Using the Wayback Machine in Litigation
About Our Webinars

- Lexbe eDiscovery Series webinars take place 1-4x per month and cover a variety of relevant topics in eDiscovery and digital forensics.

- If you experience technical issues or have questions during today’s presentation, please email us at webinars@lexbe.com.

- An archive of Lexbe webinars are available for viewing (streaming video) on YouTube:

- To receive notices to your inbox of upcoming webinars in the Lexbe eDiscovery Webinar Series, please email us at webinars@lexbe.com.

- Follow us on LinkedIn for eDiscovery best practices, links to on-demand webinars, and invitations to upcoming webinars in this series.
About Lexbe Inc.

- Over 8,665 Legal Professionals Served
- Over 17 Years of eDiscovery development and experience
- We help legal professionals, with limited resources, successfully handle document intensive cases
- Lexbe provides legal teams with a highly affordable, full-featured, and lightning-fast DIY proprietary cloud eDiscovery platform
- Delivers the industry’s fastest ROI
- CEDS-certified eDiscovery specialists, expert consultants, and a digital forensics lab

G2 finds that Lexbe “delivers best ROI in the industry and leads in 6 key metrics.”

“Secure, easy-to-use and a great review tool for consideration”

“Cost-effective eDiscovery”

“A powerful litigation document management service”

Using the Wayback Machine in Litigation
Today’s Speaker: Nicholas Taylor

- Expert Witness, Wayback Machine and Temporal Web Forensics

- Experience
  - 16+ years working in research libraries and technology firms
  - 9+ years working in the web archiving field

- Employment
  - Current: Los Alamos National Laboratory
  - Previous: Stanford University, Library of Congress, U.S. Supreme Court, Sun Microsystems, Google

- Education
  - Master of Library Science, University of Maryland
  - Master of Arts in Communication, Culture & Technology, Georgetown University
  - Bachelor of Arts, Media Studies and Psychology, Pomona College
Session Agenda

Using the Wayback Machine in Litigation

Internet Archive: "android data"

Internet Archive: "Open Handset Alliance"
the web exists in space...
...it also exists in time
A Charity Tied to the Supreme Court Offers Donors Access to the Justices

The Supreme Court Historical Society has raised more than $22 million in the last two decades, much of it from lawyers, corporations and special interests. 

While anonymity is topseller, the society has become a vehicle for those seeking access to senior-most federal judges in the nation.

Barbara Walters, a First Among TV Newswomen, Is Dead at 94

The female, a trailblazer in journalism who has interviewed all four presidents of the United States, was once called the "Dowdy Diva" of TV. Walters spent her long career covering the White House, the Oval Office and the Capitol Complex.

Print reporters, broadcast journalists and authors remember Barbara Walters as a trailblazer to journalism.

Watch some of the most memorable moments from Barbara Walters's career.

Trump's Tax Returns Undermine His Image as a Successful Entrepreneur

The release of former President Trump's private tax returns shows how much he relies on foreign entities to hide his true wealth as he continued to rack up in indebted losses.

These Musicians Made an Album. Now It's Nominated for a Grammy.

The debut album of the New York Youth Symphony, featuring some players in middle school, is up against recordings by the world's top orchestras.

What was "White Noise" in Dave Chappelle's 1996 movie, Zoolander, all about life today?

Options

In 2022? A Look at the Year, in Charts.
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 counter-indicative evidence

• material info disseminated via third-party platforms

• some questions (e.g., discoverability) can’t be definitively settled w/ client records alone
overview

• background
• interface + features
• authentication + admissibility
• temporal coherence
• timestamping content changes
IAWM background
Internet Archive

- non-profit digital library
- “universal access to all knowledge”
- provides free public access to collections of all types of digital materials
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 + public is increasingly better archived
• social media less well archived
• non-public web content not archived
IAWM interface + operation
Using the Wayback Machine in Litigation
Using the Wayback Machine in Litigation

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 a successful retrieval, with no errors
  - green means a redirect was 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 [www.supremecourt.gov](http://www.supremecourt.gov) 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 overlay banner for the resource page for the 23 March 2011 capture of www.supremecourt.gov

offers many of the same features as the calendar page: search box, captures range, timeline histogram
• 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)
Using the Wayback Machine in Litigation

