

# Web Science: Archiving the Web

## (Part 1 - Why Archive the Web?)

CS 432/532

Old Dominion University

*Permission has been granted to use these slides from Frank McCown, Michael L. Nelson, Alexander Nwala, Michele C. Weigle*



This work is licensed under a [Creative Commons](#)  
[Attribution-NonCommercial-ShareAlike 3.0 Unported License](#)

# Web Archiving

- **Web archiving** is the process of collecting pages from the Web and saving them in an archive
- Usually it's important to save all associated resources (images, style sheets, etc.) to preserve look
- Archives are typically produced using web crawlers

# Why archive the Web?

- If the Web isn't saved, we might lose a significant amount of our digital heritage
- The Web gives historians and other social scientists significant insight into our society, especially into how technology has affected it
- Serves as important resource in many lawsuits
- Some organizations want to or are legally obligated to archive their web materials
- Someone worked hard on this stuff, why not save it?

See also [WS-DL's celebration of IA's 20 year anniversary](#), Nov 2016

# Why Care About The Past?

From an anonymous WWW 2010 reviewer about our  
Memento paper (*emphasis mine*):

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

one answer: replay of contemporary pages >> summary pages

Ref: Nelson and Weigle, ["Why Care About the Past?"](#) (slides)

Requested Date: 1996 — 2010 07 / 02 / 2010

CNN Hurricane Katrina - Special Report

**cnn.com** Member Center: Sign In | Register International Edition

SEARCH THE WEB CNN.COM Search

CNN.com Home Page - U.S. - Weather More CNN.com sections: Special Reports

CNN's coverage of Katrina and the aftermath won a [2006 Peabody Award](#), the oldest honor in electronic media.

**Hurricane Katrina**

**HURRICANE KATRINA MAIN PAGE**

*Editor's Note: This special report was archived in September 2006.*

**Voices from the Gulf Coast**

August 29, 2005: Hurricane Katrina made landfall on the Gulf Coast, destroying lives, leveling homes and leaving thousands of survivors with the same story: We lost everything. One year later, many still coped with Katrina's devastation. Some had moved forward, but for others recovery was at a standstill. CNN.com visited the region to get their stories.

**Rebuilding**

- Mardi Gras
- Panoramas

**People**

- Your Stories
- One Family's Story
- Evacuation

**Rescue & Relief**

- Help Center

**Storm & Flood**

- Satellite Images

**Hurricane Season Special Report**



Transferring data from i.a.cnn.net...

**VS.**

CNN.com

http://web.archive.org/web/20050827154427/http://www.cnn.com/ google trends

1 Requested Date: 1996 2010 06 / 21 / 2010 ?

CNN.com Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 11:16 a.m. EDT (15:16 GMT), August 27, 2005

**Strengthening Katrina to hammer Gulf Coast**

A satellite image of Hurricane Katrina, showing its eye and surrounding clouds.

1,000 prisoners released from Iraq prison  
Brothers held again in Natalee Holloway case | Watch  
Vote to save bases in 2 states, lawsuit in third | Watch  
Watch: Anti-Sheehan group takes duel to Texas (2:20)  
Power-packed muscle cars come roaring back  
Oprah 'furious' at funeral snub claim  
U.S. soldier killed in Afghan attack  
Bush counsels patience in viewing Iraq war  
Criminal probe into 17 Paris fire deaths | Gallery | Watch

**MORE NEWS** Most Popular >

**WATCH FREE VIDEO** Browse/Search

Mistaken for a terrorist (2:17) Soundcheck with Faith Hill (2:35)

Now In The News (11:14 a.m. ET): Your quick news update

Featured Video More Video Picks

**BUSINESS** at CNNMONEY Markets: As of close Aug 26

DOW	53.34	10,397.29
NAS	13.60	2,120.77
S&P	7.29	1,205.10

ON CNN TV Schedule > Dead Wrong: Inside an Intelligence Meltdown

How did U.S. intelligence get it so wrong? CNN investigates how faulty intelligence led to the Iraq

Find: Next Previous Highlight all Match case

Read web.archive.org

August 27, 2005  
11:16 am EDT

CNN.com

http://web.archive.org/web/20050828031035/http://www.cnn.com/index.

Requested Date: 1996 2010 06 / 21 / 2010

CNN.com

**CNN.com** WITH FREE VIDEO Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 11:07 p.m. EDT (03:07 GMT), August 27, 2005

**Katrina drives people away from Gulf Coast**



People leaving New Orleans pack highways.

Tourists and New Orleans residents are leaving the city as Hurricane Katrina churns closer to Louisiana and Mississippi. Officials fear New Orleans is vulnerable because it sits an average of 6 feet below sea level. Katrina could be a category 4 when it makes landfall. "This is a very, very dangerous hurricane," says hurricane center Director Max Mayfield.

