

# Understanding Legal Use Cases for Web Archives

Nicholas Taylor  
@nullhandle

IIPC Web Archiving Conference  
16 June 2017

"Libro" by [Paolo Tonon](#) under [CC BY-SA 2.0](#)

# overview

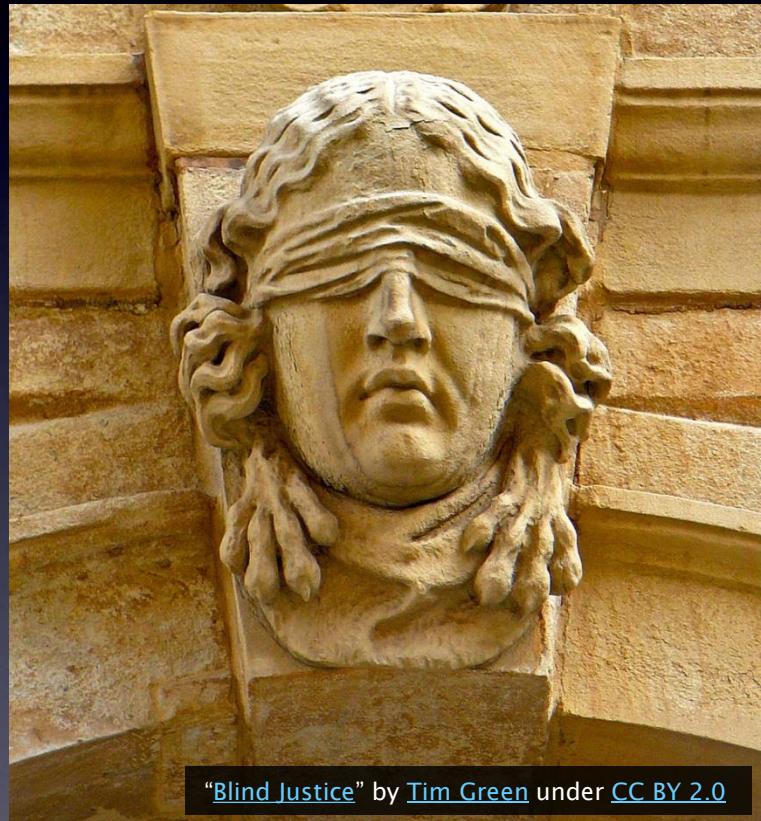
- context
- relevant rules
- example cases
- reliability factors
- what next?



"[LAX on take off](#)" by [Caribb](#) under [CC BY-NC-ND 2.0](#)

# out of scope

- preservation of web citations
- terms of service + API agreements for social media data collecting
- copyright, fair use, + legal deposit
- right to be forgotten



[“Blind Justice” by Tim Green under CC BY 2.0](#)

# CONTEXT

"Pics" by Dani Armengol Garreta under CC BY-NC 2.0

# why web archives?

- where the data is
- organizations slow to incorporate web assets in records management
- prospective records management can't manage what's already lost



# evidence from the web

*“While some look to the Internet as an innovative vehicle for communication, the Court continues to warily and wearily view it largely as **one large catalyst for rumor, innuendo, and misinformation**...Anyone can put anything on the Internet. **No web-site is monitored for accuracy**...Moreover, the Court holds no illusions that **hackers can adulterate the content on any web-site** from any location at any time. For these reasons, any evidence procured off the Internet is **adequate for almost nothing**...Instead of relying on the **voodoo information** taken from the Internet, Plaintiff must hunt for hard copy back-up documentation in admissible form”*

*St. Clair v. Johnny's Oyster & Shrimp, Inc., 76 F. Supp. 2d 773 (S.D. Tex. 1999)*

# trends for IAWM evidence

- cases w/ IAWM evidence as early as 2004
- litigants routinely challenge IAWM evidence
- courts often accept IAWM evidence
  - commonly w/ IA affidavit or testimony
  - increasingly through judicial notice
  - sometimes w/ expert testimony
- evolving understanding of web archive reliability factors

# 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.

# IA standard affidavit

...If a visitor clicks on a link on an archived page, the Wayback Machine will serve the archived file with the closest available date to the page upon which the link appeared and was clicked...

The archived data made viewable and browseable by the Wayback Machine is compiled using software programs known as crawlers that surf the Web and automatically store copies of website files, preserving these files as they exist at the point of time of capture.

...The date assigned by the Internet Archive applies to the HTML file but not to image files linked therein. Thus images that appear on the printed page may not have been archived on the same date as the HTML file...

Attached hereto as Exhibit A are true and accurate copies of printouts of the Internet Archive's records of the HTML files for the URLs and the dates specified in the footer of the printout.

# why should we care?

- legal professionals are **our users**, too
- often, we have **unique historical data**
- help courts + juries **correctly interpret** web archive evidence, leading to **more informed outcomes**
- **broaden community of practice** by bringing in legal technology professionals
- continue to **mainstream web archives**

# why might we hesitate?

- typical cases serve **private interests** rather than public goods
- **immature open source software culture** for legal technology(?)
- **market solutions** for web + social media archiving for litigation already exist

# RELEVANT RULES

# cases

## use cases

- information on individual webpages, at a point in time
- information individual webpages, over a period of time
- persistence of navigational paths, over a period of time

## types of cases

- civil litigation
  - patent infringement
  - trademark infringement
  - copyright infringement
  - class actions
- patent inter partes review (IPR)
- criminal litigation?

# authentication (FRE 901)

## Wikipedia

- “*Authentication, in the law of evidence, is the process by which documentary evidence and other physical evidence is proven to be genuine, and not a forgery.*”

[Wikipedia: Authentication \(law\)](#)

## FRE 901

- Authenticating or Identifying Evidence
  - a) In General. To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is.
  - b) Examples.
    - 1) Testimony of a Witness with Knowledge.
    - 3) Comparison by an Expert Witness...
    - 9) Evidence About a Process or System.

[Legal Information Institute: Rule 901. Authenticating or Identifying Evidence | Federal Rules of Evidence](#)

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

- trademark infringement, breach of contract
- IAWM evidence to show advertisement of partner business service after contract expiration

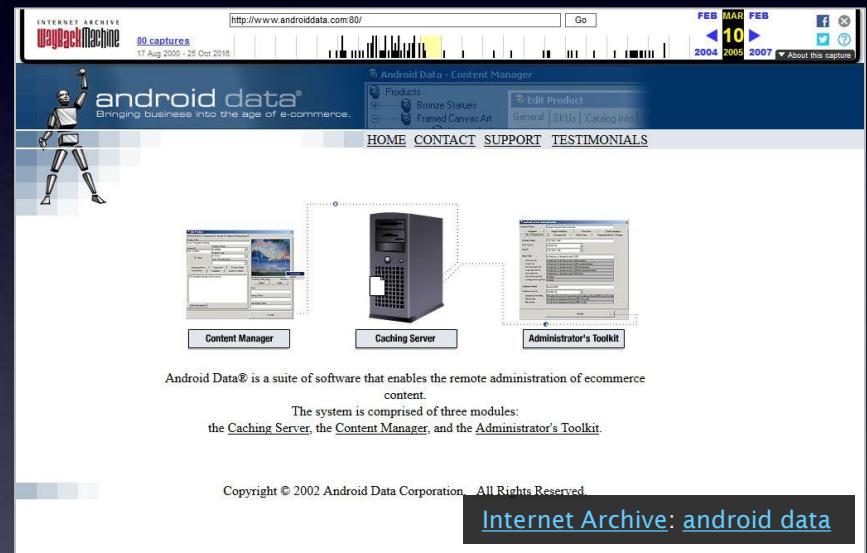
***"Plaintiff has neither denied that the exhibit represents the contents of its website on the dates in question, nor come forward with its own evidence challenging the veracity of the exhibit. Under these circumstances, the Court is of the opinion that Ms. Davis' affidavit is sufficient to satisfy Rule 901's threshold requirement for admissibility. Plaintiff is free to raise its concerns regarding reliability with the jury."***



# *Specht v. Google, Inc. (2010)*

- trademark infringement
- IAWM evidence to show ongoing use of mark over a period of time

