

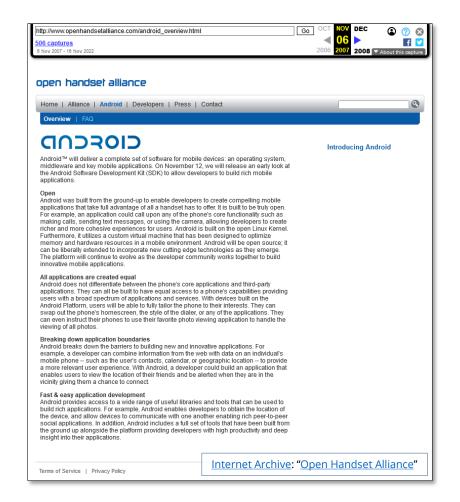
Evidentiary Use of the Temporal Web

Presented by: Nicholas Taylor and Joe Dugan

Specht v. Google, Inc. (2010)



<u>Specht v. Google, Inc., 758 F. Supp.</u> 2d 570 (N.D. III. Dec. 17, 2010)







two contemporary(?) websites





The New York Times

why do the temporal attributes of web content matter?

- cases may turn on when specific information was publicly available and/or discoverable on the web
 - when first published
 - when updated
 - when absent

- examples:
 - patent disclosures
 - use of marks or copyrighted materials
 - terms of service
 - advertised claims regarding products or services
 - business relationships

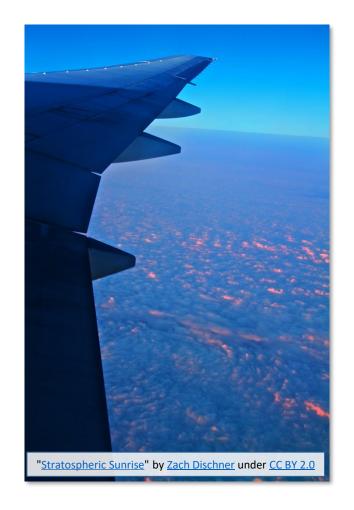
why not just rely on client's own records?

- unsystematic and/or incomplete recordkeeping
- savvy opposing counsel may check open sources for counterindicative evidence
- material info disseminated via third party platforms
- some questions (e.g., discoverability) can't be definitively settled w/ client records alone



overview

- Internet Archive Wayback Machine (IAWM)
 - background
 - interface + features
 - authentication
 - temporal coherence
- web timestamping
 - via IAWM
 - via Google
 - via Twitter



Nicholas' background

- expert witness w/ IAWM focus
- M.L.S., UMD / M.A., Comm., Culture, & Tech., Georgetown
- Google, Sun, SCOTUS, Library of Congress, Stanford, LANL
- 16+ years in research libraries + tech firms
- 9+ years in web archiving
- presentations, peer review, grants, media contributor
- website: <u>nullhandle.org</u>



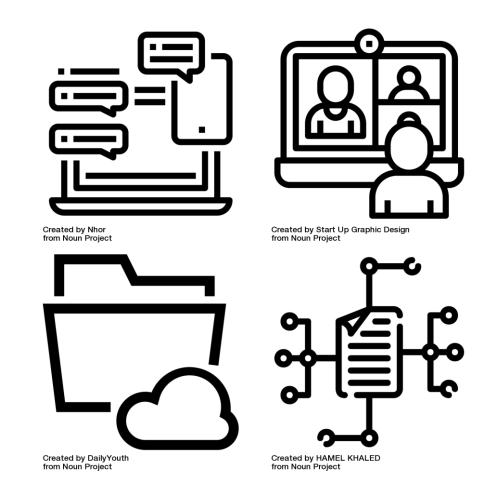
Joe's background

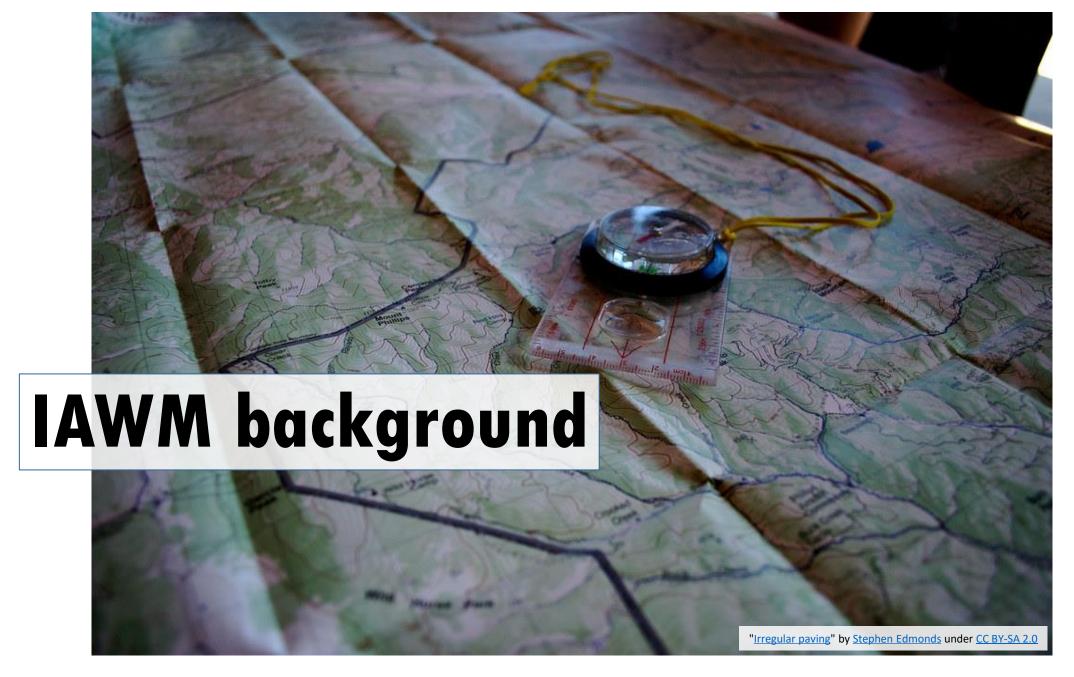
- J.D., 2015, Indiana University Maurer School of Law
- commercial litigation attorney at Gallagher Evelius & Jones in Baltimore
- previously worked as a trial attorney for DOJ, Civil Division
- clerked on U.S. District Court and U.S. Court of Appeals
- serves as his firm's unofficial go-to lawyer for ediscovery problem solving
- very recent experience navigating Slack ESI collection in the context of a discovery dispute



"<u>Document Retention, Litigation Holds, and Ediscovery Best Practices with Emerging Communication Technologies</u>"

- our 2nd session, Friday at 3:15p
- ediscovery ramifications of emerging communication tech – e.g., Zoom, Microsoft Teams, Google Drive, etc.
- topics: risk mitigation, client counseling strategies, benefits + drawbacks of DIY collection, primer on data exporting
- check out slides + CLE materials





Internet Archive

- non-profit digital library
- "universal access to all knowledge"
- provides free public access to collections of all types of digital materials

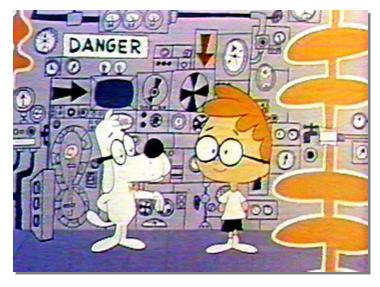


MMM'IECH2HOM'COW

IAWM

- free, open archive of websites 1996 – present
- web content is both culturally significant + ephemeral
- IAWM referenced in:
 - a 2002 patent application
 - a 2003 PTAB proceeding
 - a 2004 U.S. federal court case

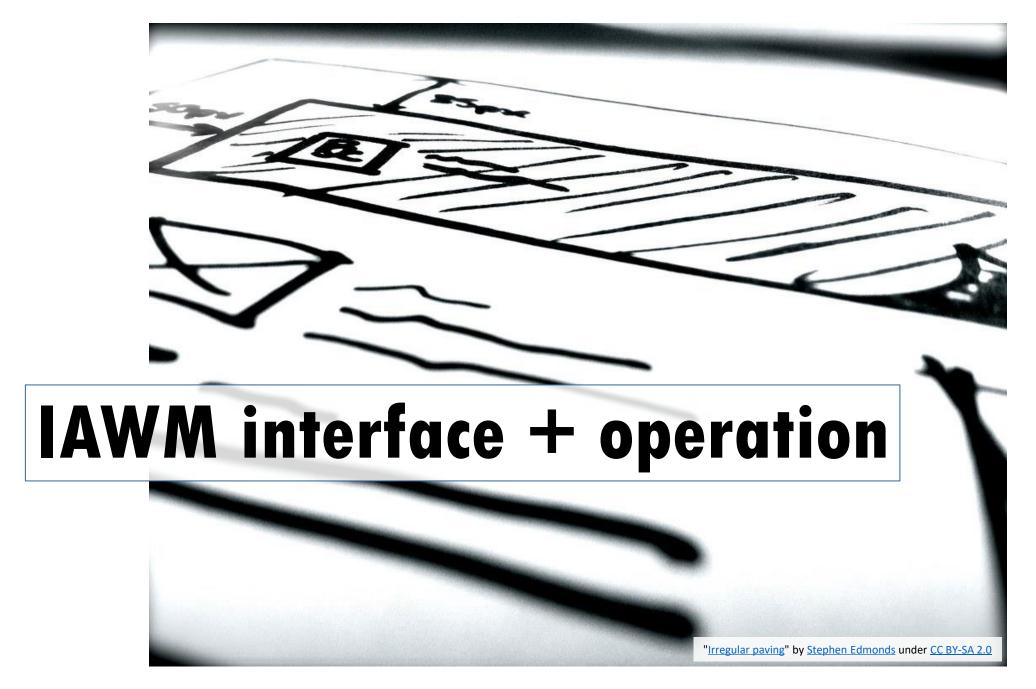




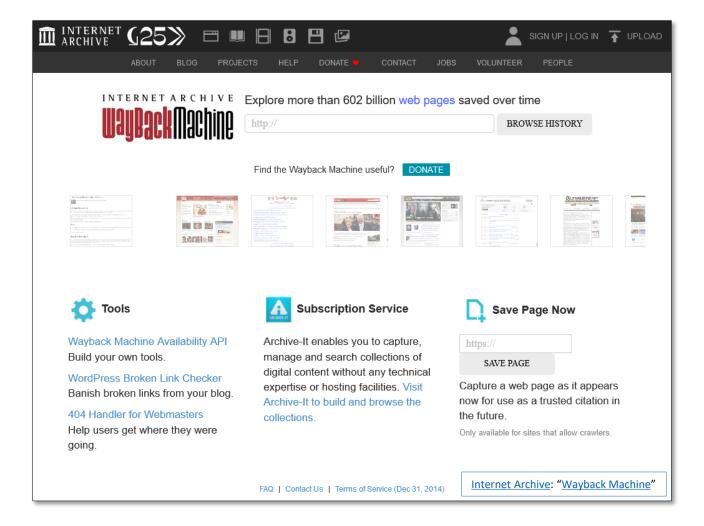
IAWM archival coverage

- "best effort" archive
- popular + more recent content is better archived
- content of interest to cultural heritage orgs is increasingly better archived
- social media less well archived
- non-public web content not archived

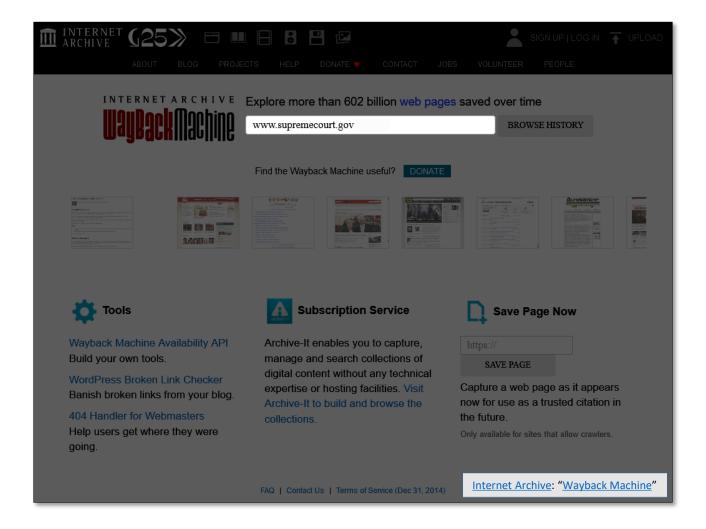




IAWM home

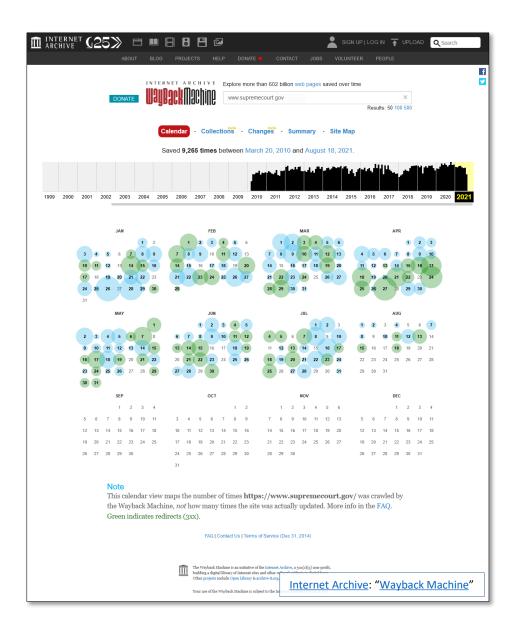


IAWM home: search box



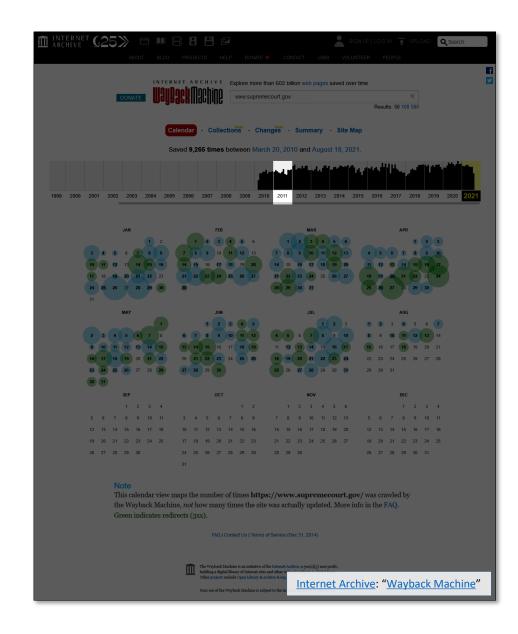
IAWM calendar page

- landing page presents archival coverage for specified web address for year selected in timeline
- bubble size indicates relative number of captures for that date
- bubble color indicates status code provided by the web server to the IA archiving agent
 - blue means <u>a successful</u> retrieval, with no errors
 - green means <u>a redirect was</u> encountered