**FULL STORY**

• Watch: 'Very concerned' I 'Bir Easy' prepares

**MORE NEWS** Most Popular >

- Northwest strikers stay on picket lines
- California files french fry lawsuit
- Dual protests square off in Crawford | Watch
- CNN/Money: Greenspan: End of home boom inevitable
- Encephalitis outbreak hits children, poor | Watch
- Oprah 'furious' at funeral snub claim
- Watch: Will aging rats lead to younger people?
- Tall, twisted building opens
- Milkshake murder trial nears end
- Dutch investigate producer's 'Big Brother' baby plan

**WATCH FREE VIDEO** Browse/Search

Living longer (2:05)  Nail polish car (.56) 

Now In The News: Your quick news update

Featured Video More Video Picks

**BUSINESS** at CNNMoney

Markets: As of close Aug 26

DOW	-53.34	10,397.29
NAS	-13.60	2,120.77

ON CNN TV Schedule >

Reasonable Doubt: Can crime labs be trusted? DNA has cleared some innocent people. But forensic

Find: Next Previous Highlight all Match case

Read web.archive.org

August 27, 2005  
11:07 pm EDT

CNN.com

http://web.archive.org/web/20050828151004/http://www.cnn.com/index.

Requested Date: 1996 2010 06 / 21 / 2010

CNN.com Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 10:56 a.m. EDT (14:56 GMT), August 28, 2005

**New Orleans orders mandatory evacuation**

**BREAKING NEWS**



With powerful and strengthening Hurricane Katrina headed straight for his city, the mayor of New Orleans today ordered the mandatory evacuation of residents. The Category 5 storm has maximum sustained winds near 175 mph and has already been blamed for nine deaths in South Florida. Looking ahead, Max Mayfield of the National Hurricane Center said: "It has the potential for a large loss of life."

**DEVELOPING STORY**

**MORE NEWS** Most Popular >

- Committee signs Iraq's draft constitution | Watch
- Suicide bomber injures 21 in Israel | Watch
- California files french fry lawsuit
- Dual protests square off in Crawford | Watch
- Watch: Will aging rats lead to younger people?
- 'Milkshake murder' trial nears end
- Report: Swazi princess whipped for loud music
- USS Monitor's cat mystery
- Rap mogul 'Suge' Knight shot at Miami party

**WATCH FREE VIDEO** Browse/Search

Lap danger (2:48) Human cannonball crosses border (48)

Now in The News (10:33 a.m. ET): Your quick news update

Featured Video More Video Picks

**BUSINESS** at CNNMoney Markets: As of close Aug 26 DOW -53.34 10,397.29

**ON CNN TV** Schedule > Reasonable Doubt: Can crime labs be trusted? DNA has cleared some

Find: Next Previous Highlight all Match case

Read web.archive.org

[Archived NWS message, Aug 28, 2005, 10am CDT](#)

August 28, 2005  
11:34 am EDT

CNN.com

http://web.archive.org/web/20050830000348/http://www.cnn.com/ Google

1 Requested Date: 1996 2010 06 / 21 / 2010 ?

CNN.com +

**CNN.com** WITH FREE VIDEO Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 7:58 p.m. EDT (23:58 GMT), August 29, 2005

World U.S. Weather Business at CNNMoney Sports at SI.com Politics Law Technology Science & Space Health Entertainment Travel Education Special Reports Video NOW FREE Autos with Edmunds.com

**Killer Katrina heading inland**

Residents in New Orleans struggle to keep above the rising flood waters.

**Killer Katrina heading inland**

- Mississippi: Three dead, "unprecedented damage"
- More than 50 rescued from floods in New Orleans
- Katrina downgraded to tropical storm
- Parts of six Louisiana parishes "devastated"
- Oil rig breaks loose, hits Mobile River bridge
- "Life-threatening storm surge flooding" occurring

**DEVELOPING STORY**

- Watch: Wind flogs Mobile | A beaten 'Big Easy'
- Watch: Katrina's destructive day | Biloxi battered
- Gallery: Photoblog | CNN reporters: In the heart of storm
- Special Report: Tracking Katrina | In depth | Gallery
- Citizen Journalist: Your e-mails | Gallery | Share your story

**MORE NEWS** Most Popular

- New Orleans refugees told stay away | Watch
- Downtown Mobile under water | Watch
- CNN/Money: Wholesale gas prices soar | Watch
- TIME.com: Is global warming fueling Katrina? | Watch
- CNN/Money: Storm cost could hit \$16B. | Watch
- Struggling airlines hit by Katrina groundings
- Eight ex-KPMG execs indicted
- Workers struggle to save seals

**WATCH FREE VIDEO** Browse/Search

Ripping and roaring through the Gulf (2:27)  The Big Easy takes a battering (2:00) 

Now in The News (6:30 p.m. ET): Your quick news update

Featured Video More Video Picks

**BUSINESS** at CNNMoney Markets: 5:16 p.m. ET, Aug 29

**ON CNN TV** Schedule > Katrina continuing coverage

Find: Next Previous Highlight all Match case Done

August 28, 2005  
11:34 am EDT

CNN.com

http://web.archive.org/web/20050830031415/http://www.cnn.com/ Google

1 Requested Date: 1996 2010 06 / 21 / 2010 ?

CNN.com +

**CNN.com** Member Services International Edition | Netscape | MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH

Home Page Updated: 5:33 a.m. EDT (09:33 GMT), August 30, 2005

World U.S. Weather Business at CNNMoney Sports at SI.com Politics Law Technology Science & Space Health Entertainment Travel Education Special Reports Video NOW FREE Autos with Edmunds.com

Levee breached in New Orleans

Residents struggle to keep above the rising flood waters.