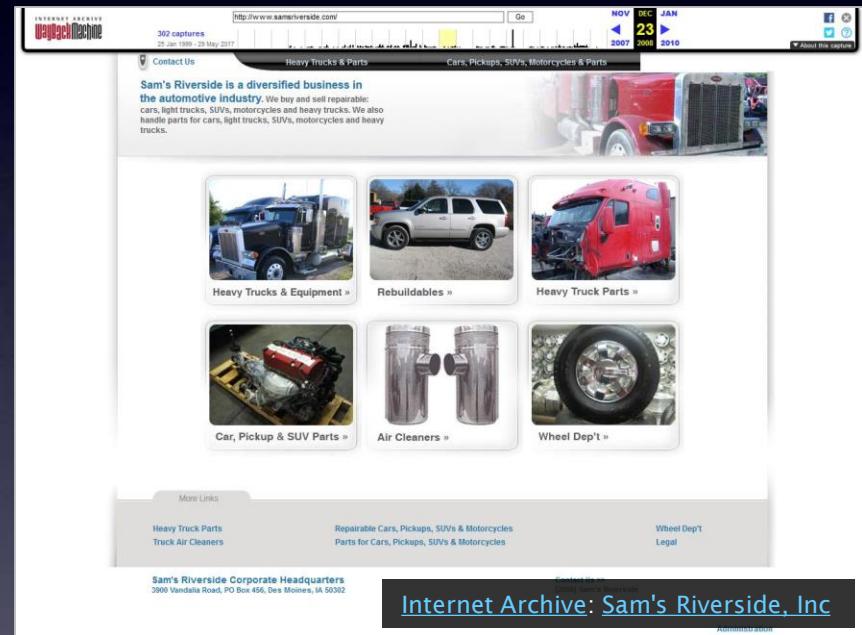
***"[T]he screen shots were not authenticated by an officer or employee of the Internet Archive, but rather through declarations of Specht... This is an improper method to authenticate screen shots from the Internet Archive... Because Plaintiffs did not properly authenticate them, the Court will not consider the Internet Archives printouts from Plaintiffs' Summary Judgment Exhibits"***



# *Sam's Riverside, Inc. v. Intercon Solutions, Inc. (2011)*

- trademark infringement,  
cybersquatting
- IAWM evidence to show ongoing  
use of mark over a period of time

*"Other courts have concluded that an affidavit from an Internet Archive employee is sufficient to authenticate screen shots taken from Archive.org...The Court finds these cases persuasive and, therefore, concludes that the Butler Affidavit is sufficient to authenticate the specific screen shots that are mentioned in and attached to the Butler Affidavit."*



# 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”*

[Wikipedia: Judicial notice](#)

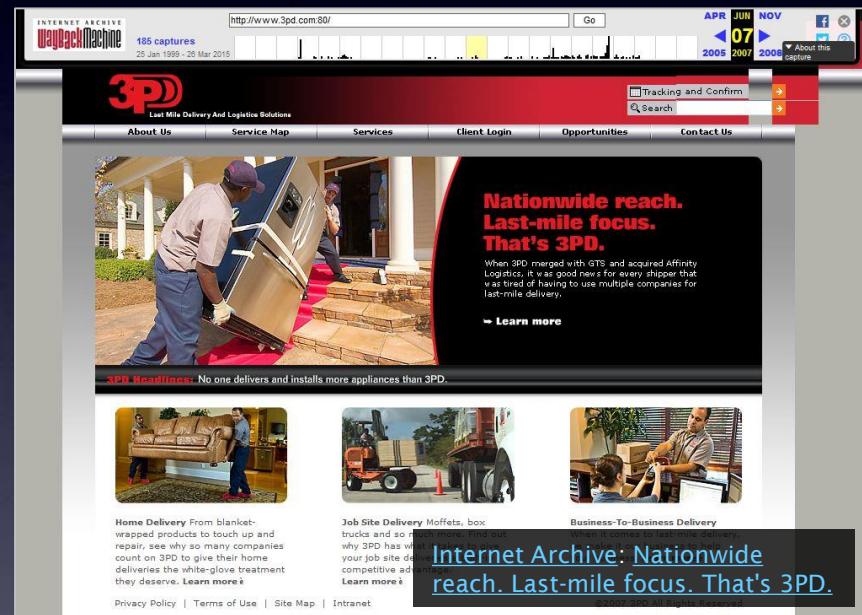
## FRE 201

- 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.

# *Martins v. 3PD, Inc. (2013)*

- employment law, class action
- IAWM evidence to show scope of business services over a period of time

*“In addition to the historical version of the 3PD website attached to Mr. Rabieh's Affidavits dated October 23, 2012, I take judicial notice of the various historical versions of the 3PD website available on the Internet Archive at Archive.org as facts readily determinable by resort to a source whose accuracy cannot reasonably be questioned.”*

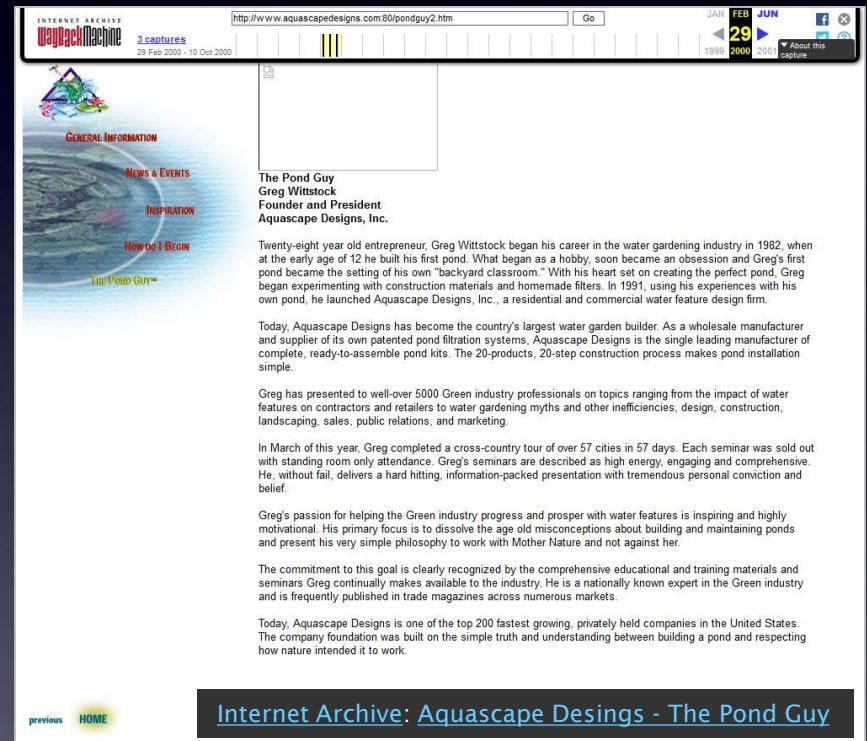


*Martins v. 3PD, Inc., No. 11 Civ. 11313, 2013 U.S. Dist. LEXIS 45753, 2013 WL 1320454, at \*16 n.8 (D. Mass. Mar. 28, 2013)*

# Pond Guy, Inc. v. Aguascape Designs, Inc. (2014)

- trademark infringement
- IAWM evidence to show earliest use of mark on a website

*“Neither party...has submitted sufficient evidence to the Court detailing exactly what their historical internet presence consisted of. Seeing as this is a key issue in dispute, the Court turns to the Internet Archive, found at <https://archive.org/web/>, for the taking of judicial notice. As a resource the accuracy of which cannot reasonably be questioned, the Internet Archive has been found to be an acceptable source for the taking of judicial notice.”*