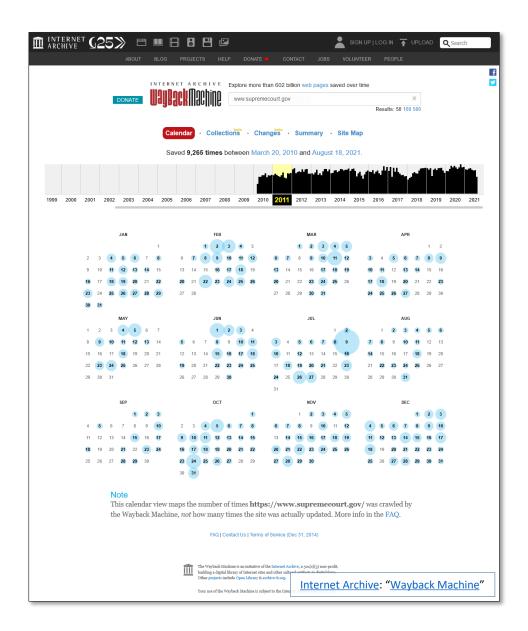
IAWM calendar page: captures by year

- click to another year in the timeline to change the calendar view below
- let's click to 2011



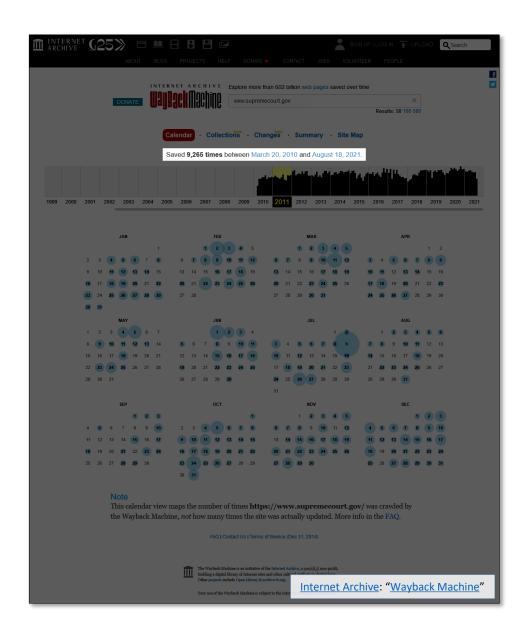
IAWM calendar page

- note that there are fewer captures in 2011, as indicated by:
 - shorter bars on the timeline histogram
 - generally smaller bubbles than in 2021
- general trend is that websites are better archived over time
- let's take a closer look at some of the other interface elements



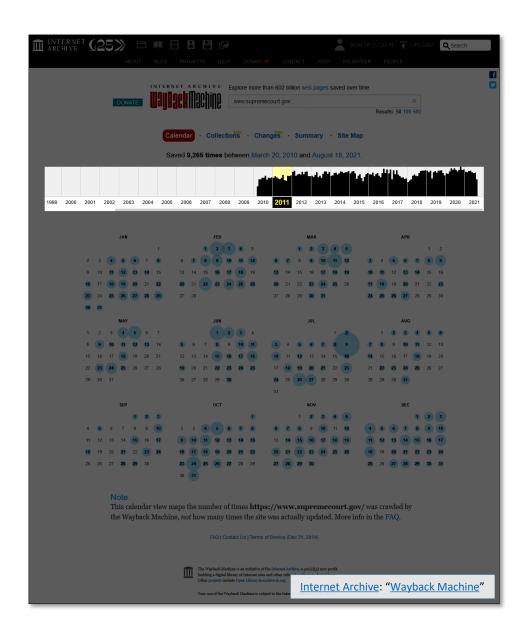
IAWM calendar page: captures range

- number of captures for the specified web address + their date range is displayed at the top
- links are provided to the first + last captures in the date range



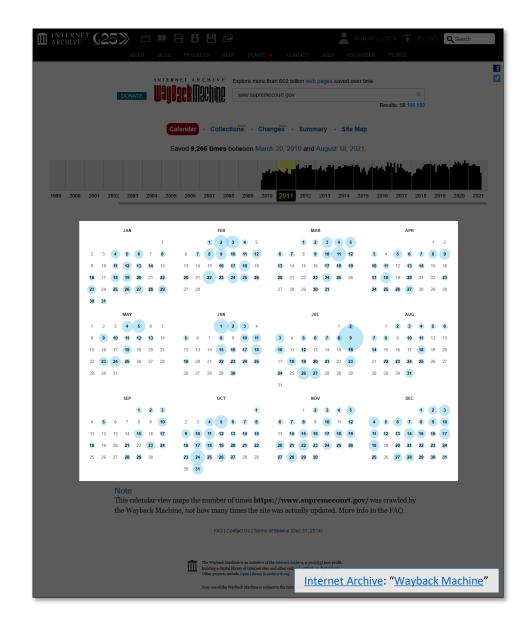
IAWM calendar page: timeline histogram

- the timeline histogram provides a visual summary of the relative number of captures over time
- each year segment has 12 columns – one for each month
- the year displayed in detail below is highlighted in yellow in the timeline



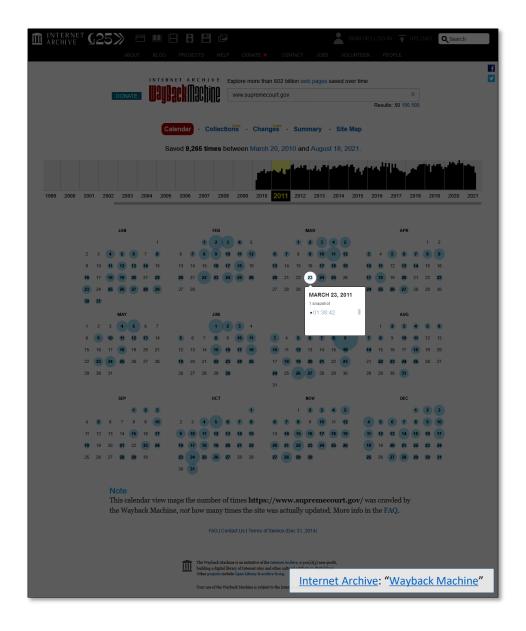
IAWM calendar page: calendar

- archival coverage for every day of the year selected in the timeline
- this is the 2011 calendar



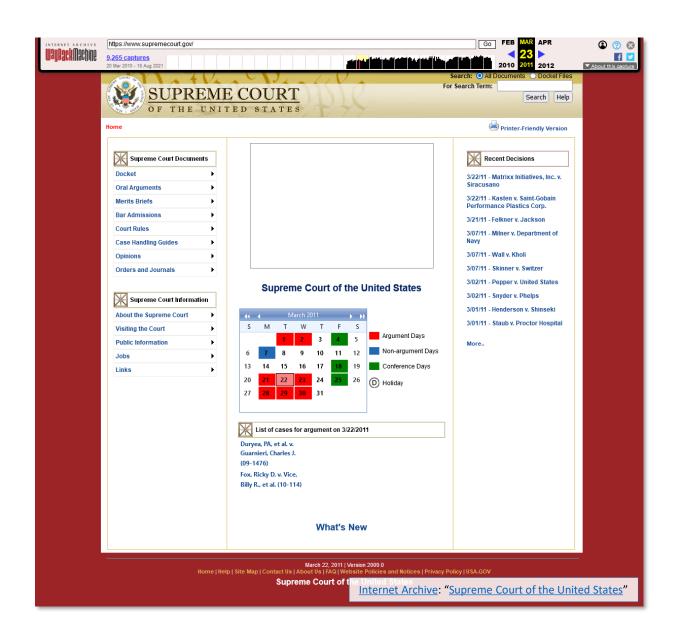
IAWM calendar page: 23 March 2011 capture

- hovering over a date for which there is at least one capture presents the number of captures + capture time(s)
- clicking on a capture time takes you to the re-presentation of the resource captured on that date at that time
- let's click through



IAWM resource page

- IAWM resource page for <u>www.supremecourt.gov</u> for the 23 March 2011 capture
- note that archived webpages have a top-justified IAWM overlay banner to:
 - remind the user they're within the IAWM context
 - facilitate navigation
 - provide additional info about the current capture + other captures of the resource located at the same web address
- let's look at the IAWM banner



IAWM banner

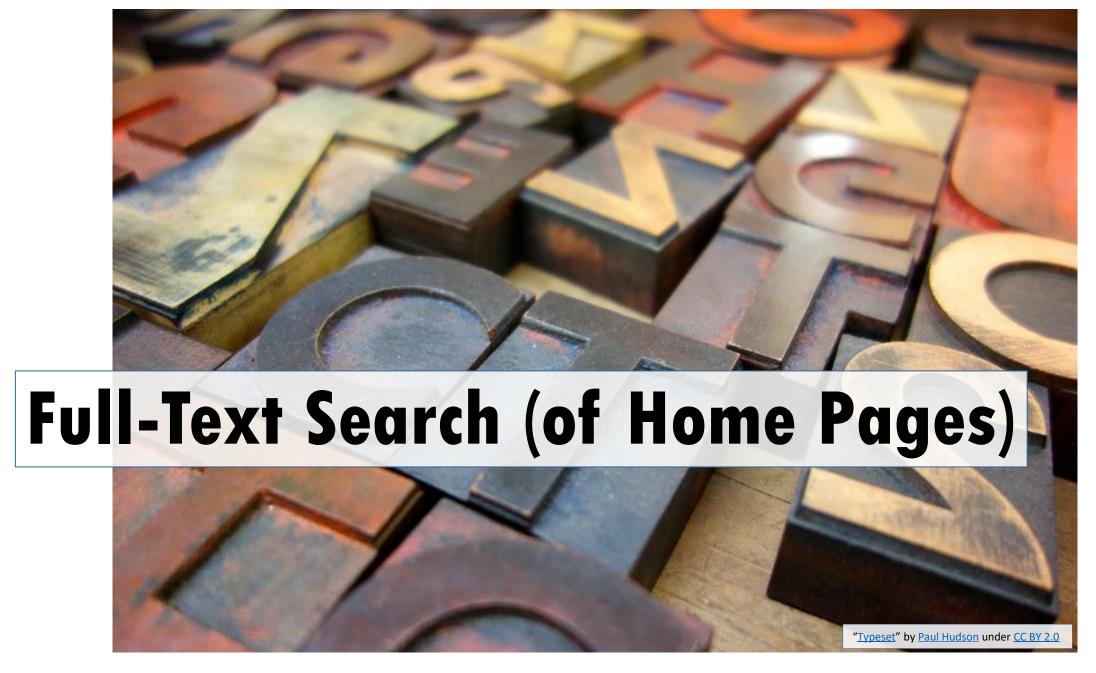


- IAWM overlay banner for the resource page for the 23 March 2011 capture of <u>www.supremecourt.gov</u>
- offers many of the same features as the calendar page: search box, captures range, timeline histogram

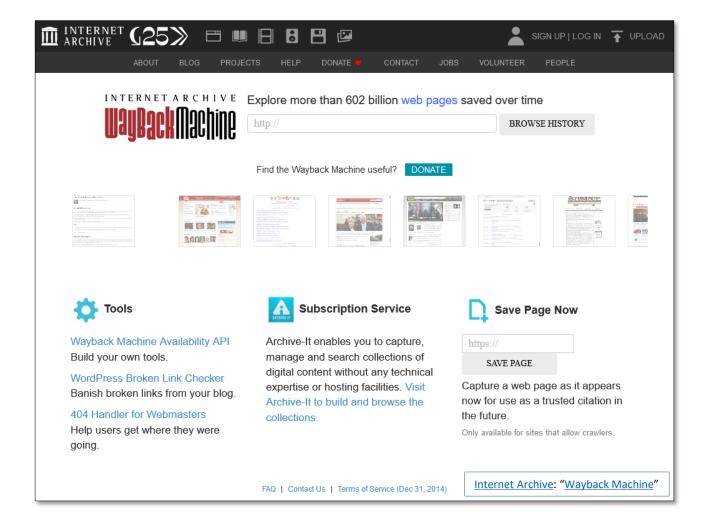
IAWM banner: temporal navigation



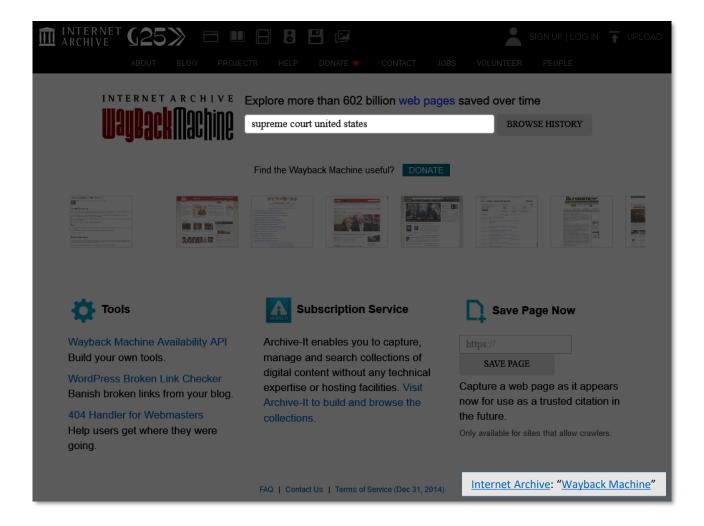
- yellow highlighted date is the capture date of the currently accessed web resource
- to the left and right are temporal navigation links for different intervals – from top to bottom:
 - one month ago / one month ahead
 - last preceding capture / next successive capture
 - one year ago / one year ahead
- for less-frequently captured web resources, intervals may be longer than a month or year