**54 DEAD IN MISSISSIPPI**

**Levee breached in New Orleans**

- Estimated 50 dead in 1 Mississippi county, officials say
- Levee breach in New Orleans forces hospital evacuation
- 1.3 million without power
- Tens of thousands to be homeless for months
- American Red Cross launching its largest relief operation

**FULL STORY**

- CNN Access: [My apartment complex is totally leveled](#)
- Watch: [Rescued from a roof](#) | [Water rises to eaves](#)
- Gallery: [Photoblog](#) | CNN reporters: In the heart of storm
- Special Report: [Tracking Katrina](#) | In depth | Gallery
- Citizen Journalist: [Your e-mails](#) | [Gallery](#) | [Share your story](#)

**MORE NEWS** Most Popular

- Huge relief effort under way
- 54 die in Mississippi, 50 in one county | Watch
- Louisiana evacuees told stay away | Watch
- Downtown Mobile under water | Watch
- CNN/Money: Wholesale gas prices soar | Watch
- TIME.com: Is global warming fueling Katrina?

**WATCH FREE VIDEO** Browse/Search

Up close on the devastation in New Orleans (1:57) A tour of the damage in New Orleans (2:42)

Now in The News: Your quick news update

Featured Video More Video Picks

Find: Next Previous Highlight all Match case

Waiting for web.archive.org...

August 30, 2005  
5:33 am EDT

CNN.com

http://web.archive.org/web/20050830175214/http://www.cnn.com/index.

1 Requested Date: 1996 2010 06 / 21 / 2010

CNN.com

International Edition | Netscape | MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH

Home Page World U.S. Weather Business at CNNMoney Sports at SI.com Politics Law Technology Science & Space Health Entertainment Travel Education Special Reports Video NOW FREE Autos with Edmunds.com

Updated: 1:44 p.m. EDT (17:44 GMT), August 30, 2005

# STATES OF DEVASTATION

This image is no longer available

Floodwaters from Hurricane Katrina cover a portion of New Orleans today.

## Death toll rising in Katrina's wake

- Impact: [Katrina's toll leaves Gulf Coast reeling](#) | [Watch](#)
- People: [Survivor describes harrowing experience](#) | [Watch](#)
- Recovery: [Aid groups sending rescuers, supplies](#) | [Watch](#)
- Assistance: [How to help storm victims](#) | [Watch](#)

### DEVELOPING STORY

- Watch: [Screams for help](#) | [Day that wouldn't end](#)
- Watch: [Katrina-spawned tornado destroys farm](#)
- Watch: ['Everything is gone'](#) | [Rooftop rescue](#) | [Katrina's fury](#)
- Special Report: [Tracking Katrina](#) | [Gallery](#) | [Map](#)
- Assisting victims: [Do you want to help?](#)
- Citizen journalist: [Your e-mails](#) | [Gallery](#) | [Share your story](#)

MORE NEWS Most Popular

- 'Significant' death toll in New Orleans | [Watch](#)
- [Mississippi governor vows to rebuild](#) | [Watch](#)
- Survivors 'screaming for help' | [Watch](#)
- Bush cuts short vacation because of Katrina
- CNN/Money: [Katrina closes 123 Wal-Marts](#)
- CNN/Money: [Oil hits record near \\$71](#) | [Watch](#)
- Katrina overwhelms weather Web sites

WATCH FREE VIDEO Browse/Search

Rescue efforts under way (2:47) Water pours into New Orleans (1:19)

Now in The News (1:31 p.m. ET): Your quick news update

Featured Video More Video Picks

Find: Next Previous Highlight all Match case

Waiting for cnn.122.2o7.net...

August 30, 2005  
1:44 pm EDT

CNN.com

http://web.archive.org/web/20050831103227/http://www.cnn.com/ google trends

Requested Date: 1996 2010 06 / 21 / 2010

CNN.com Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 6:14 a.m. EDT (10:14 GMT), August 31, 2005

# SLIPPING INTO CHAOS



A shotgun-toting New Orleans police officer has a word with a man on Canal Street.

## Unrest grows in New Orleans

- Water still rising in Crescent City, with no way out
- Looting, shootings reported
- New Orleans mayor blasts failure to patch levees
- Louisiana governor declares day of prayer
- Critical patients being airlifted, shelters being evacuated
- Paint marks indicate Mississippi homes with bodies

FULL STORY

- Watch: Chaotic aftermath | Tragedy, devastation
- Watch: The struggle to survive | Nature's fury explained
- Special Report: Trail of destruction | Gallery | Map | Blog
- Citizen journalist: Your e-mails | Gallery | Share your story
- Hotlines: Emergency info | How to help

MORE NEWS Most Popular

- Mayor blasts failure to patch levee breaches
- Tensions rising in New Orleans | Watch
- Relief efforts mobilizing | Watch
- Victims left with nothing | Watch
- States struggle with destruction | Watch
- Alabama power outage statewide | Watch
- Mississippi expects heavy death toll | Watch

WATCH FREE VIDEO Browse/Search

Lifted to safety (2:32) Hurricane Katrina's ongoing aftermath (4:00)

Now in The News: Your quick news update

Featured Video More Video Picks

Find: Next Previous Highlight all Match case

Transferring data from web.archive.org...