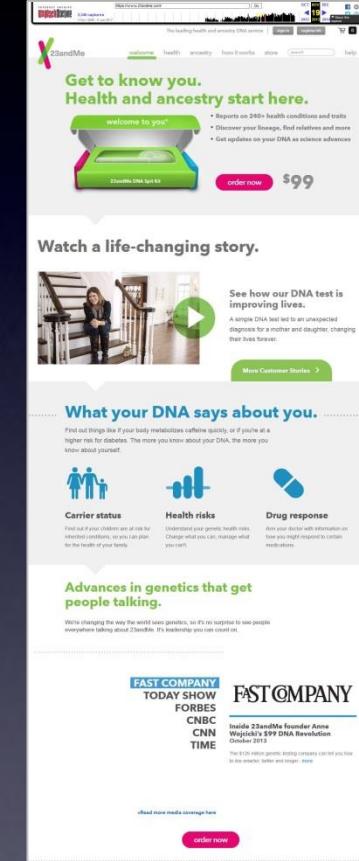
Internet Archive: [Aguascape Desings - The Pond Guy](http://www.aguascapedesigns.com:80/pondguy2.htm)

# *Tompkins v. 23andMe, Inc. (2014)*

- false advertising, consumer protection, class action
- IAWM evidence to show terms of service at point in time

*"The parties do not dispute that the key portions of the website have not changed since the relevant times when Plaintiffs allegedly performed the transactions at issue. 23andMe relies on excerpts from a February 2014 version of the website...while Plaintiffs use excerpts dated April 2014... However, the Court takes judicial notice of the Internet Archive (<http://archive.org>) version of 23andMe's website as of November 20, 2013, the full version of the website archived right before the FDA warning letter of November 22, 2013"*

*Tompkins v. 23andMe, Inc., 2014 U.S. Dist. LEXIS 88068 (N.D. Cal. June 25, 2014)*



[Internet Archive: 23AndMe - Genetic Testing for Health, Disease & Ancestry; DNA Test](#)

# expert witnesses (FRE 702)

## Wikipedia

- “*An expert witness, in England, Wales and the United States, is a person whose opinion by virtue of education, training, certification, skills or experience, is accepted by the judge as an expert.*”

[Wikipedia: Expert witness](#)

## FRE 702

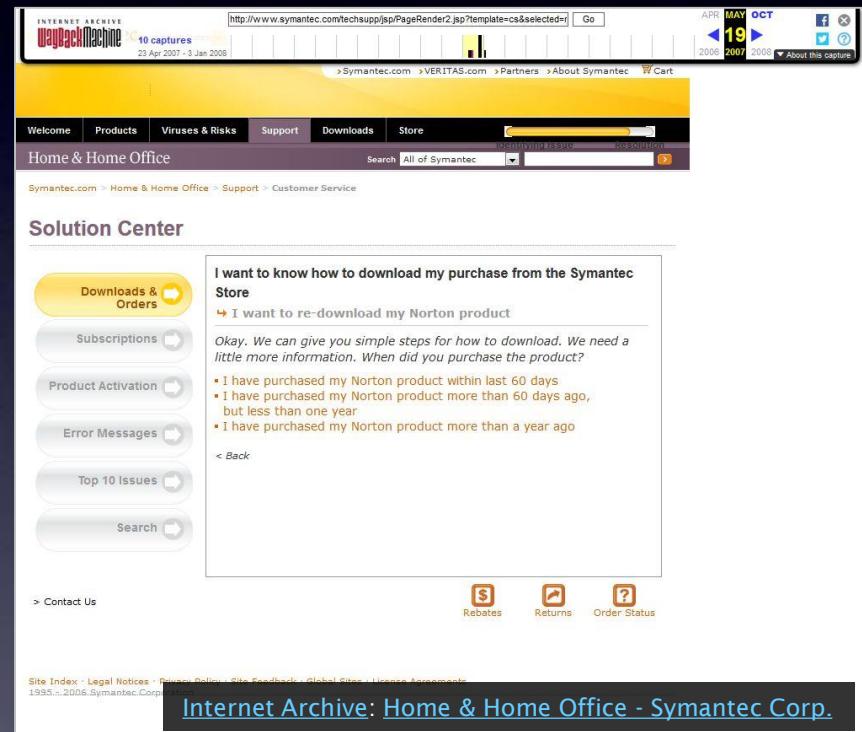
- Testimony by Expert Witnesses  
A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
  - a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand 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](#)

# *Khoday v. Symantec Corp., et al (2015)*

- consumer protection, class action
- IAWM evidence to show probable navigation paths over period of time

***"Expert testimony that merely repeats information capable of easy comprehension by a jury is excludable...The Court finds that Taylor offers conclusions not readily available to the jury without the assistance of expert testimony. Whether legitimate alternatives for redownloading Norton products were available, at no cost, to customers during the relevant time period is central to the Plaintiffs' claims, making Taylor's proffered testimony highly relevant."***





# RELIABILITY FACTORS

["Contraption - Explored"](#) by [Clint Budd](#) under [CC BY 2.0](#)

# reliability factors in cases

## raised as concerns

- IAWM disclaimer
- incompleteness
- provenance
- temporal coherence  
(discrepant embedded memento datetimes)

## no examples(?) so far

- discreteness
- temporal coherence  
(considering also original HTTP headers)
- canonicality
- client artifacts

# Nassar v. Nassar (2017)

- defamation
- IAWM evidence to show presence of defamatory information at point in time

*"[T]he organization that maintains waybackmachine.org itself disclaims any guarantee that the results it produces are accurate. See Internet Archive Wayback Machine, Internet Archive's Terms of Use, Privacy Policy, and Copyright Policy..." ("You understand and agree that the Archive makes no warranty or representation regarding the accuracy, currency, completeness, reliability, or usefulness of the content in the Collections."). In light of that, and because Plaintiff otherwise has made no showing that the accuracy of the Internet archive cannot reasonably be questioned, judicial notice of the historical contents of the websites is inappropriate."*

[Nassar v. Nassar, 2017 U.S. Dist. LEXIS 456 \(M.D. Fla. Jan. 3, 2017\)](#)



The screenshot shows a web browser window with the URL <http://www.rosananassar.com/00r>. The page title is "RosanaNassar". The content discusses a real history behind abuse, mentioning a family named Nassar. It includes a navigation menu with links to Home, Problemas com familiares, and Problemas Profissionais. A sidebar on the right shows a timeline from May to July 2010, with a yellow box highlighting the date "05 JUN 2010". Below the main content, there is a section titled "Quem é realmente rosana nassar ?" and a note about the book "O poder da libertação pela voz". The footer contains a link to "New Global Publishing".

**RosanaNassar**

A real história por trás do abuso do poder

- Home
- Problemas com familiares
- Problemas Profissionais

**Quem é realmente rosana nassar ?**

Por trás de um jeito polido, se esconde uma mente doente

Conheço a família Nassar já há vários anos. É uma família virtuosa, bonita, unida e decente... Qual não foi o meu espanto quando soube de uma publicação que teria sido feita por [Rosana Nassar](#), uma pessoa que teve todas as possibilidades de estudar, passar por um processo de graduação, teve todo auxílio dos familiares em seu processo de crescimento pessoal, mas que em dado momento de sua vida, optou por destruir suas origens, optou por difamar uma história tão bonita como é a dessa família tão bem vista em nossa cidade! Fico pensando o que aconteceu para que essa mulher resolvesse **sem nenhum fundamento** real caluniar de forma tão desacurada pessoas que outrora foram vitais para ela. Uma pessoa como essa, ao meu ver, deveria no mínimo passar por um sério **tratamento psicológico** para expurgar esses sentimentos nocivos.

Li alguns trechos do livro e qualquer pessoa em sã consciência vê que aquilo não se trata de um relato real, mas sim de uma fantasia criada por um ser afetado por um histerismo crônico, fato pelo qual foi banido das lojas em Belém! E ai ponto para as pessoas de bom senso! Para se ter uma idéia daonde uma mente psicótica pode chegar, Rosana Nassar criou uma página web de uma editora para realizar a venda desse livro sobre o pretexto de assegurar publicações baratas para profissionais que estavam começando. Realizei várias buscas na internet e encontrei várias informações evasivas sobre essa "editora" criada por ela, inclusive com vários nomes diferentes sendo um deles "[NEW GLOBAL PUBLISHING](#)".

