation
- Looking for a place to host your collection?
- PICNIC06: Avoiding the digital memory loss

my Desktop

Anonymous user
email •••••• Login

Personal
- Personal
- no collection

Why should I log in or join ?

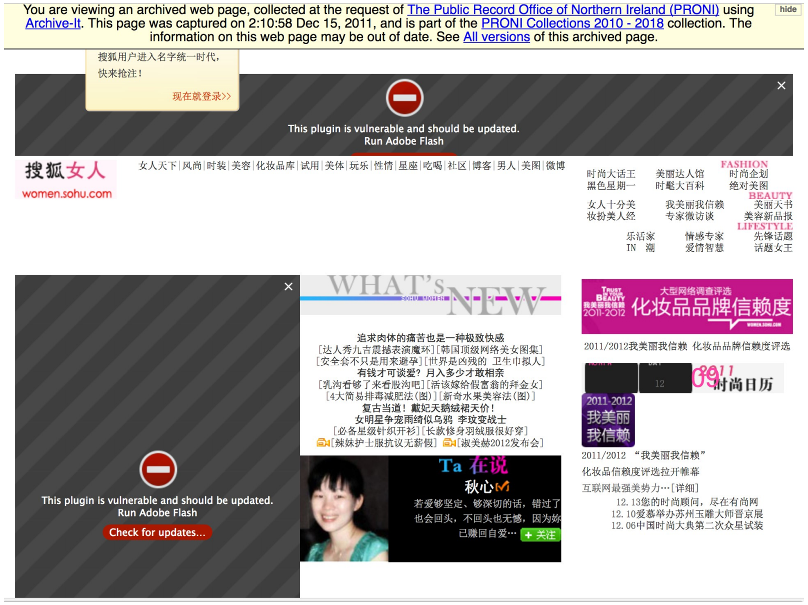
Tags

https://web.archive.org/web/20180104021440/http://europarchive.org/
See: https://tools.ietf.org/id/draft-wilde-sunset-header-03.html

OLD DOMINION UNIVERSITY

CNI Fall 2018 Membership Meeting, 2018-12-11, @phonedude_mln, @WebSciDL

webarchive.proni.gov.uk now uses Archive-It



You are viewing an archived web page, collected at the request of The Public Record Office of Northern Ireland (PRONI) using Archive-It. This page was captured on 2:10:58 Dec 15, 2011, and is part of the PRONI Collections 2010 - 2018 collection. The information on this web page may be out of date. See All versions of this archived page.

搜狐用户进入名字统一时代，
快来抢注！
现在就登录>>

This plugin is vulnerable and should be updated.
Run Adobe Flash

搜狐女人
women.sohu.com

女人天下 | 风尚 | 时装 | 美容 | 化妆品库 | 试用 | 美体 | 玩乐 | 性情 | 星座 | 吃喝 | 社区 | 博客 | 男人 | 美图 | 微博

FASHION
时尚大话王 美丽达人馆 时尚企划
黑色星期一 时髦大百科 绝对美图

BEAUTY
女人十分美 我美丽我信赖 美丽天书
妆扮美人经 专家微访谈 美容新品报

LIFESTYLE
乐活家 情感专家 先锋话题
IN 潮 爱情智慧 话题女王

WHAT'S NEW

追求肉体的痛苦也是一种极致快感
[达人秀九吉震撼表演魔环][韩国顶级网络美女图集]
[安全套不只是用来避孕][世界是凶残的 卫生巾拟人]
有钱才可谈爱？月入多少才敢相亲
[乳沟看够了来看股沟吧][活该嫁给假富翁的拜金女]
[4大简易排毒减肥法(图)][新奇水果美容法(图)]
复古当道！戴妃天鹅绒裙天价！
女明星争宠雨绮似乌鸦 李玟变战士
[必备星级针织开衫][长款修身羽绒服很好穿]
[辣妹护士服抗议无薪假] [淑美赫2012发布会]

Ta 在说
秋心
若爱够坚定、够深切的话，错过了也会回头，不回头也无憾，因为妳已赚回自爱…
+ 关注

This plugin is vulnerable and should be updated.
Run Adobe Flash
Check for updates...

大型网络调查评选
化妆品品牌信赖度
2011/2012我美丽我信赖 化妆品品牌信赖度评选

12
2011 09 时尚日历

2011-2012 我美丽 我信赖

2011/2012 "我美丽我信赖"
化妆品信赖度评选拉开帷幕
互联网最强美势力…[详细]
12.13您的时尚顾问，尽在有尚网
12.10爱慕举办苏州玉雕大师晋京展
12.06中国时尚大典第二次众星试装

The top-level site webarchive.proni.gov.uk still exists, but deep links to URI-Ms are now 404.