August 31, 2005  
6:14 am EDT

CNN.com

http://web.archive.org/web/20050831195555/http://www.cnn.com/ google trends

1 Requested Date: 1996 2010 06 / 21 / 2010 ?

CNN.com +

**CNN.com** WITH FREE VIDEO Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 3:44 p.m. EDT (19:44 GMT), August 31, 2005

World U.S. Weather Business at CNNMoney Sports at SI.com Politics Law Technology Science & Space Health Entertainment Travel Education Special Reports Video NOW FREE Autos with Edmunds.com

SURVIVORS ON APARTMENT BUILDING ROOF TRY TO GET ATTENTION OF RESCUERS

# HEALTH EMERGENCY

Health fears as bodies lie in water

- Public health emergency declared: cholera, typhoid fears
- Katrina death toll so far estimated at least 120
- New Orleans mayor says toll in his city will be hundreds
- Superdome evacuees moving to Astrodome

[FULL STORY](#)

- Watch: [Saved by a tree | Dome must be evacuated](#)
- Watch: [For Slidell, very bad bump in road](#)
- Special Report: [Trail of destruction | Gallery | Levee system](#)
- Citizen Journalist: [Your e-mails | Gallery | Share your story](#)
- Hotlines: [Where to get and give help](#)

Survivors on an apartment building roof try to get the attention of rescuers.

**MORE NEWS** **Most Popular**

- [Public health threat looming | Watch](#)
- [Superdome victims moving to Texas | Watch](#)
- [CNN reporters: Heartbreak across the coast](#)
- [Feds fan out for storm recovery efforts | Watch](#)
- [Wife makes desperate trip with husband's body](#)
- [Watch: Wife searches for missing husband](#)

**WATCH FREE VIDEO** **Browse/Search**

Health risks post-Katrina (2:18)  Clinging to a tree (3:25) 

Now In The News (3:31 p.m. ET): Your quick news update

Featured Video More Video Picks

Find: Next Previous Highlight all Match case

javascript:cnnVideo('play','/video/nitn/latest');

August 31, 2005  
3:44 pm EDT

CNN.com

http://web.archive.org/web/20050901023934/http://www.cnn.com/

1 Requested Date: 1996 2010 06 / 21 / 2010 ?

CNN.com +

**CNN.com** WITH FREE VIDEO Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 9:51 p.m. EDT (01:51 GMT), August 31, 2005

World U.S. Weather Business at CNNMoney Sports at SI.com Politics Law Technology Science & Space Health Entertainment Travel Education Special Reports Video NOW FREE Autos with Edmunds.com

**GETTING OUT**



**Evacuation of survivors begins**

- Public health emergency declared; cholera, typhoid fears
- First evacuated buses leave New Orleans
- Bush: "One of the worst national disasters"
- Extra 10,000 National Guard troops called up

**FULL STORY**

- Watch: Heartbreaking trip home | Searching for bodies
- Watch: How the levees broke | Saved by a tree
- Special Report: Trail of destruction | Gallery | Levee system
- Citizen Journalist: Your e-mails | Gallery | Share your story
- Hotlines: Where to get, give help | Tell people you're safe

Buses leave New Orleans with evacuees who had been holed up in the Superdome.

**MORE NEWS** **Most Popular**

- CNN exclusive: Police unable to aid sick woman
- Babies evacuated from ICU | Watch
- Public health threat looming | Watch
- Superdome evacuees head to Houston | Watch
- Huge sandbags planned for levee leaks |

**WATCH FREE VIDEO** **Browse/Search**

Desperate pleas (5:38)  Armed looting in New Orleans (2:30) 

Now in The News: Your quick news update

Featured Video More Video Picks

Find: Next Previous Highlight all Match case

Transferring data from web.archive.org...

August 31, 2005  
9:51 pm EDT

CNN.com

http://web.archive.org/web/20050901164946/http://www.cnn.com/ google trends

1 Requested Date: 1996 2010 06 / 21 / 2010 ?

CNN.com +

**CNN.com** WITH FREE VIDEO Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 12:30 p.m. EDT (16:30 GMT), September 1, 2005

# DESPERATION AND DANGER



Boat rescues suspended

- Boat rescuers asked to stand down due to dangers
- Huge crowd spotted at New Orleans convention center
- Gunfire, looting, small fires disrupt Superdome evacuation
- Citizens holed up in high-rises throng to buses
- Major buildings "pancaking" in Mississippi

**DEVELOPING STORY**

- Watch: 'Nobody in control' | Fires at Superdome
- Watch: Newborns rescued | Heartbreaking trip home
- Special Report: Trail of destruction | Gallery | Levee system
- Citizen Journalist: Your e-mails | Gallery | Share your story
- Hotlines: Where to get, give help | Tell people you're safe

Storm victims are stuck without food or water at the New Orleans Convention Center.

**MORE NEWS** Most Popular

- Bush tells victims: 'A lot of help coming'
- CNN/Money: Bush: 'Zero tolerance' on gouging
- Town wiped off the map by Katrina
- CNN exclusive: Police unable to aid sick woman
- Spain MP among stranded in New Orleans
- Federal rescue teams arriving

**WATCH FREE VIDEO** Browse/Search

FEMA suspends rescue operations (2:34) Anger and outrage (3:38)

Now In The News (11:48 a.m. ET): Your quick news update

Featured Video More Video Picks

x Find: Next Previous Highlight all Match case

Transferring data from web.archive.org...