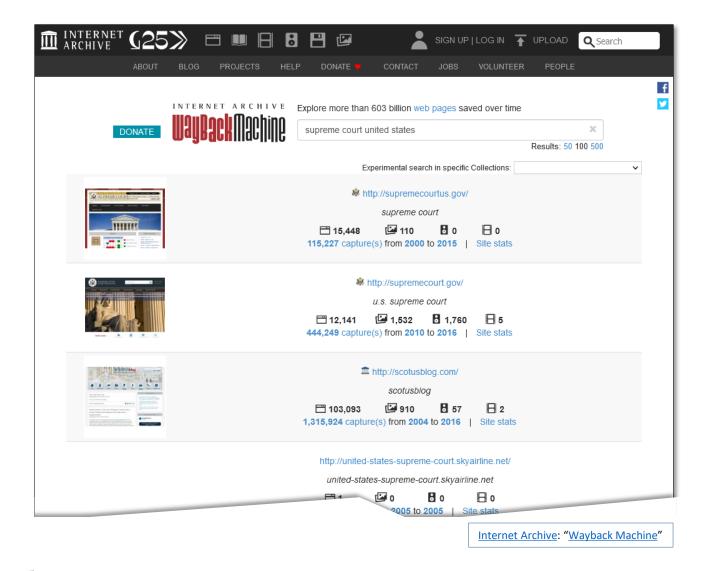
full-text search (of home pages)



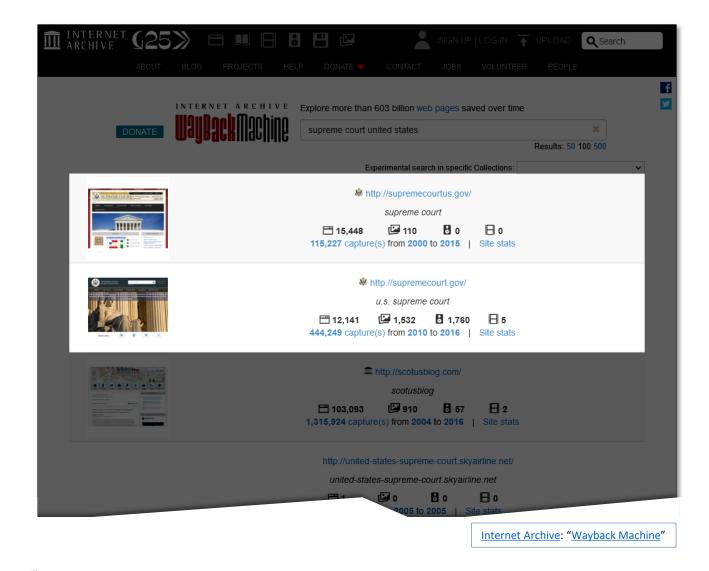
full-text search (of home pages): enter search string



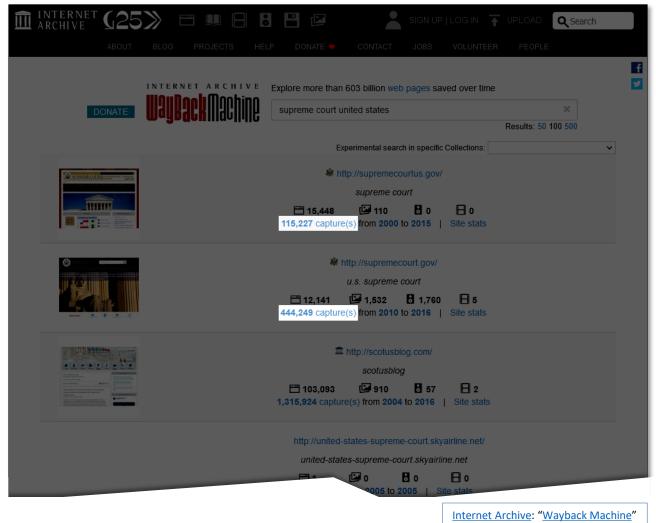
full-text search (of home pages): search results



full-text search (of home pages): top two results

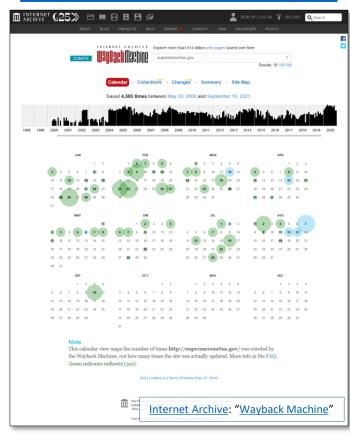


full-text search (of home pages): capture counts for top two results



IAWM calendar pages

supremecourtus.gov

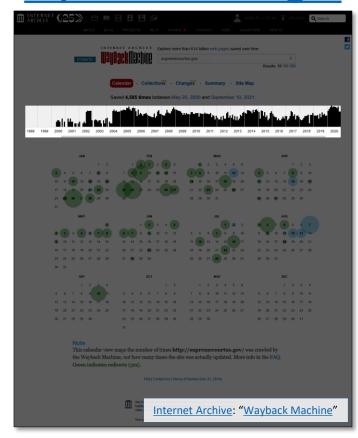


supremecourt.gov

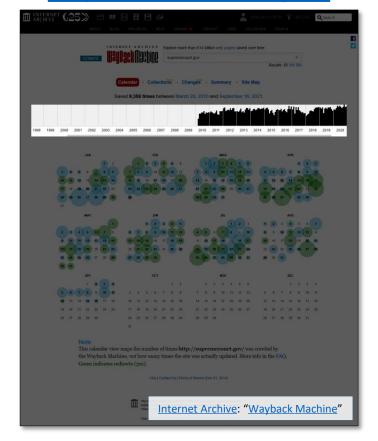


IAWM calendar pages: timeline histograms

supremecourtus.gov



supremecourt.gov



IAWM resource page

 this is the first capture for <u>supremecourtus.gov</u>, from 20 May 2000





Collections Tool + Full-Text Search

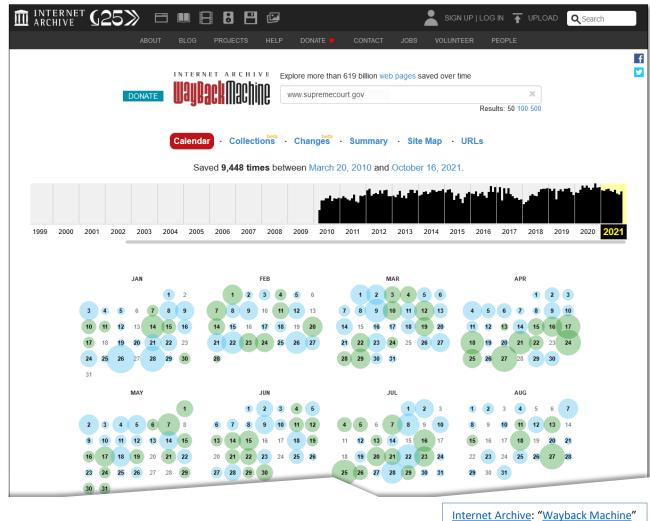


Collections tool

- IAWM captures have diverse provenance:
 - web-wide crawls
 - contract crawls
 - Archive-It crawls
 - Alexa donations
 - ArchiveTeam crawls
 - Save Page Now
- Collections tool shows sources of captures

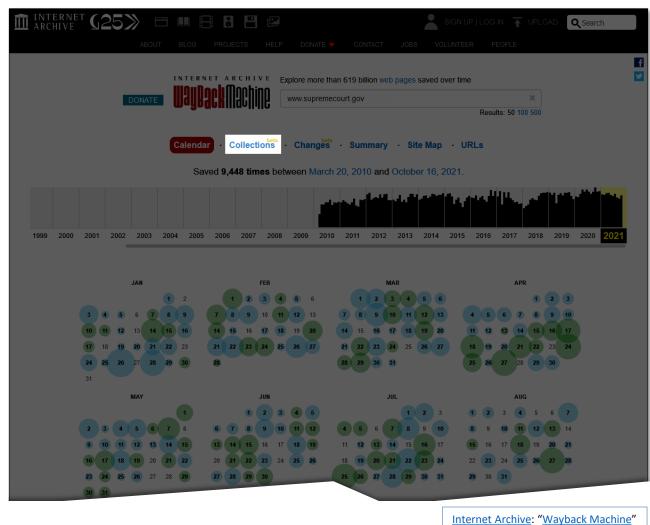
- why does provenance matter?
 - some sources have more focused crawls, indicating quality
 - some sources offer additional search + discovery options

IAWM calendar page

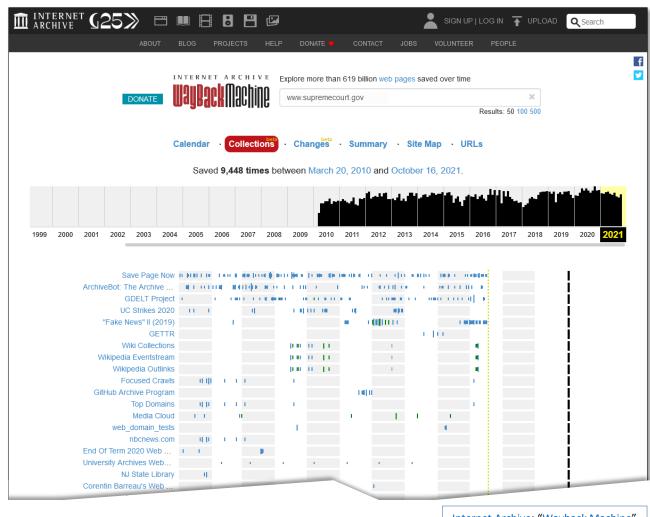


#ABATECHSHOW

IAWM calendar page



Collections tool



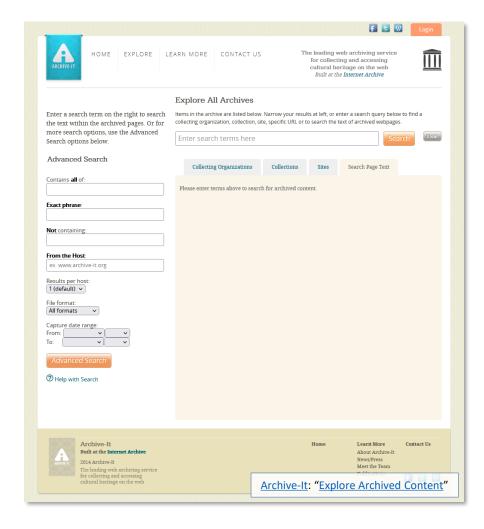
Internet Archive: "Wayback Machine"

Collections tool

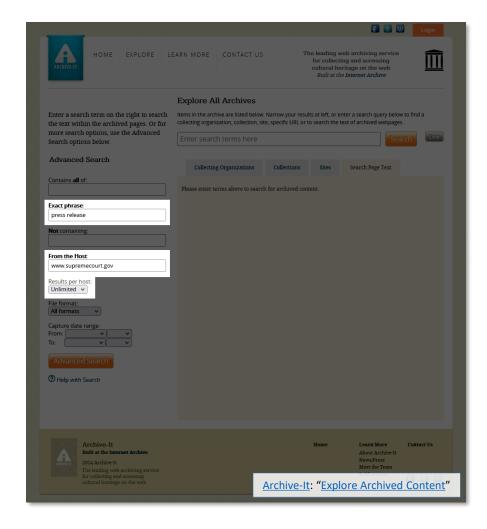
- hover over individual collection names
- look for string in link target web address: Archivelt-Collection
- Archive-It is IA's subscription web archiving service
- used by 100s of orgs to run selective web crawls + curate topical collections
- has <u>its own discovery portal</u> w/ <u>full-text search</u>



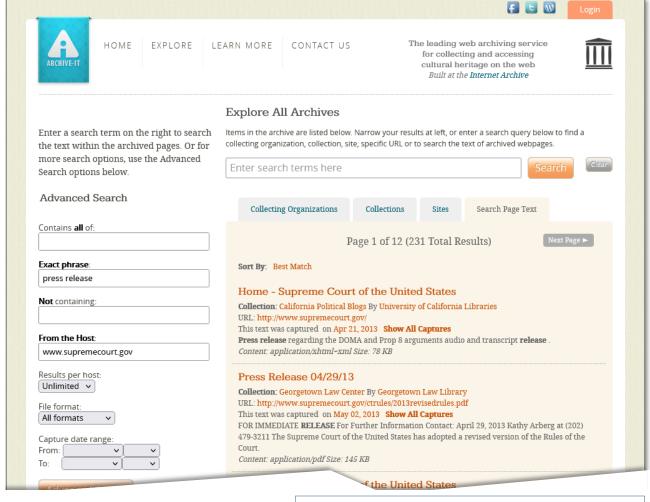
Archive-It full-text search



Archive-It full-text search



Archive-It full-text search



Archive-It: "https://archive-it.org/explore/?show=ArchivedPages..."