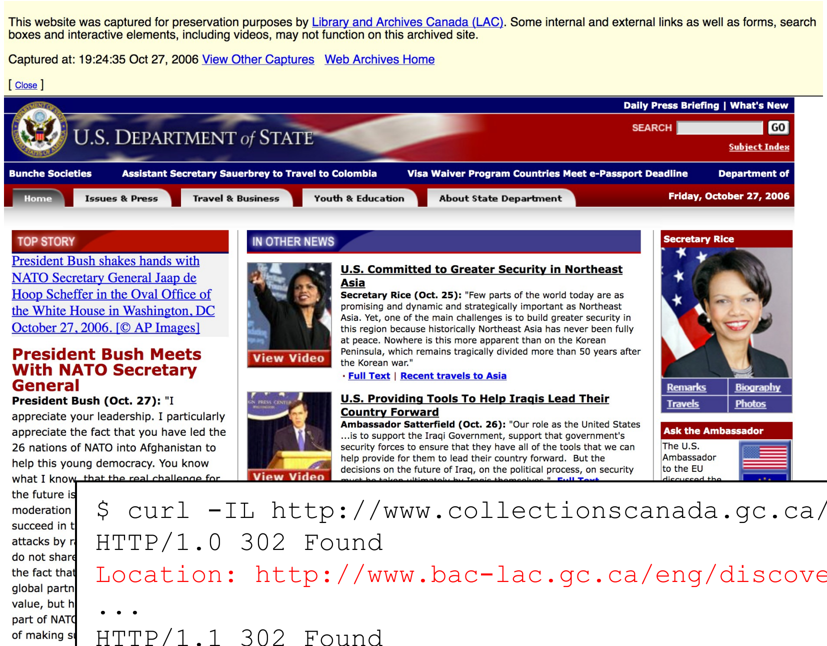
469 pages required manual rediscovery.

```
curl -I http://webarchive.proni.gov.uk/20111215021058/http://women.sohu.com/
HTTP/1.1 404 Not Found
Date: Mon, 10 Dec 2018 05:15:56 GMT
Server: Apache/2.4.18 (Ubuntu)
Content-Type: text/html; charset=iso-8859-1

curl -I https://wayback.archive-it.org/11112/20111215021058/http://women.sohu.com/
HTTP/1.1 200 OK
...
```

www.collectionscanada.gc.ca became webarchive.bac-lac.gc.ca:8080

(no, really – port 8080)

And deep links to URI-Ms now redirect to the top of the new site, which means 351 pages required manual rediscovery:



This website was captured for preservation purposes by Library and Archives Canada (LAC). Some internal and external links as well as forms, search boxes and interactive elements, including videos, may not function on this archived site.

Captured at: 19:24:35 Oct 27, 2006 View Other Captures Web Archives Home

[Close]

President Bush shakes hands with NATO Secretary General Jaap de Hoop Scheffer in the Oval Office of the White House in Washington, DC October 27, 2006. [© AP Images]

President Bush Meets With NATO Secretary General

President Bush (Oct. 27): "I appreciate your leadership. I particularly appreciate the fact that you have led the 26 nations of NATO into Afghanistan to help this young democracy. You know what I know, that the real challenge for
the future is
moderation
succeed in t
attacks by r
do not share
the fact that
global partn
value, but h
part of NATC
of making su

U.S. Committed to Greater Security in Northeast Asia

Secretary Rice (Oct. 25): "Few parts of the world today are as promising and dynamic and strategically important as Northeast Asia. Yet, one of the main challenges is to build greater security in this region because historically Northeast Asia has never been fully at peace. Nowhere is this more apparent than on the Korean Peninsula, which remains tragically divided more than 50 years after the Korean war."