[Rosana Zolin](#) é uma pessoa que tem **graves problemas psicológicos**, uma pessoa extremamente agressiva, com um sério **desvio de personalidade**, que não fala há vários anos com a mãe e os irmãos. Teve sérios **problemas e insucessos profissionais** em vários lugares que trabalhou devido sua agressividade e comportamento psicótico. Teve sérios problemas emocionais e fez consultas psiquiátricas em Belém. Em seu histórico psiquiátrico familiar teve o pai e tias acometidos por problemas psiquiátricos, o que pode auxiliar no entendimento de porque ela está cometendo tais atos, porém em nenhum momento justificou os mesmos. Pois como psicóloga que se diz (e não sou sabedor se isso é real ou mais uma invenção de sua mente conturbada) deveria saber que o caminho real de sua libertação seria pela ajuda profissional e não por um **legado malitio** que está criando para si e para seus familiares.

Caros leitores, estou colocando aqui o meu testemunho do que a família de [Rosana Zolin](#) está passando, como amigo e conhecedor do problema, muitos familiares da receberam e-mails escabrosos nas últimas semanas, fizemos o revolto que resolvi coletar dados dos familiares e deixá-los públicos como forma de tentar criar um local que desse uma resposta definitiva à essa faraônica!

- **Rosana Nassar Zolin**
- **Informações**
  - O poder da libertação pela voz
  - Símula Crítica acerca do livro
  - Quem é Rosana Nassar
  - Problemas com o irmão mais velho
  - Problemas com a mãe
  - Problemas com a terceira irmã
  - Problemas com a irmã caçula
  - Problemas profissionais
- **Links**
  - [New Global Publishing](#)

rosananassar.com é um site produzido com [Internet Archive: Rosana Nassar Zolin](#)

# *Stabile v. Paul Smith Ltd. (2015)*

- copyright infringement
- IAWM evidence to show absence of particular image over period of time

*"Stabile also objects on the ground that the Internet pages shown are not the full captures of the website, but only isolated pages. (Id.) The captures from the Wayback Machine will not be considered conclusive evidence of what was or was not on Stabile's entire website between 2001 and 2011, but will be admitted as support for the contention that there is no evidence beyond Stabile's deposition statement that Extrapolations # 3 was displayed on her website at all, and that there is no evidence that it was displayed during the relevant time period."*

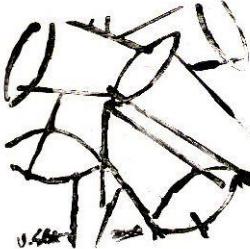
*Stabile v. Paul Smith Ltd., 2015 U.S. Dist.  
LEXIS 101291 (C.D. Cal. July 31, 2015)*

INTERNET ARCHIVE  
**Wayback Machine**  
5 captures  
14 Dec 2001 - 2 Aug 2002

<http://www.judystabile.com:80/bio.htm> | Go

## JUDY STABILE

BRIEF BIO | GALLERY | INSTALLATIONS | LINKS | CONTACT US



**Judy Stabile** is best known for her unconventional use of materials in her approach to painting. Using multiple layers of metal leaf, curved glass and a serial image, she creates a geometric dialogue and integrates it with a human emotional side.

Born in 1948. Received her education at Chouinard Institute of Art, Los Angeles, California 1966-1970. Judy Stabile presently lives and works at her studio in Venice, California.

**BORN:**  
1948 LOS ANGELES, CALIF.

**EDUCATION:**  
CHOUINARD INSTITUTE OF ART 1966-70

**EXHIBITIONS:**  
MOLLY BARNES LOS ANGELES, CA; WORKS GALLERY COSTA MESA, CA; SKIDMORE CONTEMPORARY ART LOS ANGELES, CA; TRAN AMERICA BUILDING SAN FRANCISCO, CA; DIANE NELSON FINE ART LAGUNA, CA

**COLLECTIONS:**  
ARCO CORP. DENVER, CO; CEDARS-SINAI MEDICAL CENTER LOS ANGELES, CA; XEROX DALLAS, TEXAS; SPAGO, BEVERLY HILLS, CA; VERIZON HEADQUARTERS THOUSAND OAKS, CA; PAYDEN AND RYGEL, LOS ANGELES, CA; LUCKMAN FINE ART LOS ANGELES, CA; HUGHES AIRCRAFT LOS ANGELES, CA

**COMMISSIONS:**  
LOS ANGELES COUNTY MUSEUM OF ART/ART COUNCIL-LIMITED EDITION; HEWLETT PACKARD LONDON; ROYAL CARIBBEAN CRUISE LINE, MIAMI & NORWAY

[Click here for Exhibitions / Commissions](#)

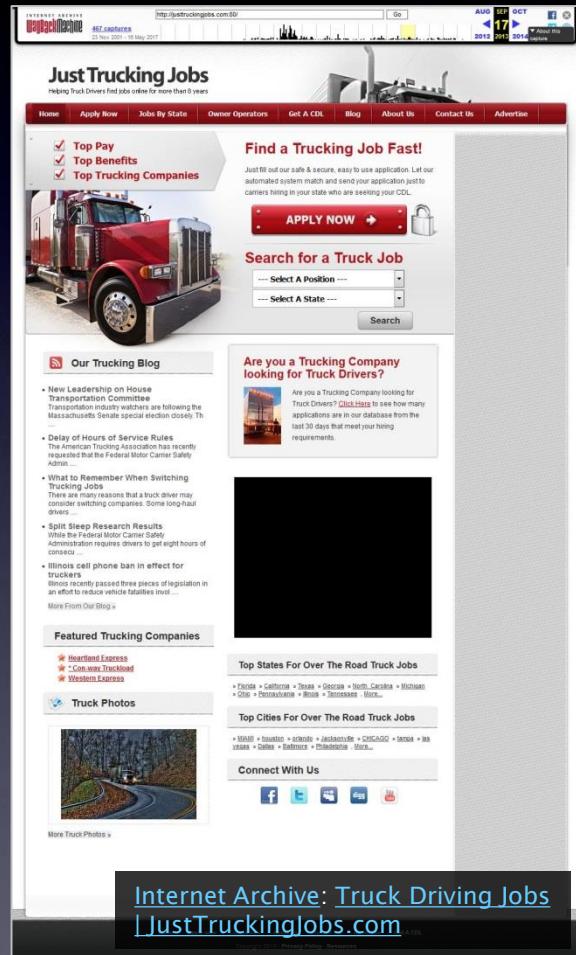
BRIEF BIO | GALLERY | INSTALLATIONS | LINKS | CONTACT US  
**Internet Archive: Judy Stabile**  
designed by [craigslaw.com](#)

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

- trademark infringement
- IAWM evidence to show ongoing use of mark over period of time

*"[T]he fact that the Wayback Machine doesn't capture everything that was on those sites does not bear on whether the things that were captured were in fact on those sites. There is no suggestion or evidence from PlattForm that the Wayback Machine ever adds material to sites (other than a Wayback Machine toolbar and coding that allows links to work)."*

[Marten Transport, Ltd v PlattForm Advertising, Inc.](#)  
2016 U.S. Dist. LEXIS 57471 (D. Kan. Apr. 29, 2016)