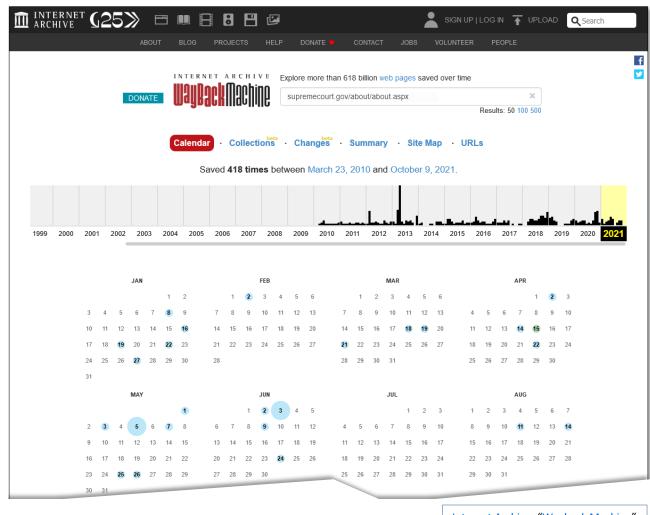


Changes tool

- flags differences in text b/t 2 captures of same web page
- caveats:
 - doesn't consider non-textual elements
 - doesn't account for intermediate changes
 - changes may not matter for legal issue

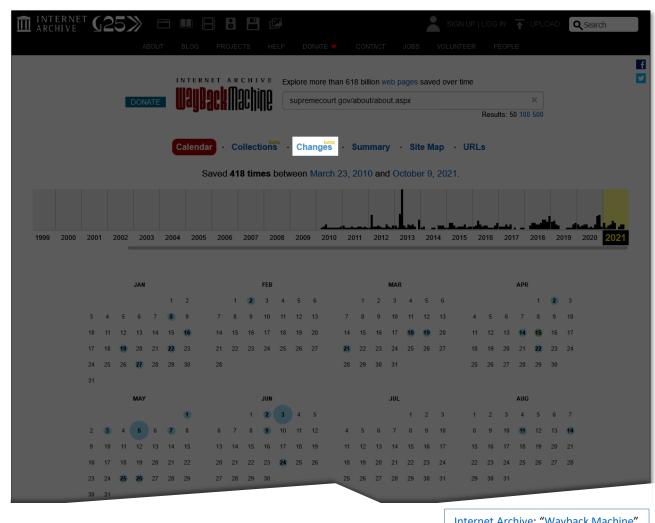
- use to:
 - identify when a web page changed
 - quickly visualize the text changes over time
 - establish when web page text was substantially identical

IAWM calendar page

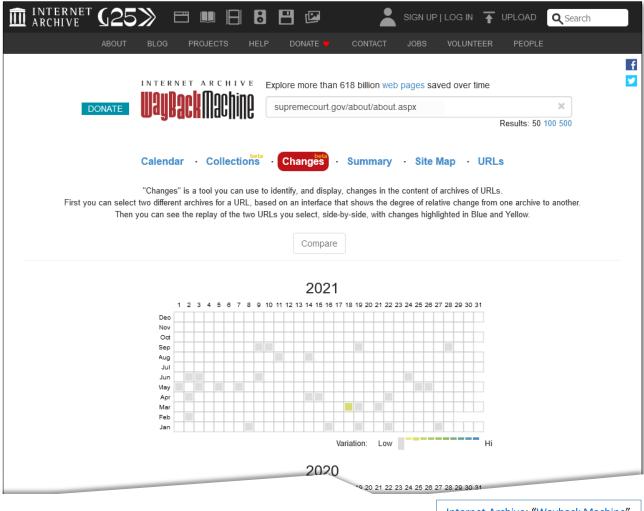


Internet Archive: "Wayback Machine"

IAWM calendar page

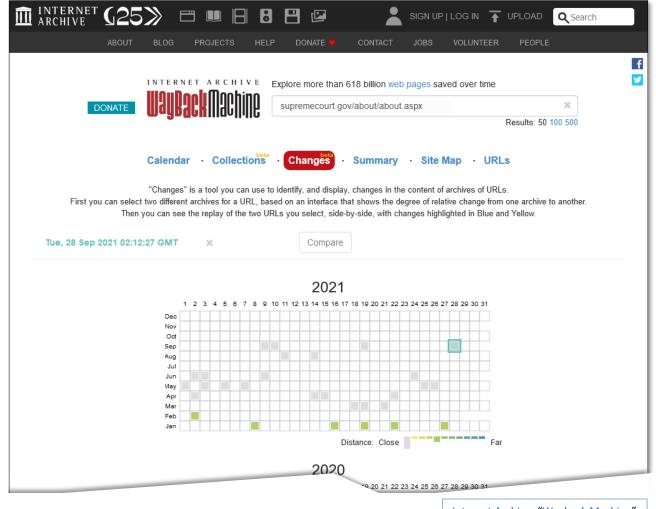


Changes tool



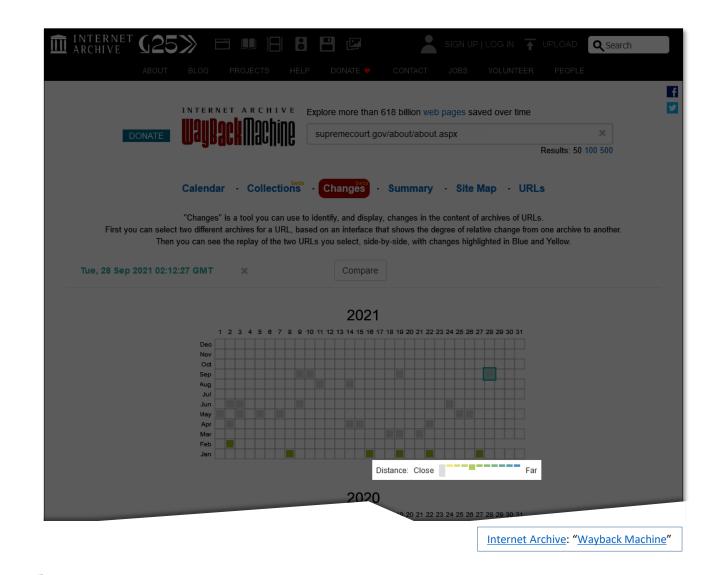
Internet Archive: "Wayback Machine"

Changes tool: choose 28 September 2021 capture

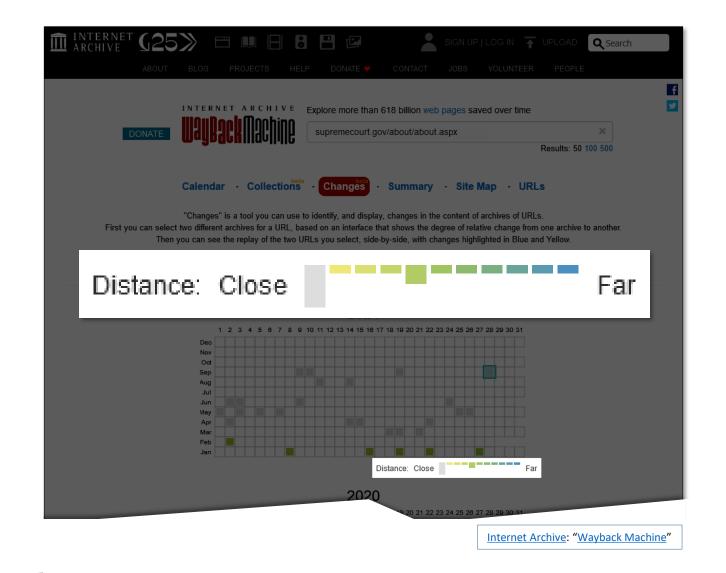


Internet Archive: "Wayback Machine"

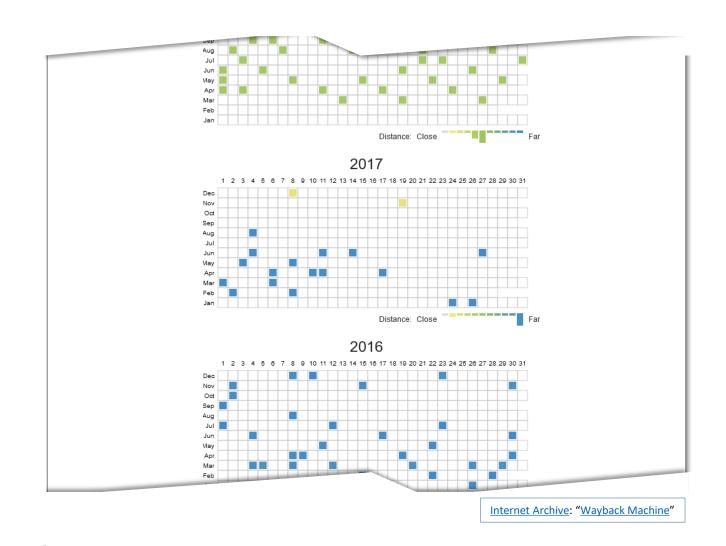
Changes tool: degree of difference key



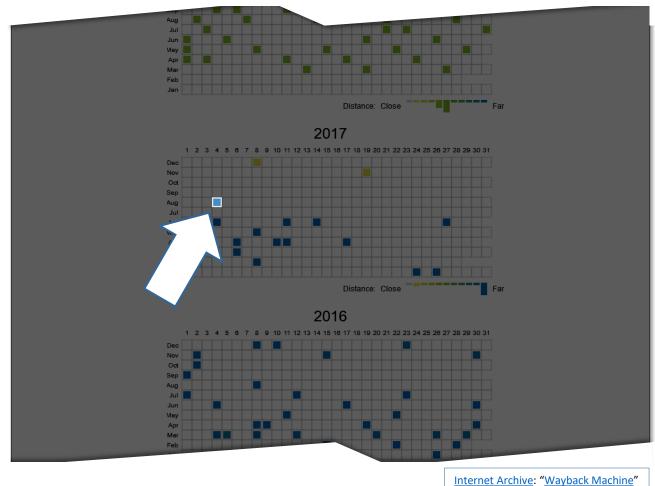
Changes tool: degree of difference key



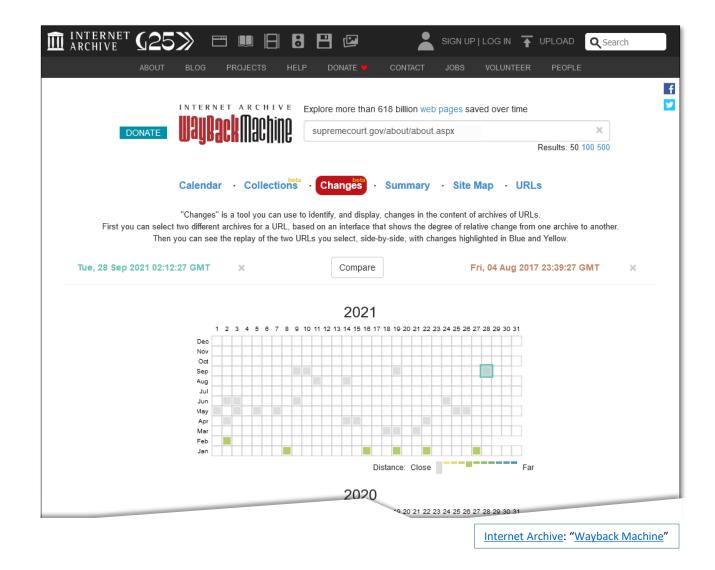
Changes tool: difference from 2017 captures



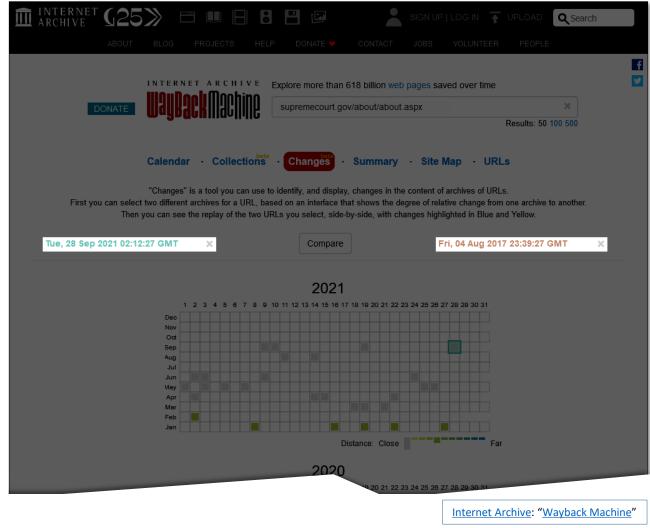
Changes tool: compare w/ 4 August 2017 capture