September 1, 2005  
12:30 pm EDT

CNN.com

Requested Date: 1996 2010 06/21/2010

CNN.com International Edition | Netscape

Member Services MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH

Home Page Updated: 12:29 a.m. EDT (04:29 GMT), September 3, 2005

World U.S. Weather Business at CNNMoney Sports at SI.com Politics Law Technology Science & Space Health Entertainment Travel Education Special Reports Video NOW FREE Autos with Edmunds.com

This image is no longer available

Military personnel distribute food and water Friday at the convention center in New Orleans.

**AID IN, PEOPLE OUT**

**Some hope returns to New Orleans**

- Food and medicine arrives for desperate survivors
- 40,000 patients airlifted from hospitals
- Bush sees destruction -- and meets angry mayor
- But thousands still stranded in homes and centers

**DEVELOPING STORY**

- Watch: Fires and floods | Aid rolls in
- Watch: Bush: It seems dark right now | The military mission
- Special Report: Hurricane Katrina | Street-by-street
- Citizen Journalist: Your e-mails | Gallery | Share your story
- Hotlines: Where to get, give help | Tell people you're safe
- CNN TV: The struggle for New Orleans, on CNN

**MORE NEWS** Most Popular

- What was said vs. What was happening
- Exhausted evacuees pack airport | Watch
- On the Scene: Hopelessness in Mississippi
- Lt. Gen. Honore a 'John Wayne dude' | Watch
- Drying New Orleans could take months | Watch
- Bush admits relief efforts 'not enough' | Watch

**WATCH FREE VIDEO** Browse/Search

Role of race questioned (2:14) Medical nightmare (1:49)

Now in The News: Your quick news update

Featured Video More Video Picks

Find: Next Previous Highlight all Match case

Waiting for web.archive.org...

September 3, 2005  
12:29 am EDT

CNN.com

http://web.archive.org/web/20050904223351/http://www.cnn.com/ google trends

1 Requested Date: 1996 2010 06 / 21 / 2010 ?

CNN.com +

**CNN.com** WITH FREE VIDEO Member Services International Edition | Netscape MAKE CNN.com YOUR HOME PAGE

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

Home Page Updated: 6:08 p.m. EDT (22:08 GMT), September 4, 2005

# HELICOPTER HEROES

**BREAKING NEWS**



A U.S. Air Force rescuer is dropped onto a rooftop to help trapped Katrina victims.

**Rescuers going roof-to-roof**

- AP: Army engineers say police shoot dead its contractors
- Helicopters dropping supplies, picking up survivors
- Convention center mostly evacuated
- Mayor: Police, firefighters traumatized

**DEVELOPING STORY**

- Watch: Coast Guard rescuers | Some refuse to leave
- Watch: Mississippi wipe out | Mayor: Officers traumatized
- Hotlines: How to get, give help | Tell people you're safe
- Citizen Journalist: Your e-mails | Gallery | Share your story
- Special Report: Timeline | New Orleans: Block by block
- On CNN TV: Dr. Phil on healing after Katrina, 9 p.m. ET

**MORE NEWS** Most Popular

- Chief Justice Rehnquist dies at 80 | Watch
- Manning brothers team up for relief | Watch
- New Orleans couple weds in Mississippi shelter
- Mobile hospital stopped by red tape
- Misery has company at Mississippi motels
- Evacuees anguished at leaving pets behind

**WATCH FREE VIDEO** Browse/Search

Misery in Mississippi (2:16) Musical tribute to New Orleans (3:08)

Now in The News (5:52 p.m. ET): Your quick news update

Featured Video More Video Picks

Find: Next Previous Highlight all Match case

Waiting for web.archive.org...

September 4, 2005  
4:08 pm EDT

# Archiving Moves At Hurricane Speed, Most News Stories Move Faster

CNN.com - Breaking News, U.S., World, Weather, Entertainment & Video News

http://web.archive.org/web/20070416111317/http://www.cnn.com/ va tech shooting

1 Requested Date: 1996 2010 06/21/2010

CNN.com - Breaking News, U.S., W...

AMERICAN MORNING SITUATION ROOM LOU DOBBS TONIGHT PAULA ZAHN NOW LARRY KING LIVE ANDERSON COOPER 360° NANCY GRACE SCHEDULES International Edition CNN Pipeline

**cnn.com** Member Center: [Sign In](#) | [Register](#)

SEARCH THE WEB CNN.COM Search powered by YAHOO! SEARCH

Home World U.S. Weather Business Sports Analysis Politics Law Tech Science Health Entertainment Offbeat Travel Education Specials Autos I-Reports

UPDATED: 6:52 a.m. EDT, April 16, 2007

[Make CNN Your Home Page](#)

 advertisement

**LATEST NEWS**

[TOP STORIES](#) [MOST POPULAR](#)

- Fortune: [Newest Fortune 500 rankings | Video](#)
- Police: [Airline pilot on cocaine, driving wrong way](#)
- Al-Sadr pulls his ministers from Iraq government
- Search intensifies for U.S. aid worker
- Madonna back in Malawi to adopt?
- SI.com: [Jackie Robinson tributes | Video](#)
- Numbers add up to difficult tax year
- Iraqis turn to TV psychic for reassurance
- Investigators try to identify 8 skeletons
- Anti-Kremlin crowd clashes with police | Video
- Surfer girl attacked by sea lion
- Celebrity foot-in-mouth club gets new member
- CNN Wire: [Latest updates on world's top stories](#)