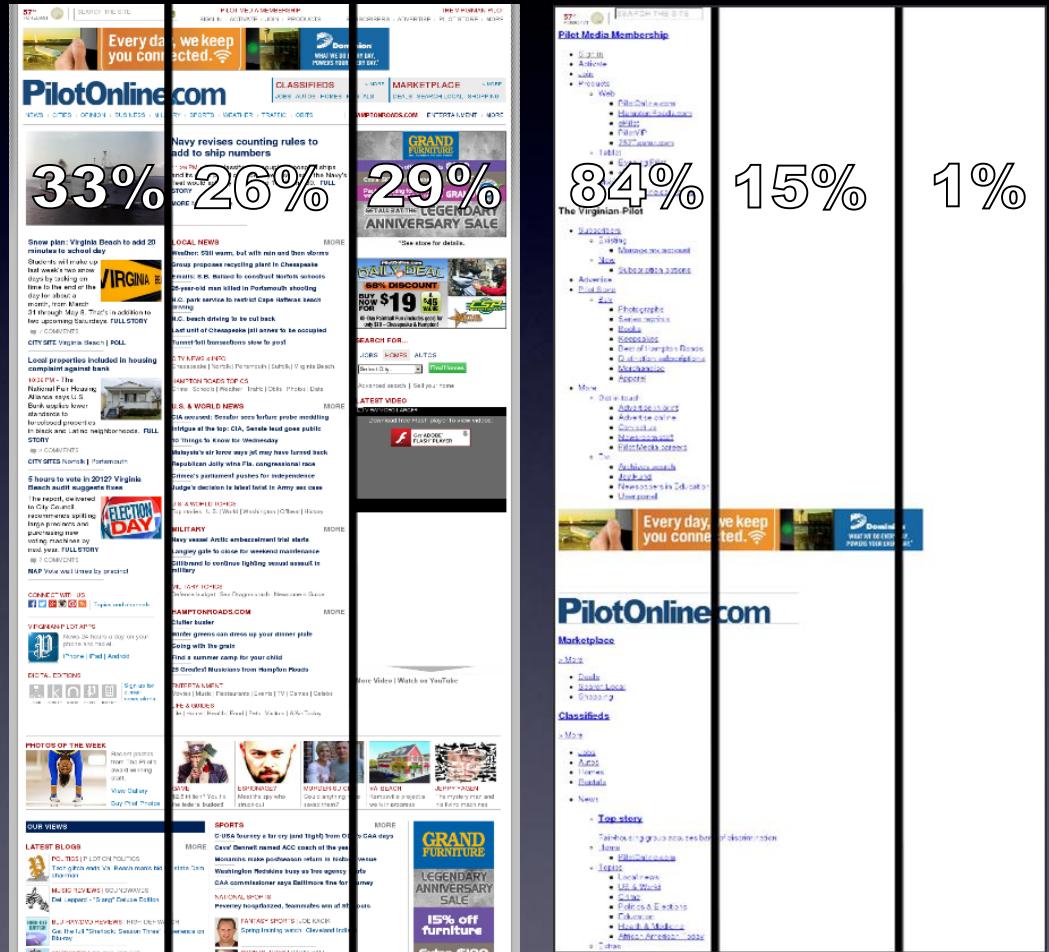
# *Marten Transport, Ltd. v. PlattForm Advertising, Inc. (2016)*

- objections to evidence admissibility:
  - any of 3 crawlers may have retrieved page
  - images sometimes missing
  - crawlers can't capture some content
  - IA doesn't guarantee accuracy
  - affiant has not prepared a crawl

*Marten Transport, Ltd v PlattForm Advertising, Inc.*  
2016 U.S. Dist. LEXIS 57471 (D. Kan. Apr. 29, 2016)

# quantifying incompleteness

- <img> + <embed> leave hints about size + centrality
- for CSS, look at distribution of background color
- missing resources connote unreliability
- materiality of missing resources subject to interpretation



Justin F. Brunelle et al: "Not all mementos are created equal: measuring the impact of missing resources"

# live site “leakage” via AJAX

INTERNET ARCHIVE  
**Wayback Machine** BETA

http://www.cnn.com Go JUL SEP OCT Close X  
15,098 captures 3 Help ?  
15 Aug 00 - 29 Jul 11

**CNN.com** Web | CNN News CNN Videos SEARCH POWERED BY Google

HOME WORLD U.S. POLITICS CRIME ENTERTAINMENT HEALTH TECH TRAVEL LIVING BUSINESS SPORTS TIME.COM VIDEO I REPORT IMPACT

Hot Topics » Hurricane Gustav » Conventions » Sarah Palin » more topics » Edition: International | Set Pref

updated 4:24 p.m. EDT, Wed September 3, 2008 Make CNN Your Home Page

**Republican National Convention 2008**



**Palin's path from city hall to governor's mansion**

When she played basketball in high school, Sarah Palin, the soon-to-be Republican vice presidential nominee, earned the nickname "Sarah Barracuda" for her fierce competitiveness. In the 21 months she has served as governor of Alaska, no one is suggesting she's lost her fighting edge. [full story](#)

- Palin ready to lead? [Analysts: Her task](#)
- Palin to call for reform in RNC speech
- Will her gender sway women to Palin?

**Latest News**

- Poll measures race in three key states [19 min](#)
- Dems blast Lieberman, say he lied to delegates
- Ticker: McCain greets Palin's daughter, fiancé
- Martin: Keep politics out of Bristol Palin issue
- Biographer: America, meet Sarah Palin
- iReport.com: Can a mom of 5 be VP?
- Cafferty: Should McCain consider different VP?
- CNNMoney: Auto sales plunge
- Gunman's bloody trail leads to six bodies
- Caylee's grandma disputes evidence [LIVE](#)
- Nagin: Residents won't be turned away
- CNNMoney: Gulf oil production scrambling back
- Hanna roams Caribbean, may hit U.S.
- Pakistan protests 'coalition air attack'
- The cartoonist beloved by GIs and regular guys
- 8 dogs die in hot truck during lunch break [LIVE](#)
- Review: Google's Chrome needs more polish
- New Marilyn Monroe film scenes found [LIVE](#)
- Beyoncé's sister aims for a path of her own
- CNN Wire: Paraguay coup plot alleged

[all news from the past 24hrs »](#) [all most popular »](#)

**Republican National Convention »**

- Bernstein: Democrats better take note
- Bush, McCain still an uneasy alliance
- Interactive: How conventions work
- McCain arrives in Minnesota for convention [LIVE](#)
- Key players | Trivia | The Forum | In Depth

**Viewer's Guide**

**Video »**



Gustav recovery briefing [LIVE](#)



Dogs die during lunch break [1:39](#)



The RNC in St. Paul [LIVE](#)

LIVE: Watch coverage of the RNC & Tropical Storm Hanna

**\* FIRST \* PRESIDENTIAL DEBATE OBAMA | ROMNEY**



ADVERTISEMENT

**CNN TV »**

Justin F. Brunelle: ["Zombies in the Archives"](#)  
Secrets to a healthy heart

Internet Archive: CNN.com

# live site “leakage” via AJAX

INTERNET ARCHIVE  
**Wayback Machine** BETA

http://www.cnn.com/ Go JUL SEP OCT  
15,098 captures 3 2007 2008 2009  
15 Aug 00 - 29 Jul 11

CNN.com

Web | CNN News CNN Videos SEARCH POWERED BY Google

HOME WORLD U.S. POLITICS CRIME ENTERTAINMENT HEALTH TECH TRAVEL LIVING BUSINESS SPORTS TIME.COM

VIDEO I REPORT IMPACT

Edition: International | Set Pref

Hot Topics » Hurricane Gustav » Conventions » Sarah Palin » more topics »

updated 4:24 p.m. EDT, Wed September 3, 2008 Make CNN Your Home Page

**Republican National Convention 2008**

**Palin's path from city hall to governor's mansion**

When she played basketball in high school, Sarah Palin, the soon-to-be Republican vice presidential nominee, earned the nickname "Sarah Barracuda" for her fierce competitiveness. In the 21 months she has served as governor of Alaska, no one is suggesting she's lost her fighting edge. full story

- Palin ready to lead? Analysts: Her task
- Palin to call for reform in RNC speech
- Will her gender sway women to Palin?