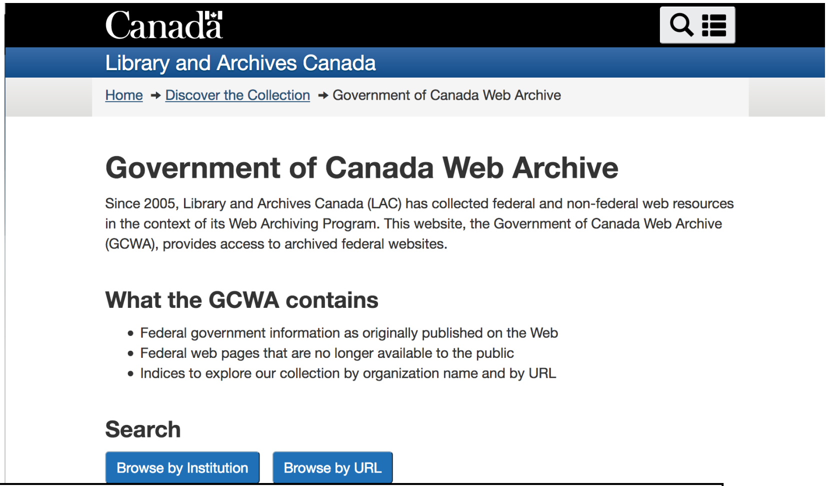
· Full Text | Recent travels to Asia

U.S. Providing Tools To Help Iraqis Lead Their Country Forward

Ambassador Satterfield (Oct. 26): "Our role as the United States ...is to support the Iraqi Government, support that government's security forces to ensure that they have all of the tools that we can help provide for them to lead their country forward. But the decisions on the future of Iraq, on the political process, on security must be taken ultimately by Iraqis themselves." Full Text

Ask the Ambassador

The U.S. Ambassador to the EU discussed the



Canada
Library and Archives Canada
Home → Discover the Collection → Government of Canada Web Archive

Government of Canada Web Archive

Since 2005, Library and Archives Canada (LAC) has collected federal and non-federal web resources in the context of its Web Archiving Program. This website, the Government of Canada Web Archive (GCWA), provides access to archived federal websites.

What the GCWA contains

- Federal government information as originally published on the Web
- Federal web pages that are no longer available to the public
- Indices to explore our collection by organization name and by URL

Search

```
$ curl -IL http://www.collectionscanada.gc.ca/webarchives/20061027192435/http://www.state.gov/
HTTP/1.0 302 Found
Location: http://www.bac-lac.gc.ca/eng/discover/archives-web-government/Pages/web-archives.aspx
...
HTTP/1.1 302 Found
Location: http://webarchive.bac-lac.gc.ca/?lang=en
Date: Mon, 10 Dec 2018 04:46:24 GMT
...
HTTP/1.1 200
Date: Mon, 10 Dec 2018 04:46:24 GMT
...
```