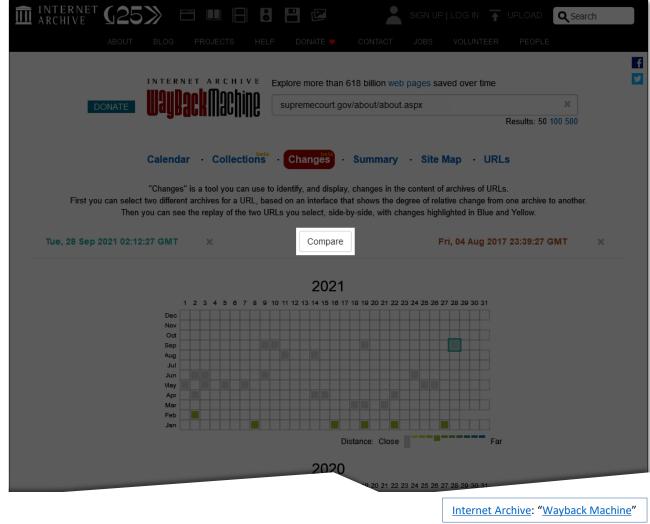
Changes tool: two captures selected



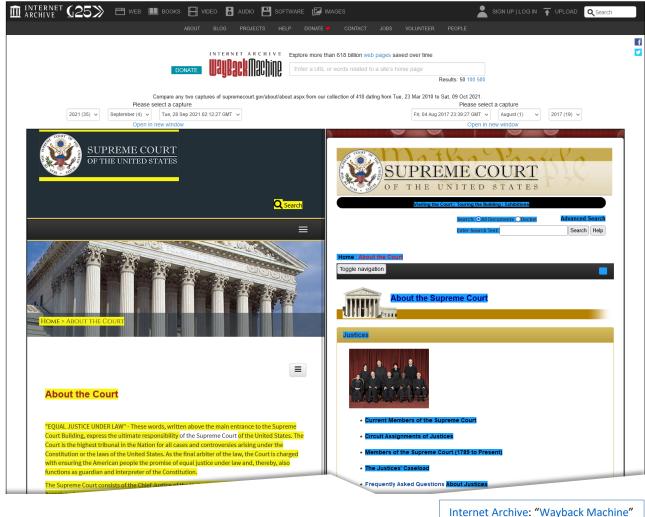
Changes tool: two captures selected



Changes tool: click *Compare*



Changes tool: compare 28 Sep. 2021 + 4 Aug. 2017 captures

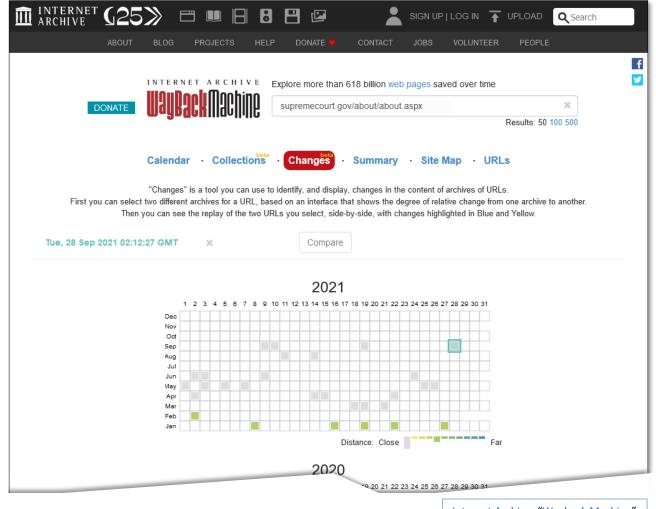


Changes tool: comparing captures

- yellow highlighted text was deleted
- blue highlighted text was added
- note that the website design has changed significantly, but this isn't specifically detected by the Changes tool
- however, major text changes often accompany major website design changes

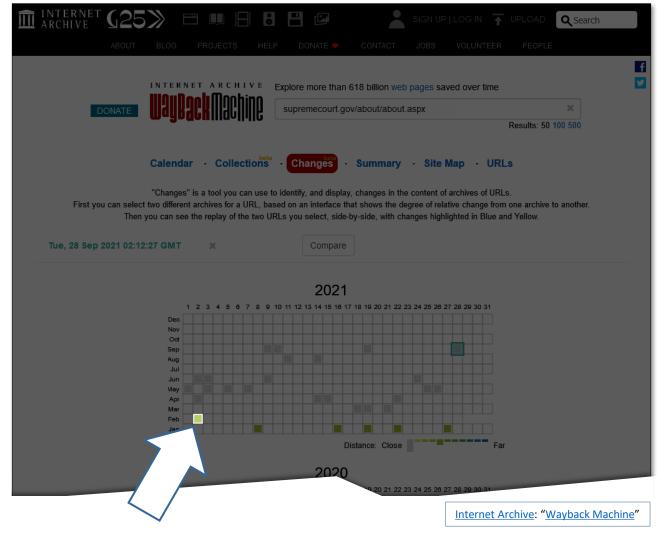


Changes tool: choose 28 September 2021 capture

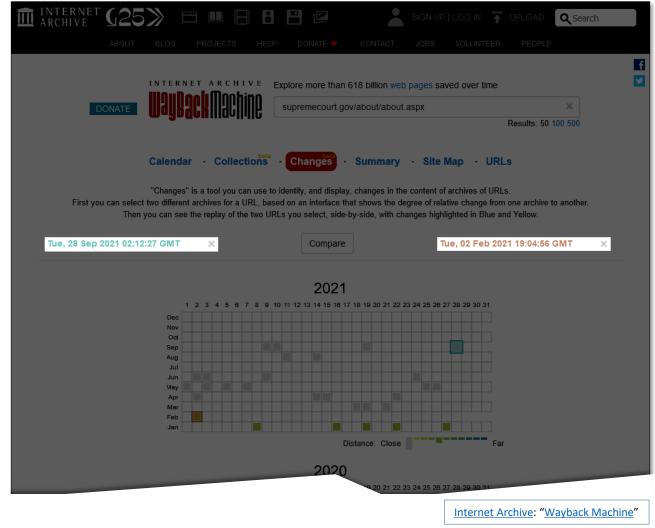


Internet Archive: "Wayback Machine"

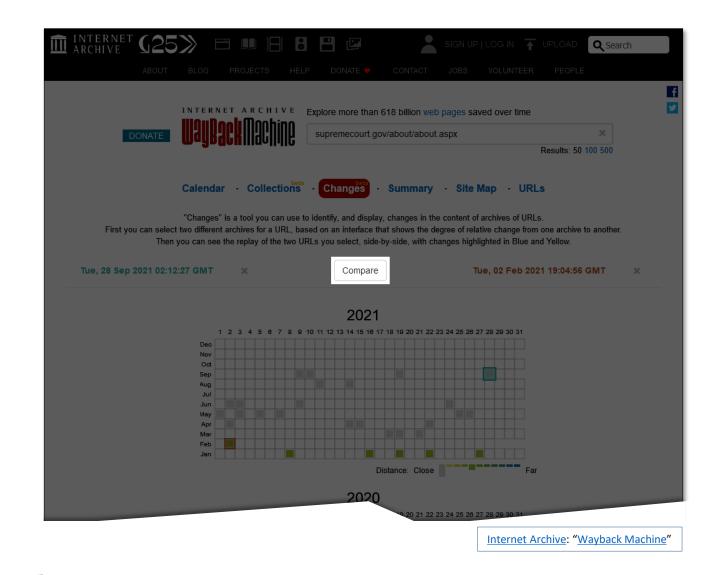
Changes tool: compare w/ 2 February 2021 capture



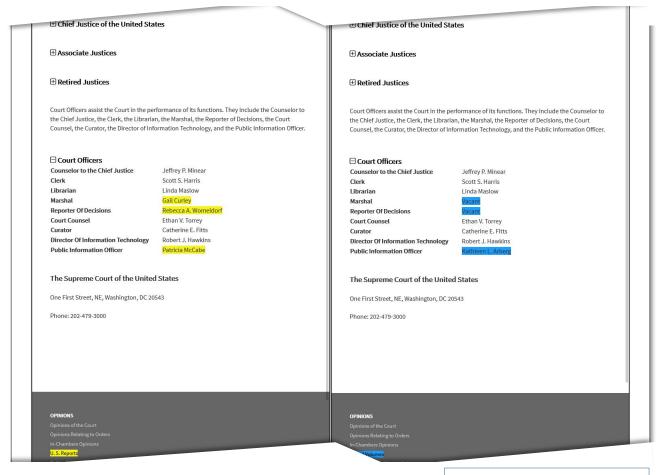
Changes tool: two captures selected



Changes tool: click *Compare*



Changes tool: compare 28 Sep. 2021 + 2 Feb. 2021 captures



Internet Archive: "Wayback Machine" #ABATECHSHOW



authentication + admissibility overview

- IAWM cited in 300+ U.S. federal court cases
- harder to authenticate than other web content
- strategies include:
 - judicial notice
 - IA affidavit
 - witness w/ personal knowledge
 - expert witness
 - mutual stipulation



judicial notice (FRE 201)

Wikipedia

 "Judicial notice is a rule in the law of evidence that allows a fact to be introduced into evidence if the truth of that fact is so notorious or well known, or so authoritatively attested, that it cannot reasonably **be doubted**...Facts and materials admitted under judicial notice are accepted without being formally introduced by a witness or other rule of evidence"

FRE 201

Judicial Notice of Adjudicative Facts

- b. Kinds of Facts That May Be Judicially Noticed. The court may judicially notice a fact that is not subject to reasonable dispute because it:
 - 2. can be accurately and readily determined from sources whose accuracy cannot reasonably be questioned.

Wikipedia: "Judicial notice"

Legal Information Institute: "Rule 201. Judicial Notice of Adjudicative Facts | Federal Rules of Evidence"

authentication via judicial notice

- efficient but inconsistent + jurisdiction-sensitive
- vital to research a given court's disposition towards IAWM evidence
- judges may themselves introduce IAWM evidence
- more likely to prevail in absence of specific assertions of unreliability

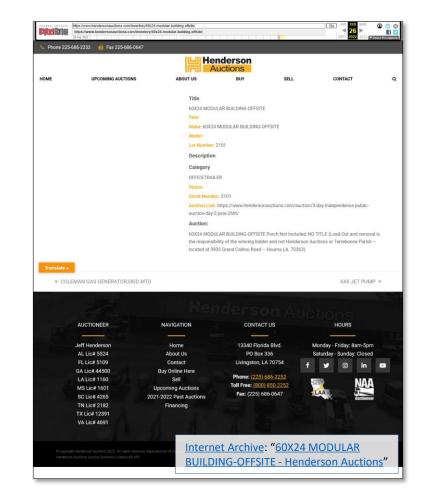


Weinhoffer v. Davie Shoring, Inc. (2022)

- IAWM used to show online auction terms of sale
- district court's admission of IAWM evidence via judicial notice reversed

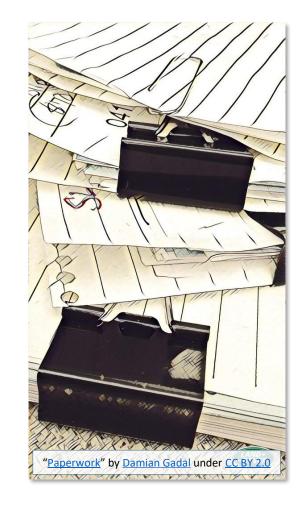
"Here, there was no testimony to authenticate the archived webpage. Our sister circuits' decisions that the Wayback Machine is not self-authenticating are persuasive in the context of judicial notice. In sum, the district court erred in taking judicial notice of the terms because a private internet archive falls short of being a source whose accuracy cannot reasonably be questioned as required by Rule 201."

Weinhoffer v. Davie Shoring, Inc., No. 20-30568, at *7 (5th Cir. Jan. 20, 2022)



authentication via IA affidavit

- reliable + efficient
- incorrect application explains most of its authentication failures
- works best for simple cases
- but less ideal when conclusions require interpretation



IA affidavit guidance

Do I really need an affidavit from the Internet Archive?

No. Please consider alternatives to an affidavit from the Internet Archive. Judicial notice and stipulation to a document's authenticity are two typical and straightforward options that might be used instead of an affidavit. Since our resources are limited, we urge you to pursue these alternatives before coming to us with authentication requests.

Does the Internet Archive's affidavit mean that the printout was actually the page posted on the Web at the recorded time?

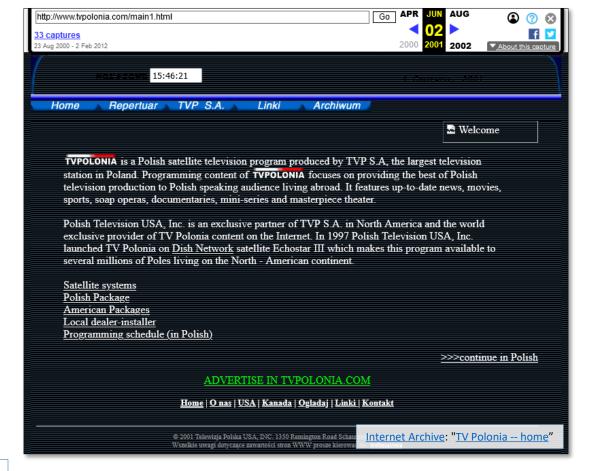
The Internet Archive's affidavit only affirms that the printed document is a true and correct copy of our records. It remains your burden to convince the finder of fact what pages were up when.

Internet Archive: "Legal FAQ"

Telewizja Polska USA, Inc. v. Echostar Satellite Corp. (2004)

- IAWM used to show advertisement of partner business service after contract expiration
- IAWM evidence admitted

"...Plaintiff has presented no evidence that the Internet Archive is unreliable or biased. And Plaintiff has neither denied that the exhibit represents the contents of its website on the dates in question, nor come forward with its own evidence challenging the veracity of the exhibit. Under these circumstances, the Court is of the opinion that [the IA affidavit] is sufficient to satisfy Rule 901's threshold requirement for admissibility. Plaintiff is free to raise its concerns regarding reliability with the jury."



<u>Telewizja Polska USA, Inc. v. Echostar Satellite</u> <u>Corp.</u>, 02 C 3293, at *14 (N.D. III., Jun. 3 2004)

witness w/ personal knowledge — FRE 901(b)(1)

Authenticating or Identifying Evidence

- b) Examples. The following are examples only not a complete list of evidence that satisfies the requirement:
 - 1) Testimony of a Witness with Knowledge. Testimony that an item is what it is claimed to be.

Legal Information Institute: "Rule 901. Authenticating or Identifying Evidence | Federal Rules of Evidence"

authentication via witness w/ personal knowledge

- more effective when leveraging witness w/ personal knowledge of historical web content
- less effective when instrumentally entering into record a comparison performed by a random person

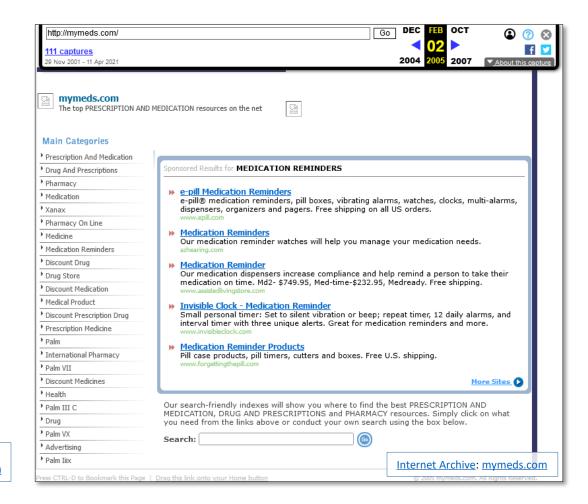


United States v. Bansal (2011)

- IAWM used to show operation of an online pharmacy
- IAWM evidence admitted

"To authenticate that the screen-shot was what it purported to be, the government called a witness to testify about how the Wayback Machine website works and how reliable its contents are. The witness also compared the screenshots with previously authenticated and admitted images from Bansal's website and concluded, based upon her personal knowledge, that the screenshots were authentic. This was evidence 'sufficient to support a finding' that the screenshots were 'what they purport[ed] to be,' rendering them admissible under Rule 901(b)(1). Accordingly, we will affirm the District Court."