**Latest News**

- Poll measures race in three key states 19 min
- Dems blast Lieberman, say he lied to delegates
- Ticker: McCain greets Palin's daughter, fiancé
- Martin: Keep politics out of Bristol Palin issue
- Biographer: America, meet Sarah Palin
- iReport.com: Can a mom of 5 be VP?
- Cafferty: Should McCain consider different VP?
- CNNMoney: Auto sales plunge
- Gunman's bloody trail leads to six bodies
- Caylee's grandma disputes evidence
- Nagin: Residents won't be turned away
- CNNMoney: Gulf oil production scrambling back
- Hanna roams Caribbean, may hit U.S.
- Pakistan protests 'coalition air attack'
- The cartoonist beloved by GIs and regular guys
- 8 dogs die in hot truck during lunch break
- Review: Google's Chrome needs more polish
- New Marilyn Monroe film scenes found
- Beyoncé's sister aims for a path of her own
- CNN Wire: Paraguay coup plot alleged

all news from the past 24hrs » | all most popular »

**Republican National Convention »**

- Bernstein: Democrats better take note
- Bush, McCain still an uneasy alliance
- Interactive: How conventions work
- McCain arrives in Minnesota for convention
- Key players | Trivia | The Forum | In Depth

Viewers' Guide

**Video »**

Gustav recovery briefing LIVE Dogs die during lunch break 1:39 The RNC in St. Paul LIVE

LIVE: Watch coverage of the RNC & Tropical Storm Hanna

**\* FIRST \* PRESIDENTIAL DEBATE OBAMA | ROMNEY**

ADVERTISEMENT

**CNN TV »**

Internet Archive: CNN.com

# temporal coherence

INTERNET ARCHIVE WayBack Machine https://web.archive.org/web/20041209190926/http://www.wunderground.com/cgi-bin/ Go Google AUG DEC MAR Close X 2003 9 2004 2006 Help ?

wunderground.com PDA/Mobile

Member Benefits: No Ads Weather Email \$5 a year Signup Here

Email Password Login Print This Page Maps Temperature Heat Index

Find the Weather for any City, State or ZIP Code, or Airport Code or Country

Link to Wunder Photos

Features: Wunder Photos NEXRAD radar Regional Radar NEW Zoom Satellite Maps Trip Planner Tropical / Hurricane Marine Severe Astronomy Ski Education Favorites Personal Weather Stations

**Varina, Iowa**  Add this page to your Favorites  Weather Email Local Time: 1:09 PM CST Set My Timezone Lat/Lon: 42.6° N 94.8° W | MSN Map

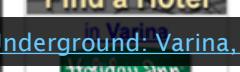
Current Conditions Updated: 12:55 PM CST on December 09, 2004 Observed at Storm Lake, Iowa ([History](#)) Elevation: 1486 ft / 453 m

 **41 °F / 5 °C** Light Drizzle Windchill: 37 °F / 3 °C Humidity: 100% Dew Point: 41 °F / 5 °C Wind: 6 mph / 9 km/h from 

5-Day Forecast for ZIP Code 50593

Thu	Fri	Sat	Sun	Mon
 43°   29°	 36°   22°	 47°   34°	 41°   15°	 31°   15°
Chance of Rain	Mostly Cloudy	Mostly Cloudy	Partly Cloudy	Partly Cloudy
<a href="#">Detail</a>	<a href="#">Detail</a>	<a href="#">Detail</a>	<a href="#">Detail</a>	<a href="#">Detail</a>

Click [Detail](#) for hourly wind, temperature, humidity and UV forecasts. Alternate Computer Forecast: [AVN MOS Weather Graph](#) | [Local Allergy Info from Pollen.com](#)

 [Local Radar](#)  [Regional Radar](#)  [Find a Hotel](#)  [Wunder Photos](#)

Scott G. Ainsworth et al: "Only One Out of Five Archived Web Pages Existed as Presented" 56 in / 1001 hPa

Internet Archive: Weather Underground: Varina, Iowa Forecast

# temporal coherence

https://web.archive.org/web/20041209190926/http://www.wunderground.com/cgi-bin/

INTERNET ARCHIVE Wayback Machine http://www.wunderground.com/cgi-bin/findweather/getForecast?query= Go Google

21 captures 27 Oct 02 - 25 Jan 11

Find the weather for any City, State or ZIP Code, or Airport Code or Country Link to Wunder Photos

wunderground.com PDA/Mobile missing

-15 hours

Member Benefits: No Ads Weather Email \$5 a year Signup Here

Email Password Login Print This Page Maps Temperature Heat Index

Varina, Iowa

Local Time: 1:09 PM CST Set My Timezone

Add this page to your Favorites Weather Email Lat/Lon: 42.6° N 94.8° W | MSN Map

Current Conditions

Updated: 12:55 PM CST on December 09, 2004  
Observed at Storm Lake, Iowa (History)  
Elevation: 1486 ft / 453 m

41 °F / 5 °C Light Drizzle Windchill: 37 °F / 3 °C Humidity: 100% Dew Point: 41 °F / 5 °C Wind: 6 mph / 9 km/h from

5-Day Forecast for ZIP Code 50593

Thu	Fri	Sat	Sun	Mon
43°   29°	36°   2	15°	31°   15°	
Chance of Rain	Mostly Cl			
<a href="#">Detail</a>				

+9 hours

Click Detail for hourly wind, temperature, humidity and UV forecasts.

Alternate Computer Forecast: AVN MOS Weather Graph | Local Allergy Info from Pollen.com

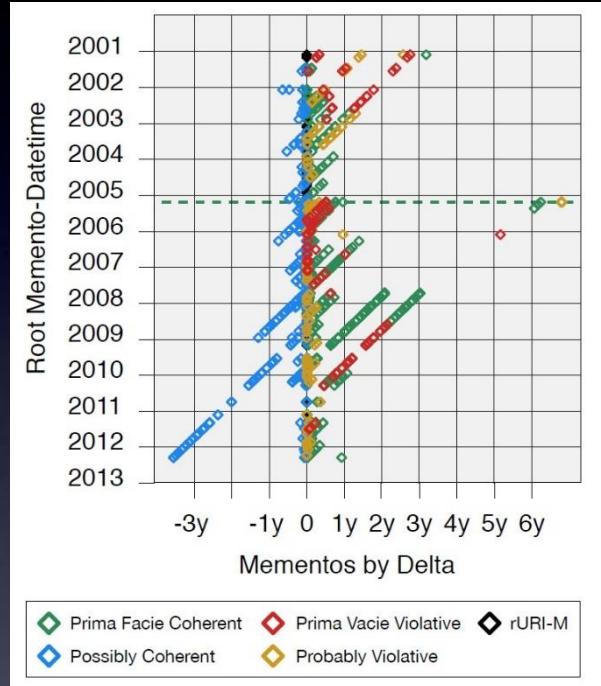
Scott G. Ainsworth et al: "Only One Out of Five Archived Web Pages Existed as Presented"

+9 months

Internet Archive: Weather Underground: Varina, Iowa Forecast

# visualizing temporal coherence

The screenshot shows the Wayback Machine's "Time Travel" interface. At the top, there is a red suitcase icon with a "1" on it, followed by the text "Time Travel". Below this is a dropdown menu. To the right are links for "About API Privacy Terms". A search bar contains the URL "http://stanford.edu". Below the search bar are two date inputs: "2010-04-28" and "10:34:32". Next to them are "Find" and "Reconstruct" buttons, and a "Request" button. To the right of these is a timestamp "2010-04-27 16:17:59 GMT" with a play button icon. Below these controls is a horizontal timeline with markers for "4 months before", "12 days before", "a day before", "3 hours before", "17 minutes before", "2 minutes before", "9 seconds before", and the current timestamp "2010-04-27 10:34:32". A vertical blue bar highlights the "a day before" marker. To the right of the timeline is a legend: "wayback.archive-it.org (39)" (blue square), "web.archive.org (2)" (light blue square), and "wayback.vefsafn.is (3)" (red square). Below the timeline, a message states: "The page below assembled using 44 Mementos from 3 archives, spanning a year". The main content area displays the Stanford University homepage from April 2010. The header features the Stanford University logo and navigation links for "Maps", "A-Z Index", and "Search...". The main content includes sections for "About Stanford", "Admission", "Academics", "Research", and "Life On Campus". Below this are sections for "EVENTS" (listing events like "Panel: SEC v. Goldman Sachs 5 p.m.", "Hal Holbrook: Mark Twain Tonight! 8 p.m.", and "Wellness Fair 10 a.m."), "UNIVERSITY NEWS" (articles about the "Admit Weekend 2010", "Campus rolls out the Cardinal carpet for more than 1,300 prospective freshmen and their families.", "Ash and aircraft", "Danger to airliners from volcano's debris is difficult to assess, Stanford expert says.", and "Researcher: Indoor air may be hazardous to women's health"), and "GATEWAYS FOR..." (links to "Students", "Faculty & Staff", "Alumni", "Parents", and "Visitors & Neighbors"). A sidebar on the right lists "TOP DESTINATIONS" under the heading "SCHOOLS", showing images of various university buildings.