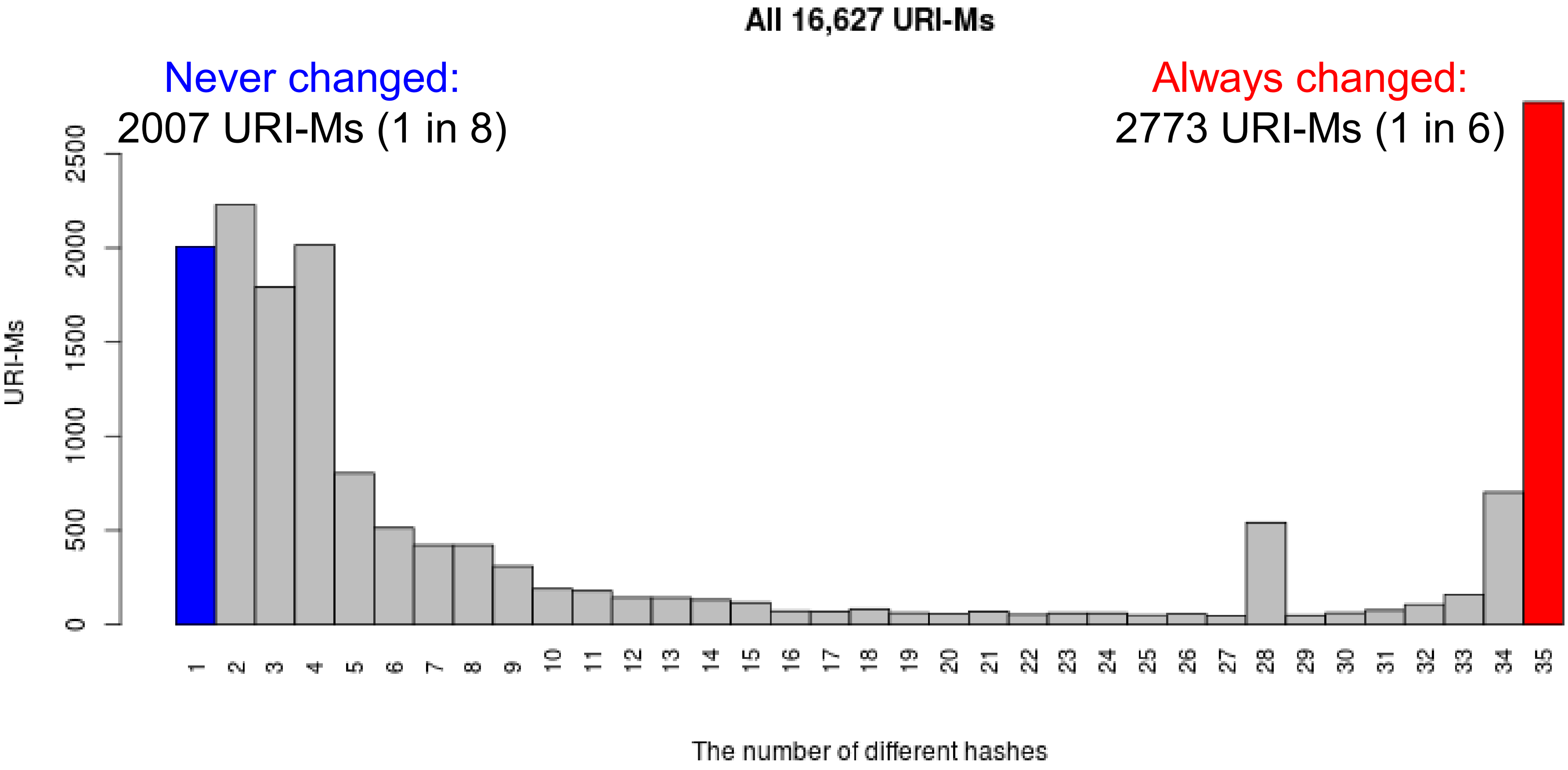
7 out of 8 pages produced > 1 hash over 11 months

```
                                    URI-Ms with at
Archive Name                URI-Ms least two hashes
------------------------------------------------------
webarchive.loc.gov           1,594    1,235 (77.47%)
vefsafn.is                   1,589    1,133 (71.30%)
webcitation.org              1,585      981 (61.89%)
arquivo.pt                   1,569    1,563 (99.61%)
archive.org                  1,566    1,430 (91.31%)
archive.is                   1,396    1,364 (97.70%)
archive-it.org               1,383    1,383   (100%)
swap.stanford.edu            1,222    1,005 (82.24%)
nationalarchives.gov.uk        994      978 (98.39%)
internetmemory.org             979      979   (100%)
webharvest.gov                 712      712   (100%)
digar.ee                       488      308 (63.11%)
proni.gov.uk                   469      469   (100%)
bac-lac.gc.ca                  351      351 (100.0%)
webarchive.org.uk              349      348 (99.71%)
archive.bibalex.org            199      199   (100%)
perma-archives.org             182      182   (100%)
------------------------------------------------------
total                       16,627   14,620 (87.92%)
```

You cannot replay twice the same archived page

(apologies to Heraclitus)