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

http://supremecourt.gov/
supreme court

http://supremecourt.gov/
U.S. Supreme Court

http://scotusblog.com/
scotusblog

http://united-states-supreme-court.ukayainline.net/
United States Supreme Court.
full-text search (of home pages): top two results
Using the Wayback Machine in Litigation

full-text search (of home pages): capture counts for top two results
Using the Wayback Machine in Litigation

IAWM calendar pages

supremecourtus.gov

supremecourt.gov

Internet Archive: "Wayback Machine"
IAWM calendar pages: timeline histograms

**supremecourtus.gov**

**supremecourt.gov**
IAWM resource page

• this is the first capture for supremecourtus.gov, from 20 May 2000
Collections Tool + Full-Text Search
Using the Wayback Machine in Litigation

- 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

Using the Wayback Machine in Litigation

Internet Archive: "Wayback Machine"
Collections tool

Using the Wayback Machine in Litigation
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 its own discovery portal w/ full-text search
Archive-It full-text search

Using the Wayback Machine in Litigation
Archive-It full-text search

Explore All Archives

Enter search terms on the right to search the text within the archived pages. Try the more advanced search options below.

Advanced Search

Contains all of:

Exact phrase:

Search within:

From the files:

Advanced Search Help

Explore Archived Content

Archive-It: “Explore Archived Content”

Using the Wayback Machine in Litigation

eDiscovery Webinar Series
Using the Wayback Machine in Litigation

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
Using the Wayback Machine in Litigation

IAWM calendar page

Internet Archive: "Wayback Machine"
Using the Wayback Machine in Litigation

IAWM calendar page

Internet Archive: "Wayback Machine"
Changes tool

"Changes" is a tool you can use to identify and display changes in the content of archives of URLs. First you can select two different archives for a URL, based on an interface that shows the degree of relative change from one archive to another. Then you can see the replay of the two URLs you select, side-by-side, with changes highlighted in Blue and Yellow.
Changes tool: choose 28 September 2021 capture
Using the Wayback Machine in Litigation

Changes tool: degree of difference key

"Changes" is a tool you can use to identify, and display, changes in the context of archives of URLs. First you can select two different archives for a URL, based on an interface that shows the degree of relative change from one archive to another. Then you can see the replay of the two URLs you select, side-by-side, with changes highlighted in Blue and Yellow.

Results: 50 100 500

Explore more than 618 billion web pages saved over time

supremecourt.gov/about/about.aspx

Tue, 28 Sep 2021 02:12:27 GMT

Compare

Distance: Close
Changes tool: degree of difference key
Changes tool: difference from 2017 captures
Changes tool: compare w/ 4 August 2017 capture
Using the Wayback Machine in Litigation

Changes tool: two captures selected

Internet Archive: "Wayback Machine"
Changes tool: two captures selected

Using the Wayback Machine in Litigation
Changes tool: click Compare
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
Changes tool: compare with 2 February 2021 capture
Changes tool: two captures selected
Changes tool: click Compare
Authentication + Admissibility
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.

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.”
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
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.
Using the Wayback Machine in Litigation


- 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."
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.
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.”
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.
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."
Frye + Daubert standards

**Frye standard**
- established in *Frye v. United States*, 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**
- “…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.”
# Admissibility Trends Summary – U.S. Federal Courts

<table>
<thead>
<tr>
<th></th>
<th>Judicial Notice</th>
<th>IA Standard Affidavit</th>
<th>Witness w/ Personal Knowledge</th>
<th>Expert Witness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admitted?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>U.S. District Court</td>
<td>51</td>
<td>13</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>U.S. Circuit Court</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

- 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 recent ruling will likely make judicial notice less successful going forward
Temporal (In)coherence
Using the Wayback Machine in Litigation

Scott G. Ainsworth, et al.：“Only One Out of Five Archived Web Pages Existed as Presented”

Internet Archive: “Weather Underground: Varina, Iowa Forecast”
temporal (in)coherence