<u>United States v. Bansal, 663 F.3d</u> 634, 652, at *667-668 (3d Cir.2011)



expert witnesses — FRE 702

Testimony by Expert Witnesses

- a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- b) the testimony is based on sufficient facts or data;
- c) the testimony is the product of reliable principles and methods; and
- d) the expert has reliably applied the principles and methods to the facts of the case.

Legal Information Institute: "Rule 702. Testimony by Expert Witnesses | Federal Rules of Evidence"

authentication via expert witness

- reliable but more involved + less-used mechanism
- use when interpretation, explanation, or qualification needed
- expert witnesses have included:
 - digital forensics specialists
 - patent examiners
 - web archiving professional(s)



Marten Transport, Ltd. v. PlattForm Advertising, Inc. (2016)

- IAWM used to show ongoing use of mark over period of time
- IAWM evidence admitted

"PlattForm argues that Mr. Fischer, in opining that certain websites referenced Marten in the past as shown by the Wayback Machine, improperly acts as a surrogate for the experts who created the Internet Archive.

The Court rejects this challenge. PlattForm relies on Khoday v.

Symantec Corp., 93 F.Supp.3d 1067 (D.Minn.2015), but in that case, the court rejected an argument that the expert was not qualified because he had not worked at the Internet Archive to design the Wayback Machine archives or the web crawler that it used...PlattForm notes that the expert in Khoday at least had experience creating web archives, but the court in that case did not suggest that such archiving experience was required...In this case, Mr. Fischer has experience using tools to recover electronic data, including the use of the Wayback Machine, and the Court finds such experience to be sufficient to withstand a Daubert challenge."

Marten Transp., Ltd. v PlattForm Advertising, Inc., 184 F. Supp. 3d 1006, at *1009 (D. Kan. 2016)



Frye + Daubert standards

Frye standard

- established in <u>Frye v. United</u> <u>States</u>, 293 F. 1013 (D.C. Cir. 1923)
- "...the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."

Daubert standard

- established in <u>Daubert v. Merrell</u> <u>Dow Pharmaceuticals, Inc.</u>, 509
 U.S. 579 (1993)
- "...the trial judge...must make a preliminary assessment of whether the testimony's underlying reasoning or methodology is scientifically valid and properly can be applied to the facts at issue."

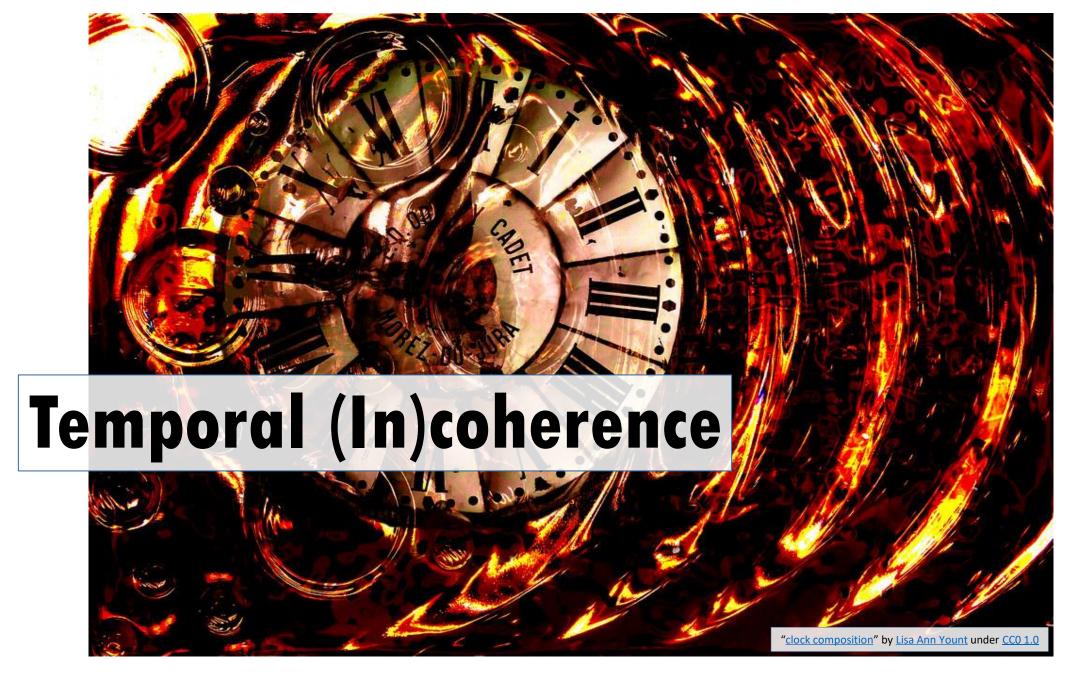
<u>Legal Information Institute</u>: "Frye Standard"

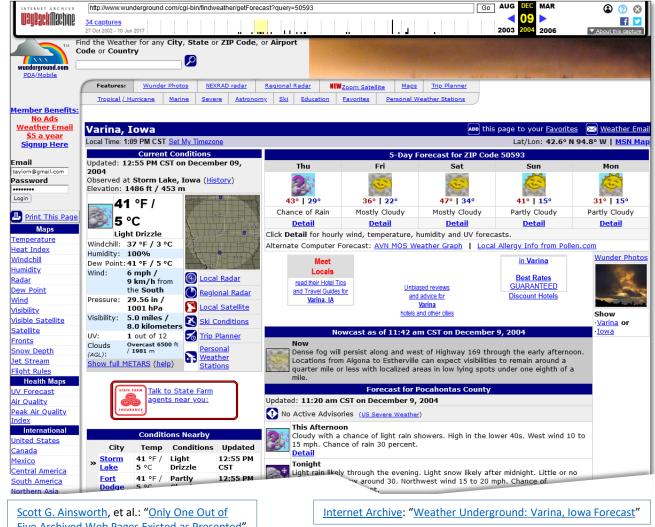
<u>Legal Information Institute: "Daubert Standard"</u>

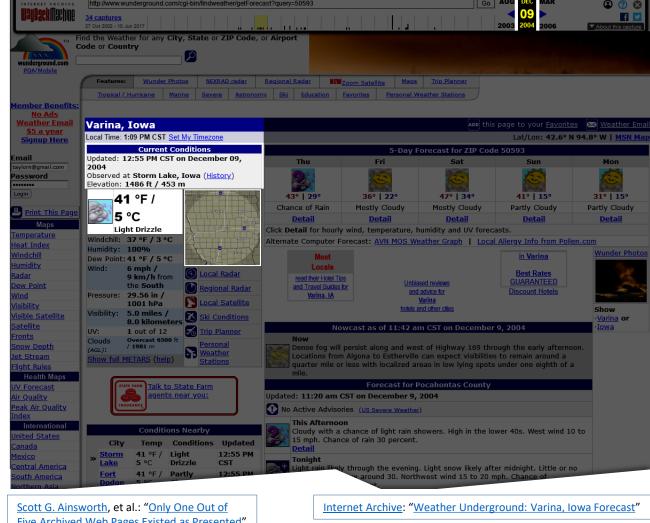
admissibility trends summary — U.S. federal courts

	judicial notice		IA standard affidavit		witness w/ personal knowledge		expert witness	
admitted?	yes	no	yes	no	yes	no	yes	no
U.S. district court	47	12	14	3	6	4	7	0
U.S. circuit court	1	2	2	0	3	1	2	0

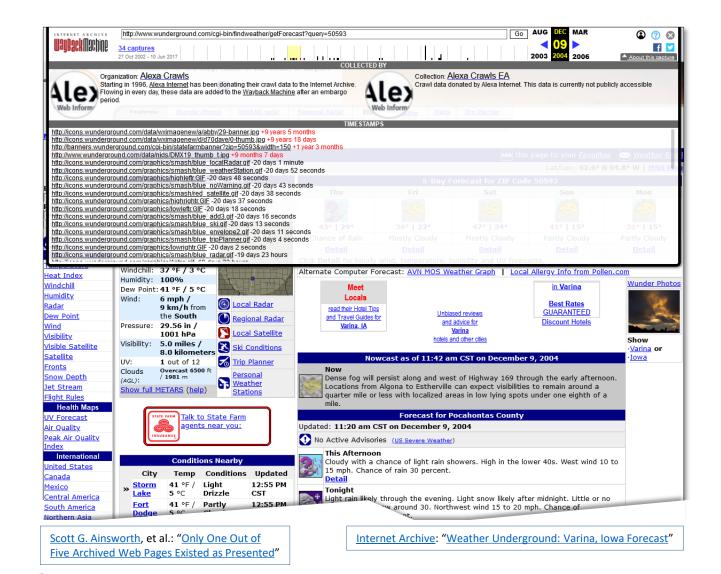
- based on analysis of ~300 U.S. federal court cases mentioning IAWM
- affidavit + expert witness succeed most consistently
- examples of all strategies succeeding at circuit court level, but <u>recent ruling</u> will likely make judicial notice less successful going forward