**WATCH VIDEO**

[MOST POPULAR](#) [BEST VIDEO](#)

- 1. [The year of celebrity slurs \(2:31\)](#)  
Don Imus' outburst is the latest in a string of celebrity slurs. CNN's Sibila Vargas reports. (April 14)  


**CNN Pipeline** live and commercial free video [Sign in | Help](#)

[LIVE](#) [WHAT IS PIPELINE?](#)

**CNN International**  
Live newscasts and in-depth programs from a global perspective, plus the Best of Pipeline.  


Find: Next Previous Highlight all Match case

Read web.archive.org

CNN.com - Breaking News, U.S., World, Weather, Entertainment & Video News

http://web.archive.org/web/20070416234959/cnn.com/ va tech shooting

1 Requested Date: 1996 2010 06/21/2010 ?

CNN.com - Breaking News, U.S., W... +

AMERICAN MORNING SITUATION ROOM LOU DOBBS TONIGHT PAULA ZAHN NOW LARRY KING LIVE ANDERSON COOPER 360° NANCY GRACE SCHEDULES International Edition CNN Pipeline

**cnn.com** Member Center: [Sign In](#) | [Register](#)

SEARCH THE WEB CNN.COM Search powered by YAHOO! SEARCH

Home World U.S. Weather Business Sports Analysis Politics Law Tech Science Health Entertainment Offbeat Travel Education Specials Autos I-Reports

UPDATED: 7:27 p.m. EDT, April 16, 2007 Make CNN Your Home Page

# BLOODBATH IN CLASS

**BREAKING NEWS**



PHOTO BY AP/WIDEWORLD

**Student: There was blood everywhere**

- 32 killed, 29 injured by lone gunman at Virginia Tech
- All I can tell you is he's a male, says police chief
- Killer reportedly dressed in Boy Scout-type outfit
- Student: I was one of only four that made it out
- It is the deadliest shooting rampage in U.S. history

**DEVELOPING STORY**

- [Virginia Tech shootings](#) (LIVE) | [Campus map](#)
- [Students describe their 'College Columbine'](#)
- [Cops draw guns, hunt shooter](#) | [Gallery](#)
- [I-Report: Gunshots captured on a cell phone](#)
- [Witness: We were all frightened](#) | [Send your videos](#)

**LATEST NEWS**

[TOP STORIES](#) [MOST POPULAR](#)

- [Officials thought killer left campus after first shooting](#)
- [Student captures video of shooting](#) | [Video](#)

**WATCH VIDEO**

[MOST POPULAR](#) [BEST VIDEO](#)

- [1. Gunfire on university campus](#) (5:03)  
A student describes gunfire on the [Gunfire on](#)

Find:  Next Previous  Highlight all  Match case

http://web.archive.org/web/20070416111317/http://www.cnn.com/2007/WORLD/africa/04/16/madonna.returns.reut/index.html

advertisement

CNN Pipeline live and commercial free video [Sign in](#) | [Help](#)

LIVE WHAT IS PIPELINE?

(LIVE) [LIVE BREAKING NEWS](#)  
Police and VA Tech officials hold a

CNN.com - Breaking News, U.S., World, Weather, Entertainment & Video News

http://web.archive.org/web/20070417071941/http://www.cnn.com/ va tech shooting

1 Requested Date: 1996 2010 06/21/2010 ?

CNN.com - Breaking News, U.S., W... +

AMERICAN MORNING SITUATION ROOM LOU DOBBS TONIGHT PAULA ZAHN NOW LARRY KING LIVE ANDERSON COOPER 360° NANCY GRACE SCHEDULES International Edition CNN Pipeline

**cnn.com** Member Center: [Sign In](#) | [Register](#)

SEARCH THE WEB CNN.COM Search powered by YAHOO! SEARCH

Home World U.S. Weather Business Sports Analysis Politics Law Tech Science Health Entertainment Offbeat Travel Education Specials Autos I-Reports

UPDATED: 3:17 a.m. EDT, April 17, 2007 Make CNN Your Home Page

# STUDENTS SLAUGHTERED



Students attend a Mass in Blacksburg, Virginia.

**Police have preliminary ID of gunman**

- Police not releasing name of Virginia Tech gunman
- Gunman massacres 32 at two Virginia Tech sites
- Some students jumped from windows to escape
- Professor is among dead at academic hall
- It is the deadliest shooting rampage in U.S. history

**FULL STORY**

- [Horrible day ends with sleepless night | Map](#)
- [Student: He shot almost the whole class | Gallery](#)
- [Time.com: Too few warnings? | Gallery](#)
- [I-Report: Gunshots captured on a cell phone | Gallery](#)
- [Witness: We were all frightened | Your photos](#)
- [Special Report: Massacre at Virginia Tech](#)

**LATEST NEWS**

[TOP STORIES](#) [MOST POPULAR](#)

- [Officials thought killer left campus after first shooting](#)

**WATCH VIDEO**

[MOST POPULAR](#) [BEST VIDEO](#)

- 1. [Eyewitness describes shooter \(1:23\)](#)

Find:  Next Previous  Highlight all  Match case

CNN Pipeline live and commercial free video [Sign in](#) | [Help](#)

LIVE  WHAT IS PIPELINE?