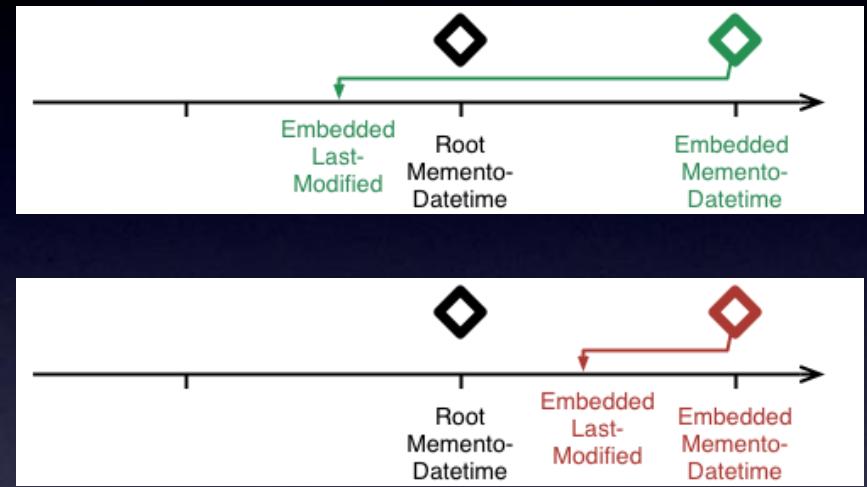


[Scott Ainsworth et al: "Only One Out of Five Archived Web Pages Existed as Presented"](#)

again, the materiality of temporal incoherence is subject to interpretation

# other datetime estimation strategies

- Carbon Dating (+ constituent services)
- comparing X-Archive-Orig-last-modified w/ Memento datetime
- contemporaneous inlinks via IAWM beta search



Scott G. Ainsworth: “Evaluating the Temporal Coherence of Composite Mementos”

# interpreting datetimes

- static datetime in text of archived webpage
- Memento datetime
- X-Archive-Orig-last-modified
- datetimes of root vs. embedded mementos
- datetimes of contiguously-navigable mementos
- effect of crawl deduplication on memento datetimes



[“Dali Time” by PhotoAteller under CC BY 2.0](#)

# *Servicenow, Inc. v. Hewlett-Packard Co. (2015)*

- inter pares review
- IAWM evidence to show presence of prior art

*“Initially, we note that, even relying on Exhibit A to the Butler Affidavit, the webpage for the Introducing Collaborate Reference was archived on November 1, 2002, less than one year prior to the May 14, 2003 filing date of the ’981 patent and, thus, fails to qualify as a prior-art, printed publication under 35 U.S.C. § 102(b)”*

[Servicenow, Inc. v. Hewlett-Packard Co., Case IPR2015-00707, slip op. at 5-18 \(PTAB August 26, 2015\)](#)

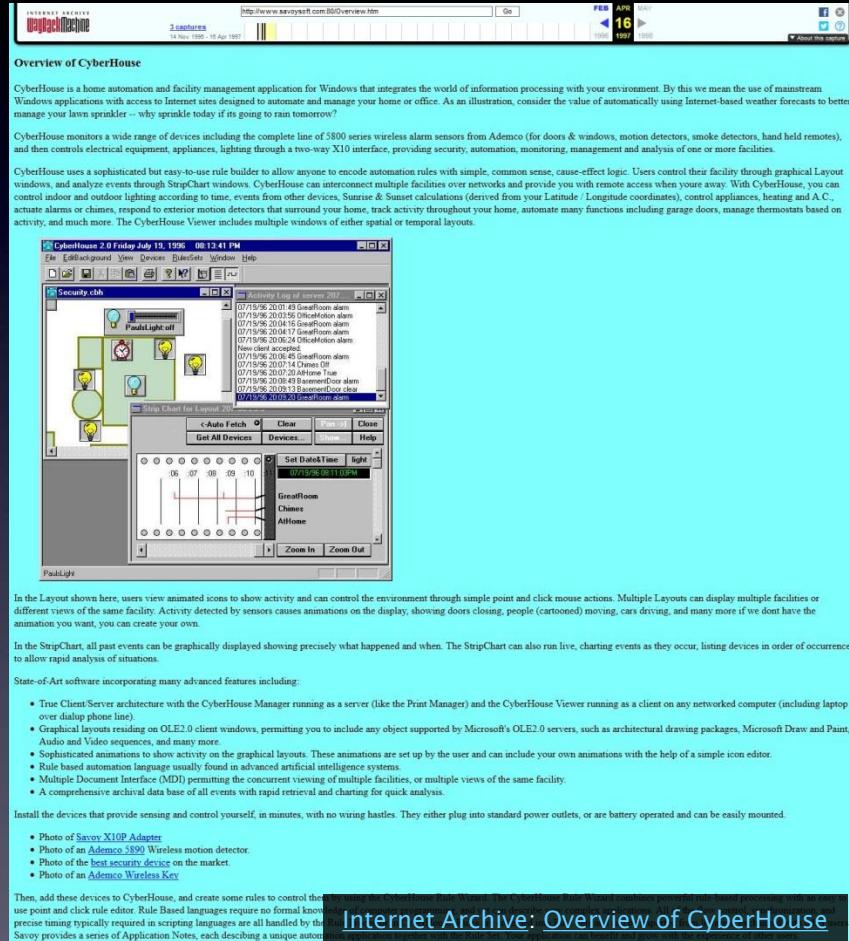
The screenshot shows a Wayback Machine capture of the BEA Collaborate Release 2.0 website. The page has a yellow header bar with the BEA Logo and the title 'BEA Collaborate Release 2.0'. Below the header is a navigation menu with links to 'BEA Home', 'Events', 'Solutions', 'Partners', 'Products', 'Services', 'Download', 'Developer Center', and 'WebSUPPORT'. The main content area features a large orange banner with the text 'Welcome to Introducing BEA WebLogic Collaborate! This document:' followed by a bulleted list of features. The list includes: 'Provides an overview of WebLogic Collaborate architecture and features.', 'Describes how WebLogic Collaborate meets your business-to-business integration needs.', and 'Uses an end-to-end example to describe the tasks necessary to start a business-to-business transaction among business partners.' At the bottom of the page, there are sections for 'Overview' and 'Getting Started Using Weblogic Collaborate', each with a list of links to various documentation pages.

# *Crestron Electronics, Inc. v. Intuitive Building Controls, Inc. (2016)*

- inter pares review
- IAWM evidence to show prior art

*“Although Patent Owner is correct that the Butler Affidavit asserts that embedded images may not be archived on the same date as the web page that embeds the image, Mr. Butler does not state that such embedded images cannot be archived on the same date. Thus, we have no evidence on this record that the specific web pages produced in Exhibit 1002 do not in fact include the embedded images, as asserted by Patent Owner.”*