Scott G. Ainsworth, et al.: "Only One Out of Five Archived Web Pages Existed as Presented"

Internet Archive: "Weather Underground: Varina, Iowa Forecast"
Using the Wayback Machine in Litigation

Scott G. Ainsworth, et al.: "Only One Out of Five Archived Web Pages Existed as Presented"

Internet Archive: "Weather Underground: Varina, Iowa Forecast"
temporal (in)coherence
Using the Wayback Machine in Litigation

Scott G. Ainsworth, et al.: "Only One Out of Five Archived Web Pages Existed as Presented"

Internet Archive: "Weather Underground: Varina, Iowa Forecast"
temporal (in)coherence

Using the Wayback Machine in Litigation

Scott G. Ainsworth, et al.: "Only One Out of Five Archived Web Pages Existed as Presented"

Internet Archive: "Weather Underground: Varina, Iowa Forecast"
temporal (in)coherence

Webpage was captured on 9 December 2004 at 1:09p CST.

Scott G. Ainsworth, et al.: "Only One Out of Five Archived Web Pages Existed as Presented"

Internet Archive: "Weather Underground: Varina, Iowa Forecast"
temporal (in)coherence
Using the Wayback Machine in Litigation

Webpage was captured on 9 December 2004 at 1:09p CST.

Text on webpage and associated infographic indicate current conditions as raining. Infographic captured on 12 December 2004 at 8:54a CST.

Internet Archive: "Weather Underground: Varina, Iowa Forecast"
Using the Wayback Machine in Litigation


Internet Archive: “Weather Underground: Varina, Iowa Forecast”

Temporal (in)coherence

Webpage was captured on 9 December 2004 at 1:09p CST.

Text on webpage and associated infographic indicate current conditions as raining. Infographic captured on 12 December 2004 at 8:54a CST.

Timestamp in text on webpage corroborates time of capture (i.e., 9 December 2004 at 1:09p CST).

Radar graphic reflects weather conditions (clear) at the time of its capture: 12 September 2005 at 5:34p CDT.

Internet Archive: “Weather Underground: Varina, Iowa Forecast”
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
Timestamping Content Changes
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 before: operator
• Twitter until: 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., Exif, XMP)
timestamping content changes

**when first published**

- easier when there are 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
“Editing” IAWM Contents
can IAWM contents be edited?

- not as such; archive is append-only
- missing content may be captured later
- content can be excluded from access
- IA has made no public claims that content is ever deleted
- individual webpages or web resources can be manually archived
exclusion requests + crawler directives

exclusion requests
• domain owner can request exclusion of their content from IAWM
• criteria unclear, but IA exercises discretion
• if granted, IAWM access to content will be disabled + IA archiving will cease

crawler directives
• domain owner can configure a robots.txt file to ward off IA archiving agents
• again, IA exercises discretion
• updates to robots.txt directives can restore access

- IAWM (would be) used to show expressed or implied licenses for copyrighted software

"[P]laintiffs oppose defendants’ motion for an order compelling them to temporarily disable the robot.txt file so that access to the archived records may be had...Here, plaintiffs argue that they do not have exclusive contractual or statutory rights to control the Internet Archive’s 'archiving activities.'...The issue, however, is not control over 'archiving activities,' but control over access to the archived Netbula web pages. There can be no serious dispute that plaintiffs control that information. Indeed, they have unilaterally blocked access to the same."
IAWM Save Page Now feature

• manually archive to IAWM a real-time webpage snapshot
• good habit when relying on live web evidence
• works for error pages (i.e., document absence)
• when signed in, can archive current webpage + 1 hop out
Takeaways
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
Thank You
Learn More About Lexbe:

Request a personalized platform demo today!

1-800-401-7809 x22 | sales@lexbe.com

‘Lexbe’s cost advantages, SaaS convenience and search capabilities appeal to many small firms’

‘Secure, easy-to-use and a great review tool for consideration’

“A powerful litigation document management service”

‘Cost-effective eDiscovery’

“Because of the Lexbe software, the entire playing field has been leveled for my firm.”

“Lexbe is the easiest eDiscovery software I have ever used”

Using the Wayback Machine in Litigation