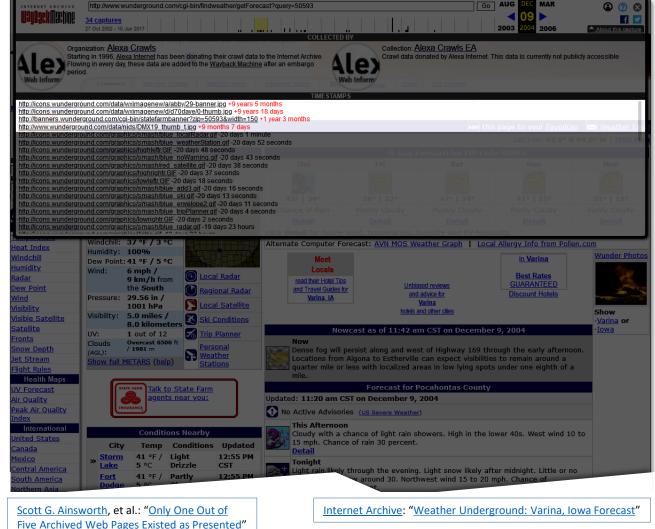


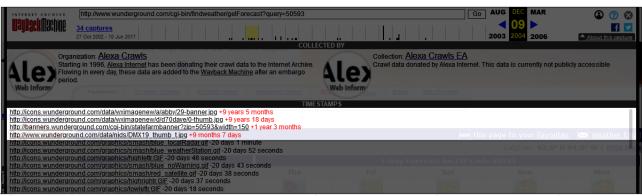






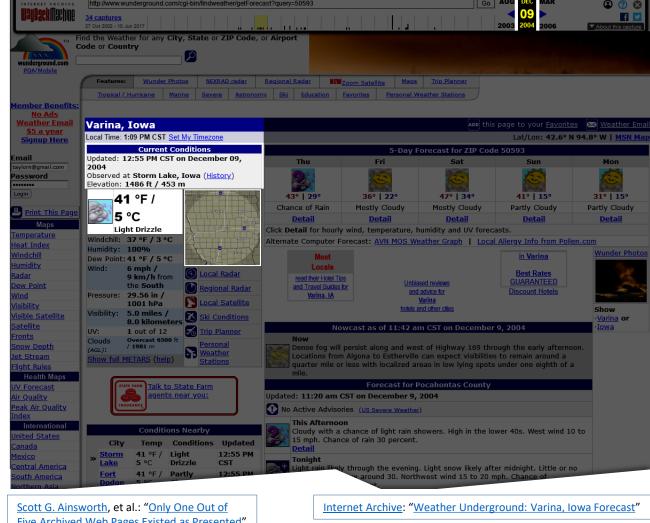


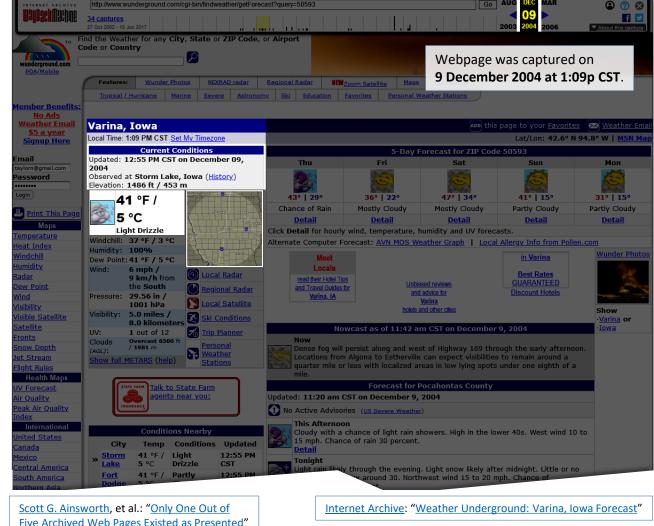


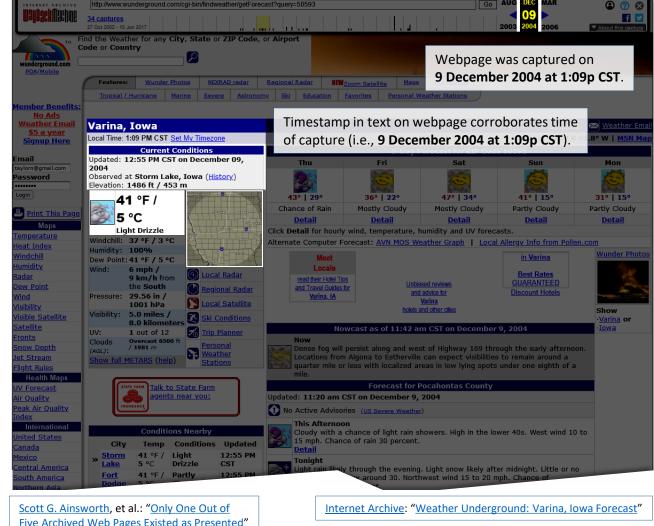


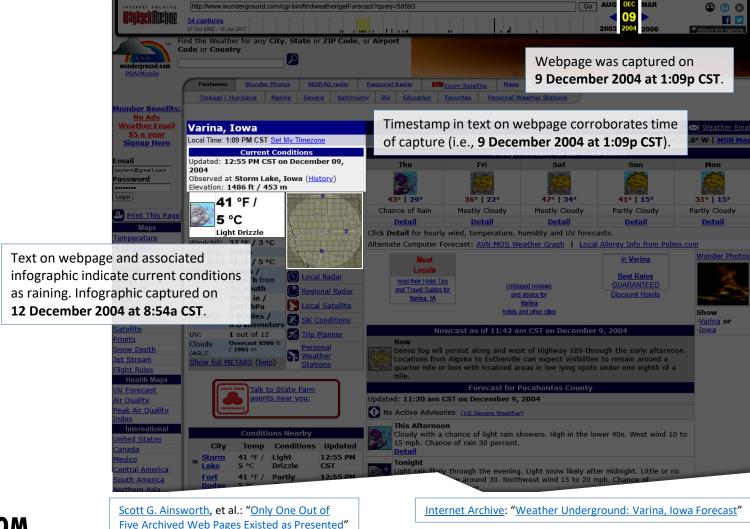
http://icons.wunderground.com/data/wximagenew/a/abby/29-banner.jpg +9 years 5 months http://icons.wunderground.com/data/wximagenew/d/d70dave/0-thumb.jpg +9 years 18 days http://banners.wunderground.com/cgi-bin/statefarmbanner?zip=50593&width=150 +1 year 3 months http://www.wunderground.com/data/nids/DMX19_thumb_t.jpg +9 months 7 days

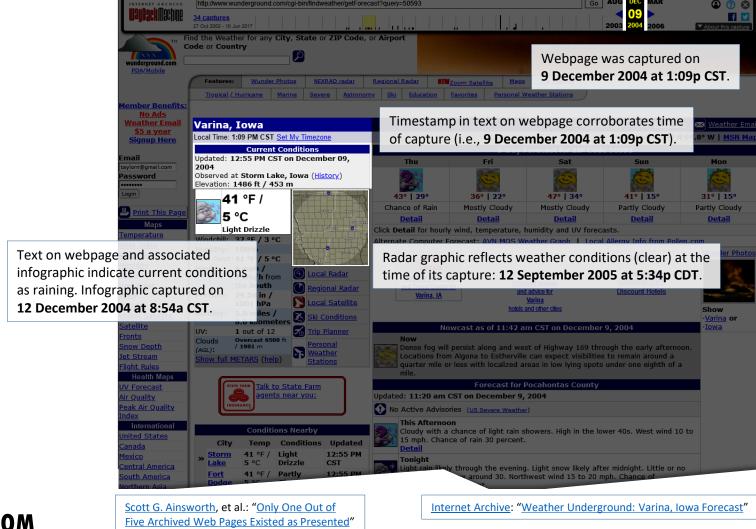






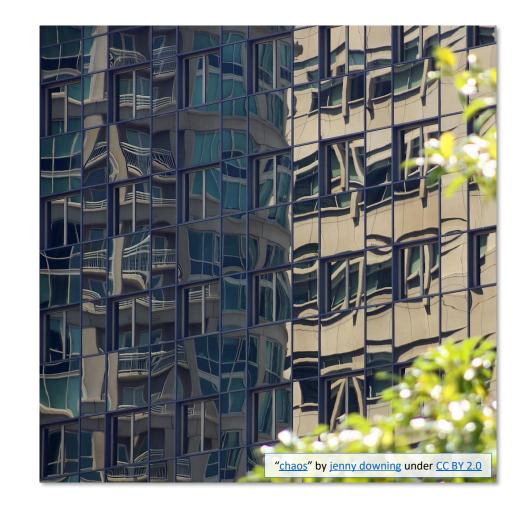






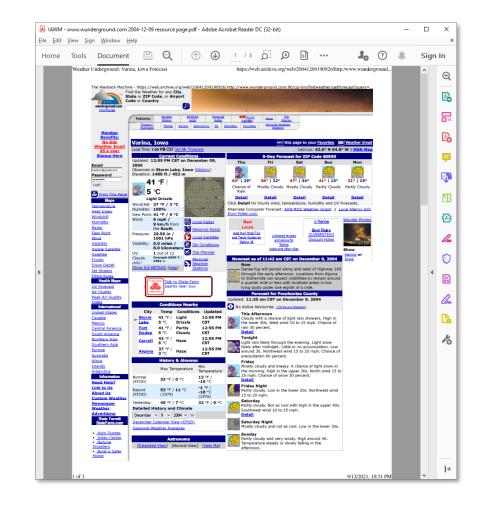
what's going on here?

- archiving of web resources is time-staggered
- IAWM reconstitutes archived webpage w/ most temporally proximate elements
- IAWM may therefore re-present fictitious web pages
- examine capture dates + assess materiality of temporal incoherence



problems w/ PDF

- IAWM evidence submitted as PDF screenshots
- temporal variations no longer apparent
- if temporal coherence is an issue, provide supplementary info



how to have admitted

- IA will still provide an affidavit
- however, this IAWM evidence is more easily contested
- or viewed as less credible by judge
- consider an expert witness





sources of temporal information

conspicuous

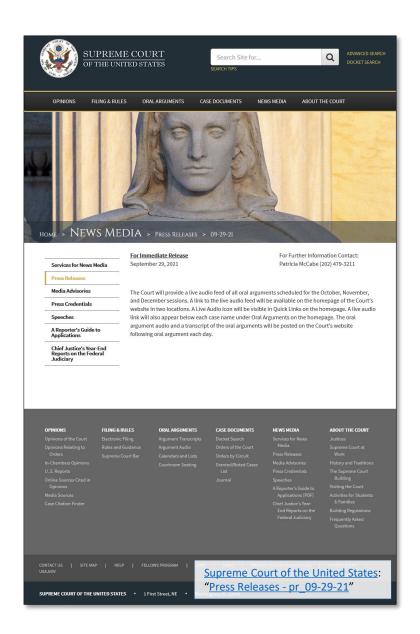
- (system-generated) datetime string in webpage text
- IAWM capture timestamp

less obvious

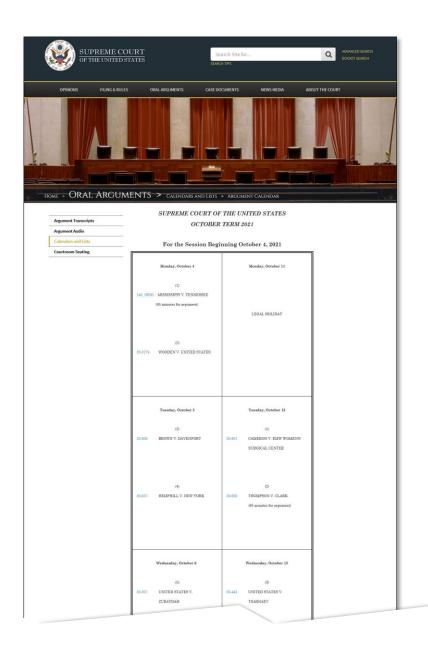
- (system-generated) datetime string in webpage source code
- Google <u>before</u>: operator
- Twitter <u>until</u>: operator
- IAWM capture datetime of webpage w/ inlinks
- web server

 client network
 headers (e.g., X-Archive-Orig-last modified)
- timestamps in embedded image metadata (e.g., <u>Exif</u>, <u>XMP</u>)

- when was the webpage for the oral argument live audio stream first made available?
- press release said it would go live at the start of October Term 2021
- link would be published on the home page



- first oral arguments of the term were scheduled for 4 October 2021
- stream presumably went live no later than that, but what about the webpage?



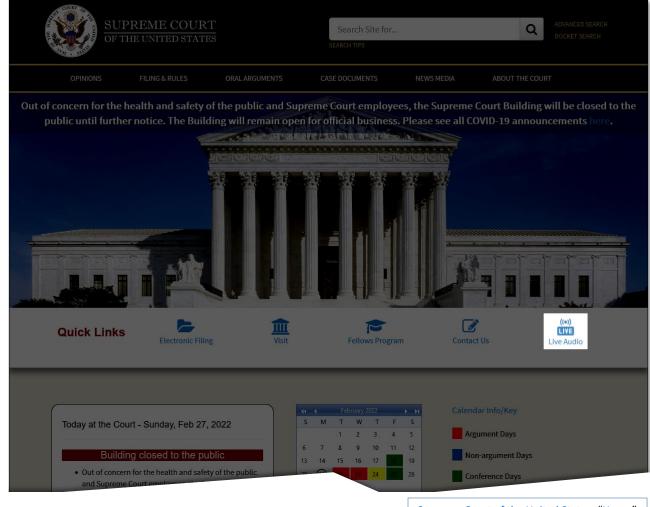
Supreme Court of the United States: "Monthly Argument"



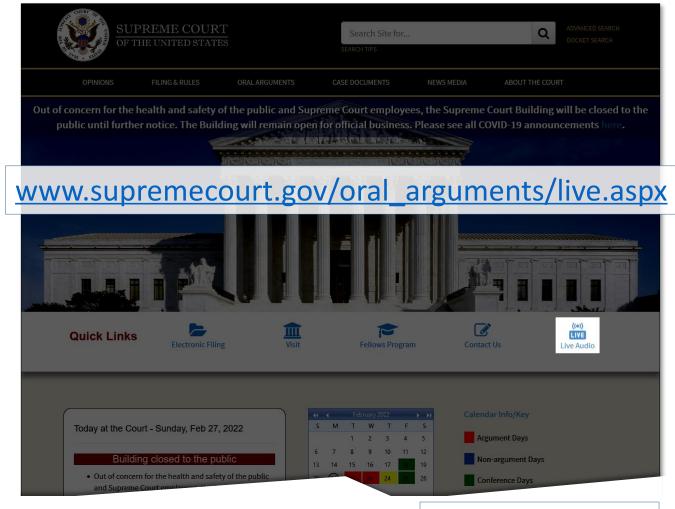
WWW.TECHSHOW.COM

Supreme Court of the United States: "Home"

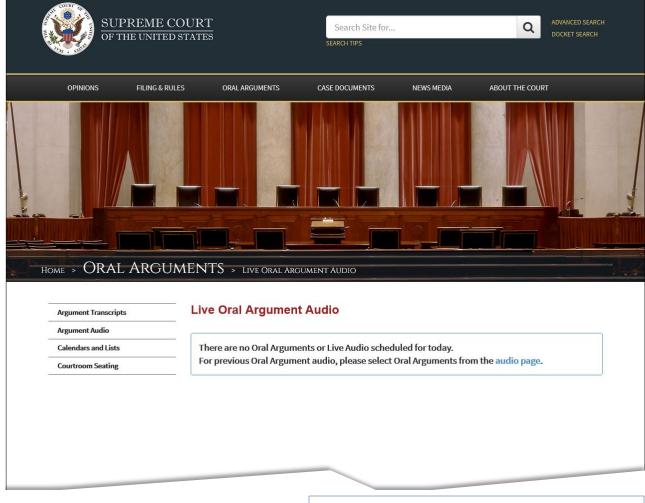
#ABATECHSHOW



Supreme Court of the United States: "Home"



Supreme Court of the United States: "Home"



Supreme Court of the United States: "Live Oral Argument Audio"

#ABATECHSHOW WWW.TECHSHOW.COM

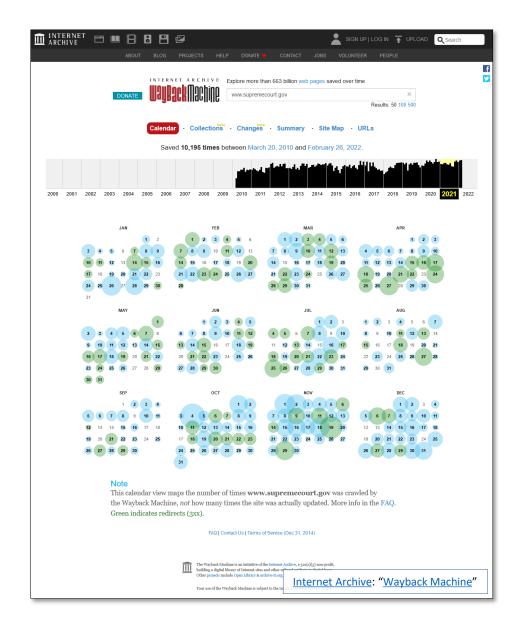
when did link appear on home page?



Supreme Court of the United States: "Home"

when did link appear on home page?

- brute force (i.e. try insideoutside dates, then narrow down)
- use <u>Archive-It</u> to <u>search for</u> <u>"live audio"</u>
- use <u>CDX Server API</u> to bulk retrieve captures, then run local full-text search for "live audio"

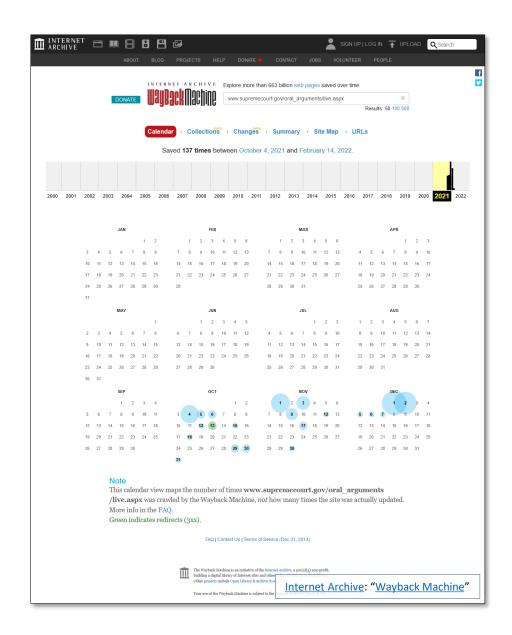


when did link appear on home page?