More Archived Pages Changed Every Time Than Never Changed

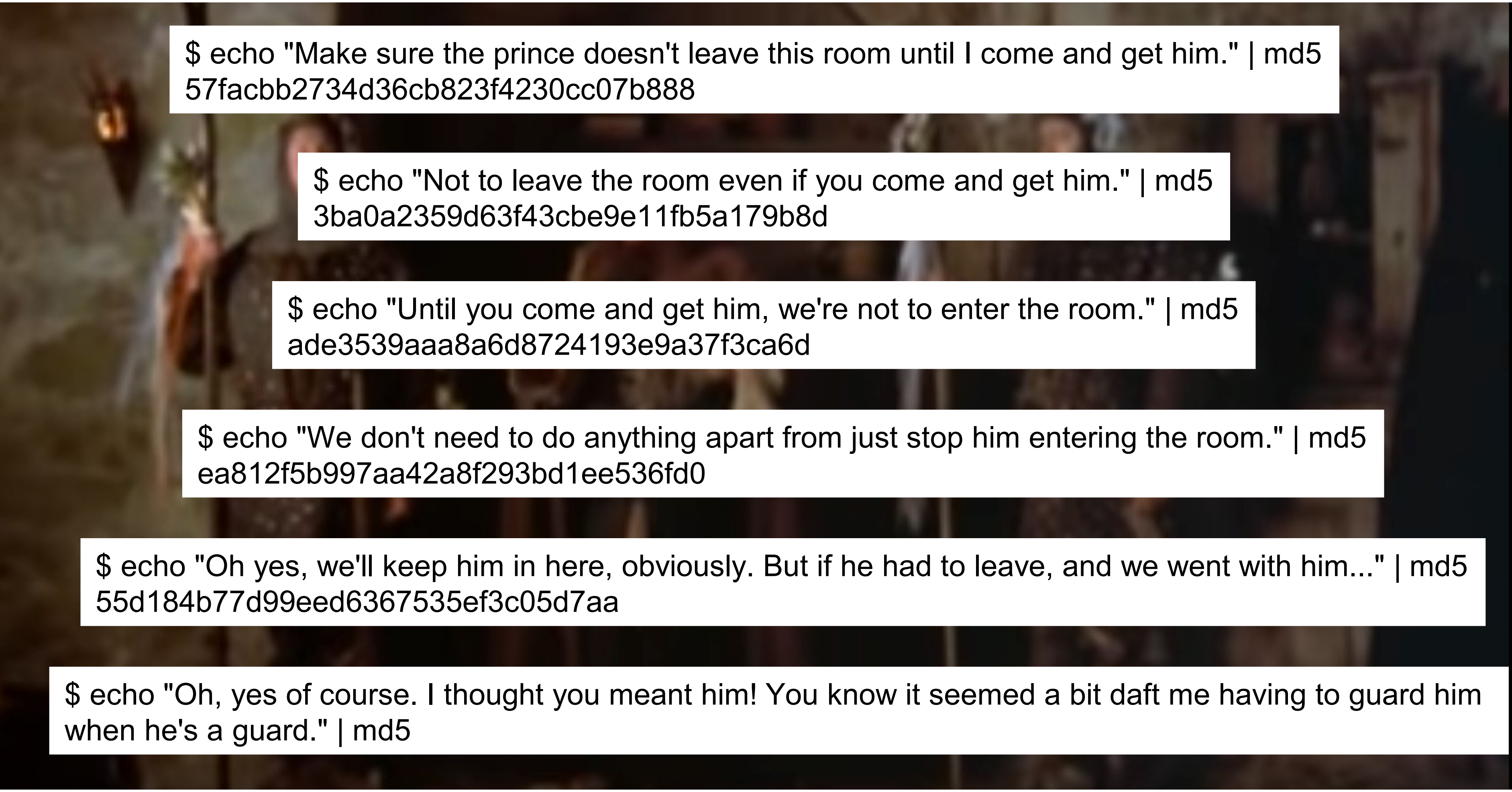


A metaphor for replaying archived web pages

https://www.youtube.com/watch?v=ekO3Z3XWa0Q

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

King of Swamp Castle: live web/ground truth
Guard: archival replay



```
$ echo "Make sure the prince doesn't leave this room until I come and get him." | md5
57facbb2734d36cb823f4230cc07b888
```

```
$ echo "Not to leave the room even if you come and get him." | md5
3ba0a2359d63f43cbe9e11fb5a179b8d
```

```
$ echo "Until you come and get him, we're not to enter the room." | md5
ade3539aaa8a6d8724193e9a37f3ca6d
```

```
$ echo "We don't need to do anything apart from just stop him entering the room." | md5
ea812f5b997aa42a8f293bd1ee536fd0
```

```
$ echo "Oh yes, we'll keep him in here, obviously. But if he had to leave, and we went with him..." | md5
55d184b77d99eed6367535ef3c05d7aa
```

```
$ echo "Oh, yes of course. I thought you meant him! You know it seemed a bit daft me having to guard him when he's a guard." | md5
```

https://www.youtube.com/watch?v=ekO3Z3XWa0Q

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

CNI Fall 2018 Membership Meeting, 2018-12-11,
@phonedude_mln, @WebSciDL

OLD DOMINION UNIVERSITY

Archival replay & blockchain: building a castle in a swamp

- Fixity checks only work when it’s clear what to hash
  - Hash only the root HTML and modifications are possible via embedded resources (false negatives)
  - Recursively hash all embedded resources and you’ll rarely get the same hash (false positives)
- Replay is working as designed, it’s not something that will be “fixed”
  - we need server-side support for auditing, and archive-aware hashing functions
- There is increasing incentive to attack existing archives and create networks of fake archives
  - http://bit.ly/Weaponized-Web-Archives