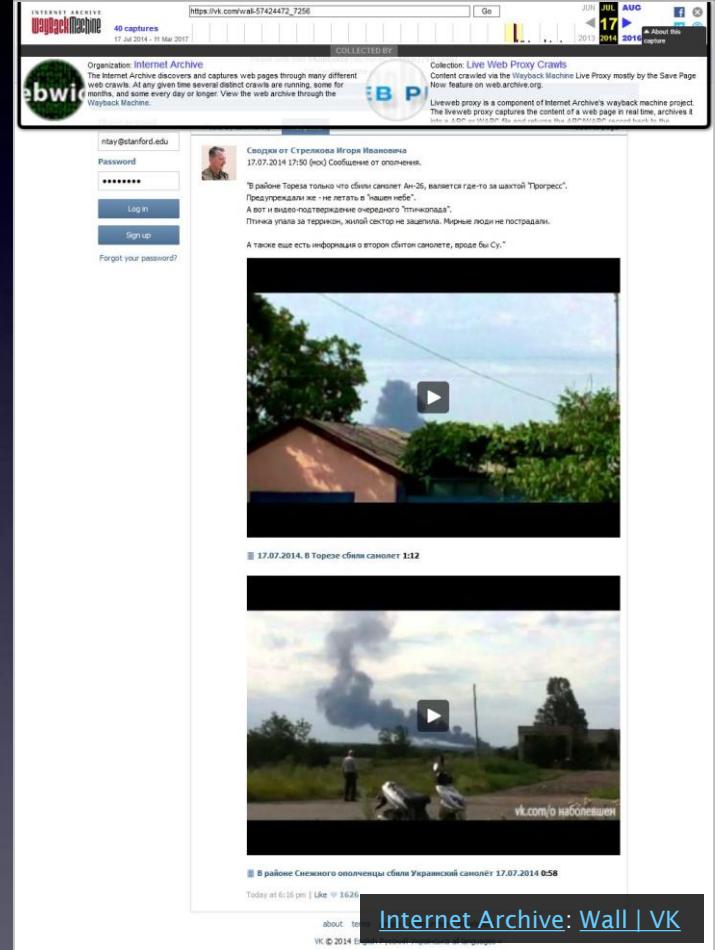
*Crestron Electronics, Inc. v. Intuitive Building Controls, Inc.,  
Case IPR2015-01460, slip op. at 12-22 (PTAB Jan. 14, 2016)*



[Internet Archive: Overview of CyberHouse](#)

# IAWM provenance

- IAWM sources:
  - web-wide crawls
  - contract crawls
  - Archive-It crawls
  - Alexa donations
  - ArchiveTeam crawls
  - Save Page Now
- how does mixed provenance affect acceptance of reliability assertions?



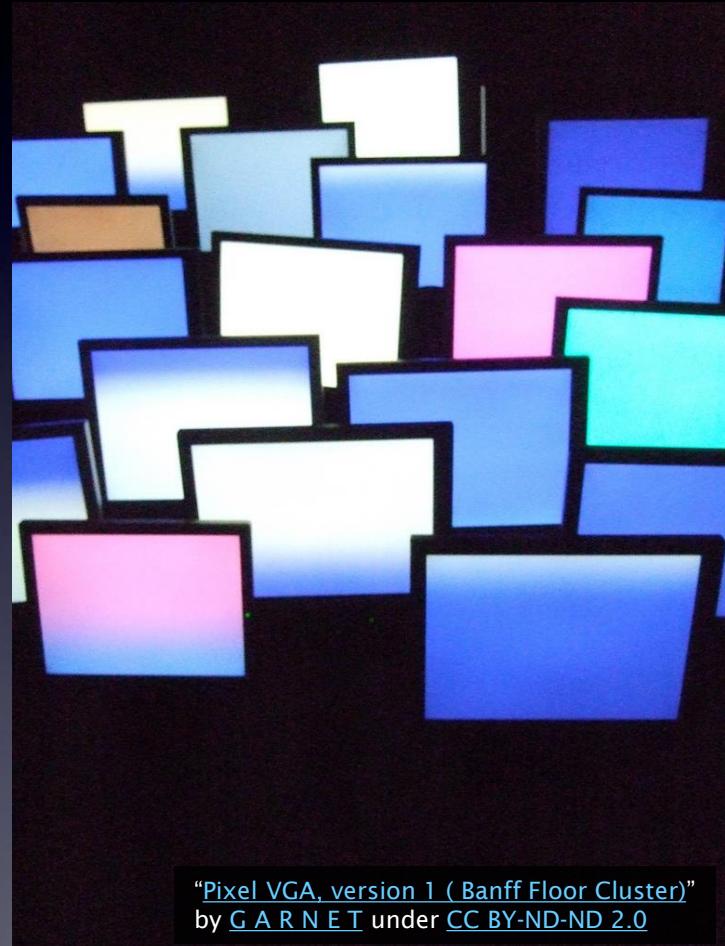
# other web archive repositories

The image displays a collage of screenshots from several web archive repositories, illustrating the diversity of such collections:

- UK Web Archive**: A screenshot of the UK Web Archive homepage, showing a timeline of archived websites from 2004 to 2009. It includes sections for "Quick search", "Explore the Special Collections" (with thumbnails for "Blogs", "Credit Crunch", "Live Art", "London Terrorist Att...", "Olympic & Paralympic...", "Personal Experiences...", "Quakers", "Queen's Diamond Jubil...", "UK General Election ..."), and "Visualisations". The UK Web Archive is provided by the British Library.
- CyberCemetery**: A screenshot of the CyberCemetery homepage, which is part of the Stanford Web Archive Portal. It features a banner for "CyberCemetery Home" and a timeline of archived websites from 2004 to 2009. The page includes sections for "Welcome to the UK Web Archive", "Quick links", "Browse by Subject", and "Visualisations".
- Stanford Web Archive Portal**: A screenshot of the Stanford Web Archive Portal homepage. It shows a search bar, a timeline of archived websites from 1995 to 1999, and a section for "Info" featuring a thumbnail of the SLAC home page from 1999.
- University of North Texas Libraries**: A screenshot of the University of North Texas Libraries homepage, showing a search bar and a timeline of archived websites from 1995 to 1999.
- SLAC home page 1999**: A thumbnail image of the SLAC home page from 1999, showing the Stanford Linear Accelerator Center website.

# canonicity

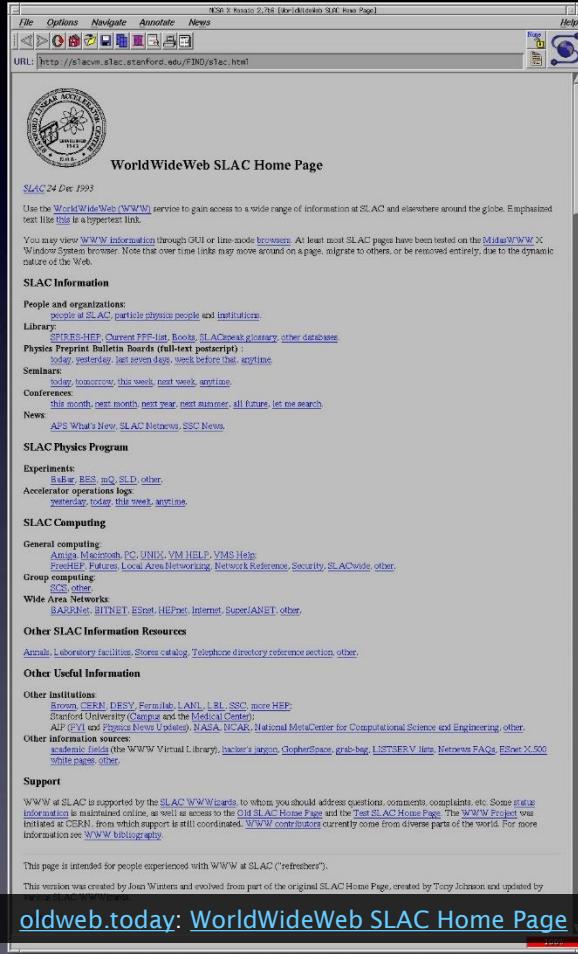
- personalization  
of content served  
to archival agent?
  - geolP
  - user-agent
  - cookies
  - authentication
  - split testing



[“Pixel VGA, version 1 \( Banff Floor Cluster\)”](#)  
by [G A R N E T](#) under [CC BY-ND-ND 2.0](#)

# client artifacts

- access via contemporaneous browser/OS
- enable plugins (e.g., Flash)
- disable client-side utilities affecting presentation (e.g., adblocker)





**WHAT NEXT?**

"Spiral Staircase - 11/52" by [Phil Roeder](#) under [CC BY 2.0](#)

# better serving legal (+ research?) use cases

- surface more provenance information
- improve interfaces to understand temporal coherence
  - toggle visibility of violative elements
- more broadly implement original HTTP header replay
- conduct more research on identifying personalized representations

# discussion

- **why else might we care**, or not, about legal use cases?
- **what other reliability factors** are relevant?
  - what is the **relative importance** of different reliability factors?
  - **for what use cases** are different reliability factors relevant?
- how are web archives being used in litigation in **other jurisdictions**?
- **what else could we do** to support legal use cases?