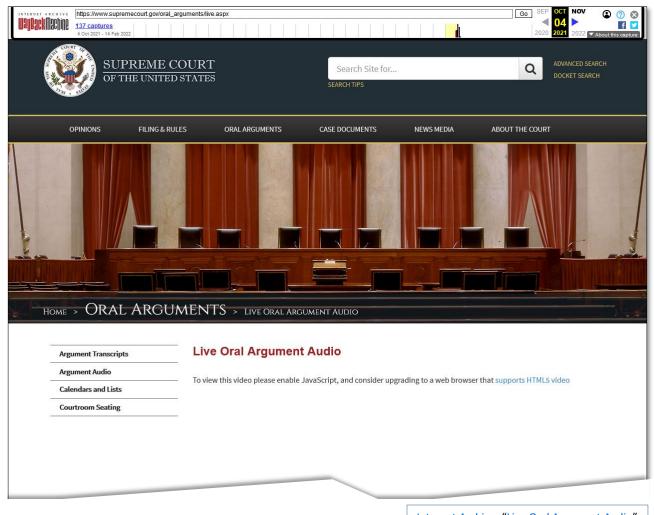


captures of webpage itself

- timeline histogram only shows captures for calendar years 2021 + 2022, up to current date
- click back to 2021
- IAWM calendar page shows captures dating back to 4 October 2021



first capture of webpage: 4 October 2021

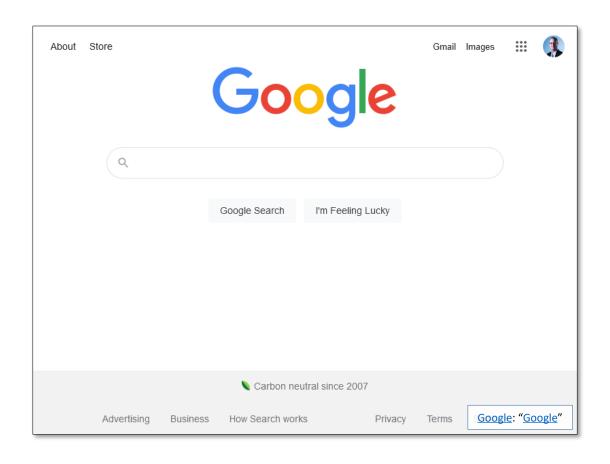


Internet Archive: "Live Oral Argument Audio"
#ABATECHSHOW

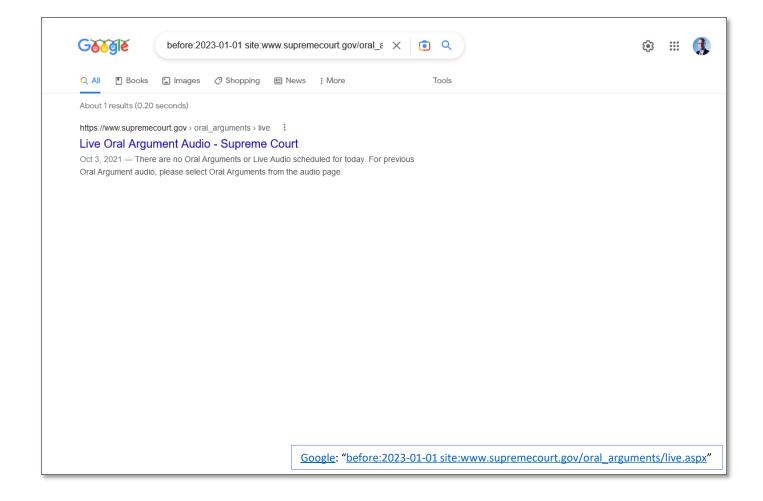


Google search operators

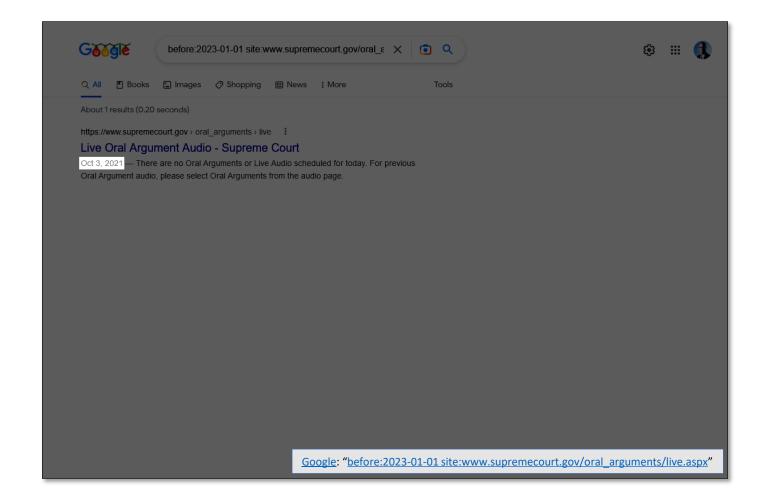
- Google search operators are specific commands that may be included in a search to refine the returned results
- the before: operator limits results to those that were updated before the specified date
- the after: operator limits results to those that were updated after the specified date
- the site: operator limits results to those w/ web addresses starting w/ the specified path



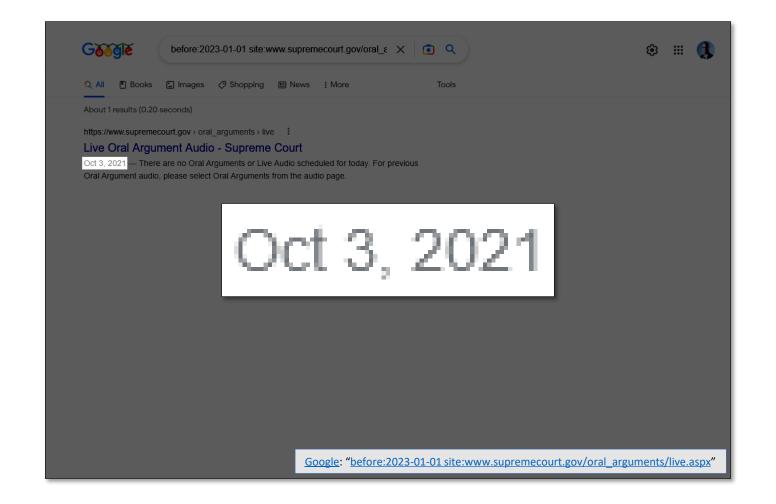
Google search before: 1 January 2023



Google search before: 1 January 2023



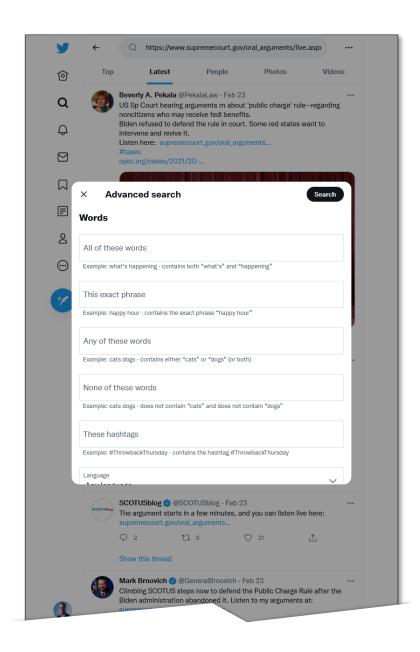
Google search before: 1 January 2023





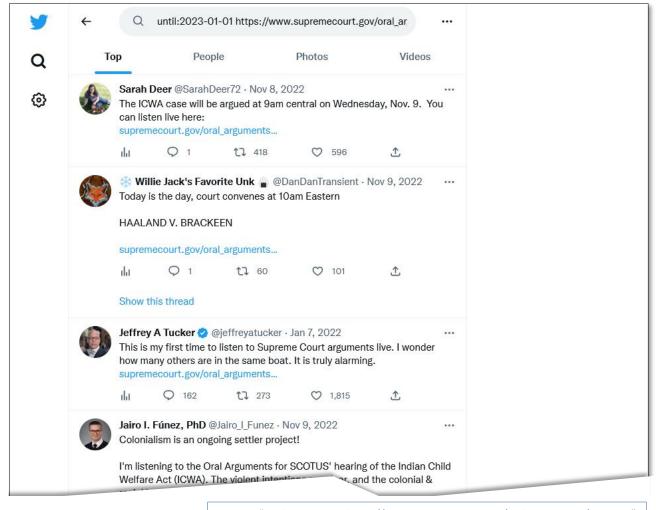
Twitter search operators

- <u>Twitter search operators</u> are specific commands that may be included in a search to refine the returned results
- the until: operator limits returned tweets to those published before the specified date
- the since: operator limits returned tweets to those published after the specified date



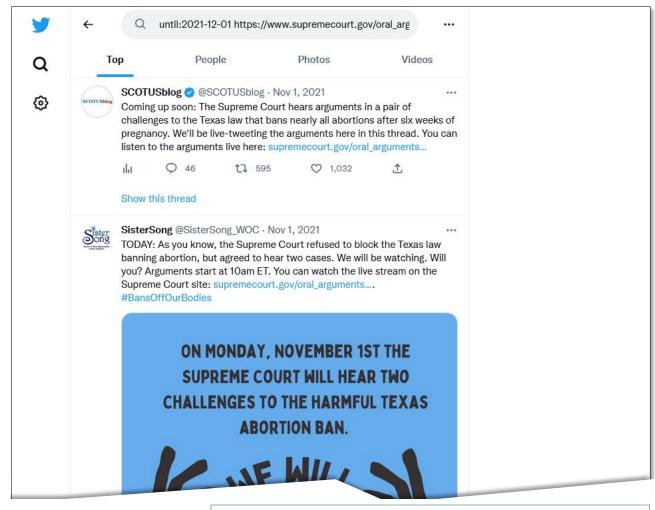
Twitter: "Twitter Search"

Twitter search *until:* 1 January 2023



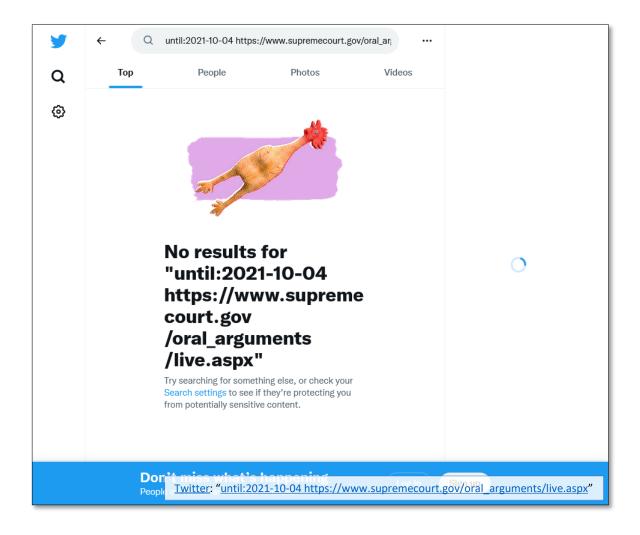
Twitter: "until:2023-01-01 https://www.supremecourt.gov/oral_arguments/live.aspx"

Twitter search until: 1 December 2021

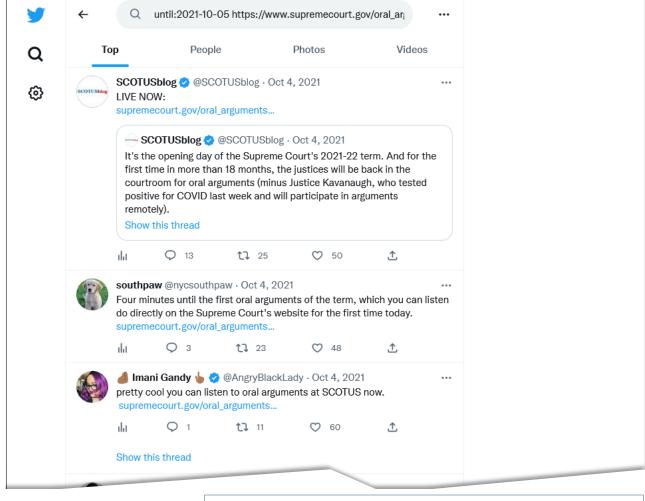


Twitter: "until:2021-12-01 https://www.supremecourt.gov/oral_arguments/live.aspx"

Twitter search until: 4 October 2021



Twitter search until: 5 October 2021



<u>Twitter</u>: "until:2021-10-05 https://www.supremecourt.gov/oral_arguments/live.aspx"

timestamping content changes

when first published

- in this example, lucky to have many traces + via diverse services
- more typical to be working w/ fragments
- key limitations on assertions:
 - generally only establish *no later than*
 - don't necessarily establish the moment when

when continuously available

- use similar techniques to document availability across successive short intervals
- continuous intermediate availability more reasonable to assume than up-down-up
- can get complicated for content situated deep in a website across redesigns or platform changes
 - content may persist, but specific web address may change



takeaways

- temporal attributes of web content can matter (a lot)
- IAWM is a powerful tool for surfacing historical web info
- don't take archived webpages at face value
- there are multiple ways to timestamp web content

- employ strategies for authentication of IAWM evidence
 - IA affidavit
 - witness
 - judicial notice
- for any strategy, don't neglect to provide foundation for admission





YOU play the most important part in keeping TECHSHOW exciting.

Please complete the Speaker evaluation before you leave.

See You Next Year!

TECHSHOW 2024: February 14-17, 2024