- CNN International

Read web.archive.org

# SHARI

The screenshot shows a web browser window titled "dsa-puddles" displaying the "Storygraph" category of the DSA Puddles project. The URL is <https://oduwSDL.github.io/dsa-puddles/categories/storygraph/>. The page has a dark background and features a grid of six news items, each with a thumbnail image, date, and a summary of the story's main points.

- 2020, SEP 13:** StoryGraph Biggest Story 2020-09-13 -- morbidity and mortality weekly reports (4), former trump campaign official michael caputo arrives at (4), department of health and human services (4), centers for disease control and prevention (4), bob woodward that (4)
- 2020, SEP 12:** StoryGraph Biggest Story 2020-09-12 -- joe biden (5), vice president (3), the covid 19 pandemic has sparked (3), in new york city (3), if the black lives matter movement (3)
- 2020, SEP 11:** StoryGraph Biggest Story 2020-09-11 -- joe biden (11), the former vice president (7), the american people (6), new york times (6), journalist bob woodward (6)
- 2020, SEP 10:** StoryGraph Biggest Story 2020-09-10 -- bob woodward (22), trump told woodward (16), i don't want to create panic as you say (11), joe biden (10), i still like playing it down (10)
- 2020, SEP 09:** StoryGraph Biggest Story 2020-09-09 -- bob woodward (13), trump told woodward (11), joe biden (10), covid 19 (10), of the coronavirus pandemic (9)
- 2020, SEP 08:** StoryGraph Biggest Story 2020-09-08 -- joe biden trump campaign works to (6), said biden (4), press conference (4), former vice president (4); biden trump campaign works to set (4)

WS-DL blog, "[SHARI: StoryGraph Hyperlane ArchiveNow Raintale Integration -- Combining WS-DL Tools For Current Events Storytelling](#)"  
[Storygraph stories at DSA Puddles](#)      [@StormyArchives](#) on Twitter



**StoryGraph Biggest Story 2020-04-10 -- the centers for disease control and prevention, new york, social distancing, public health, health care**

A news brief made from 25/85 recently archived news articles

[« Previous Story](#) [View the StoryGraph of this Story](#) [Next Story »](#)

<a href="#">Most frequent entities</a>	trump, the united states, americans, u.s., cdc
<a href="#">Most frequent sumgrams</a>	the centers for disease control and prevention, new york, social distancing, public health, health care

[StoryGraph Biggest Story 2020-04-10](#)



## StoryGraph Biggest Story 2020-05-31 -- of george floyd (24), tim walz (16), 46 year old black man killed by police during an (15), national guard (14), police officers (13)

A news brief made from 27/85 recently archived news articles

[« Previous Story](#)

[View the StoryGraph of this Story](#)

[Next Story »](#)

Most frequent entities

minneapolis, floyd, george floyd, trump, minnesota

Most frequent sumgrams

of george floyd (24), tim walz (16), 46 year old black man killed by police  
during an (15), national guard (14), police officers (13)

[StoryGraph Biggest Story 2020-05-31](#)



**StoryGraph Biggest Story 2020-06-09 -- new york (13), george floyd's death (13), death george (8), civil rights (7), the coronavirus pandemic (6)**

A news brief made from 15/85 recently archived news articles

[« Previous Story](#) [View the StoryGraph of this Story](#) [Next Story »](#)

<a href="#">Most frequent entities</a>	trump, george floyd, new york, the united states, zachary halaschak
<a href="#">Most frequent sumgrams</a>	new york (13), george floyd's death (13), death george (8), civil rights (7), the coronavirus pandemic (6)

[StoryGraph Biggest Story 2020-06-09](#)



## StoryGraph Biggest Story 2020-08-12 -- kamala harris (25), joe biden's (20), sen kamala (18), running mate (16), vice presidential (15)

A news brief made from 28/85 recently archived news articles

[« Previous Story](#)

[View the StoryGraph of this Story](#)

[Next Story »](#)

Most frequent entities

harris, kamala harris, joe biden, trump, biden

Most frequent sumgrams

kamala harris (25), joe biden's (20), sen kamala (18), running mate (16), vice presidential (15)

[StoryGraph Biggest Story 2020-08-12](#)

# Web Science: Archiving the Web

## (Part 2 - How is the Web Archived and How Much of the Web Is Archived?)

CS 432/532

Old Dominion University

*Most slides are courtesy Dr. Frank McCown, Dr. Michael L. Nelson, and Alexander Nwala*



This work is licensed under a [Creative Commons](#)  
[Attribution-NonCommercial-ShareAlike 3.0 Unported License](#)

# Who is Archiving the Web?

# Heroic solo efforts

TMW 172 doomed BBC website... x

← → C thethenextweb.com/uk/2011/02/10/172-doomed-bbc-websites-saved-by-one-geek-for-3-99

Navigation PREV IN UK 10/02/2011, Martin Bryant NEXT IN UK

ALL STORIES TOP STORIES CHANNELS APPS APPLE DAILY DOSE DESIGN & DEV ENTREPRENEUR EVENTS FACEBOOK GADGETS GOOGLE INDUSTRY LIFEHACKS LOCATION

## 172 doomed BBC websites saved by one geek, for \$3.99

Tweet 253 SHARE 155 172 doomed BBC website

**Update at foot of post.**

As we recently [reported](#), the BBC is set to close down 200 of its websites in the near future as part of cost-cutting measures. Hearing that 172 of these sites would be deleted from the Web entirely, an anonymous (and possibly even "Anonymous", judging by the picture on the website) individual has [taken matters into his or her own hands](#).



good? yes. complete story? no.

1TB endowment = ~\$4700: Rosenthal, "[Paying for Long-Term Storage](#)", Feb 2011

# Who is archiving the Web?

- Internet Archive
  - Founded by Brewster Kahle in 1996
  - Largest web archive in the world (~470B webpages)
    - Pages available to public via the [Wayback Machine](#)



-200k software programs, 20M books, 4M videos, 4.5M audio recordings, 3M images (as of Jan 2020)

Image sources: ["About IA"](#)

← → ⌂ blog.archive.org/2013/01/09/updated-wayback/ ⋮

# Internet Archive Blogs

A blog from the team at archive.org



Home Announcements About

← Thanks a Million!

Aaron Swartz, hero of the open world, dies →

## Wayback Machine: Now with 240,000,000,000 URLs

Posted on January 9, 2013 by brewster

Today we updated the [Wayback Machine](#) with much more data and some code improvements. Now we cover from late 1996 to December 9, 2012 so you can surf the web as it was up until a month ago. Also, we have gone from having 150,000,000,000 URLs to having 240,000,000,000 URLs, a total of about 5



### Recent Posts

- Knight Foundation Strengthens Support for Television News Research Service
- National Security Agency ♥ ♥ Internet Archive?
- Brewster Kahle to be Honored with 2013 Amer Lib Assoc LITA/Library Hi Tech Award

85B in 2006: [Wayback Machine has 85 Billion Archived Webpages](#)

532B URIs, 279B pages in Jan 2017: [Brewster Kahle on Twitter, Jan 2018](#)

703B URIs, Jan 2019: [Brewster Kahle on Twitter, Jan 2019](#)

881B URIs, Jan 2020: [Brewster Kahle on Twitter, Jan 2020](#)

468B pages, Sep 2020: [Internet Archive on Twitter, Sep 2020](#)

Books to Borrow : Free Texts : F X

https://archive.org/details/inlibrary

INTERN ARCHIVE WEB BOOKS VIDEO AUDIO SOFTWARE IMAGES

ABOUT BLOG PROJECTS HELP DONATE CONTACT JOBS VOLUNTEER PEOPLE

SIGN UP | LOG IN UPLOAD Search



## Books to Borrow

Share  
Favorite  
RSS

Books in this collection may be borrowed by logged in patrons. You may read the books online in your browser or, in some cases, download them into [Adobe Digital Editions](#), a free piece of software used for managing loans.

ABOUT

COLLECTION

1,729,317 RESULTS

Search this Collection

- Metadata  
 Text contents

Media Type

texts 1,729,317

Availability

Lending Library 1,729,055

Year

- 2015 16,544  
 2014 22,632  
 2013 27,160  
 2012 31,808  
 2011 32,654  
 2010 35,978  
[More ▶](#)

Topics & Subjects

SORT BY VIEWS • TITLE • DATE PUBLISHED • CREATOR



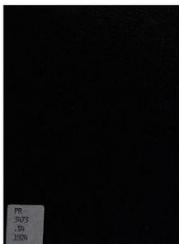
The sands of time,  
recollections and  
1924



Victory  
1924



The masters of modern  
art  
1924



The shepherd's week :  
in six pastorals  
1924



Jonathan Swift :  
selections  
1924



[Books to Borrow](#)

Internet Arcade : Free Software X +

https://archive.org/details/internetarcade

SIGN UP | LOG IN UPLOAD Search

ABOUT CONTACT BLOG PROJECTS HELP DONATE JOBS VOLUNTEER PEOPLE

INTERNET ARCADE

Internet Arcade

The Internet Arcade is a web-based library of arcade (coin-operated) video games from the 1970s through to the 1990s, emulated in JSMESS, part of the JSMESS software package.

Share Favorite RSS Play All

ABOUT COLLECTION

1,722 RESULTS

SORT BY VIEWS TITLE DATE ARCHIVED CREATOR

Search this Collection

Metadata Text contents (no results)

Media Type software 1,722

Year

2016	1
2014	1
2008	3
2007	1
2006	2
2003	4
More ▾	

Topics & Subjects

arcade	3
maze	3
Berzerk Franchise	2
shooter	2
speech	2
stem	2
More ▾	

DEFENDER Internet Arcade: Defender by Williams 636,818 227 19

ASTRO BLASTER Internet Arcade: Astro Blaster by Sega 488,363 54 8

THE THREE STOOGES Internet Arcade: The Three Stooges In Brides Is by MyStar 470,942 38 2

BERZERK Internet Arcade: Berzerk by Stern Electronics 462,327 100 17

JOUST Internet Arcade: Joust by Williams 454,130 255 14

INDIANA JONES AND THE TEMPLE OF DOOM Internet Arcade: Indiana Jones and the Temple of Doom by Atari Games 436,251 129 10

BATMAN Internet Arcade: Batman by Valadon Automation (Stern Electronics License) 426,550 62 4

BIO ATTACK Internet Arcade: Bio Attack by Taito Corporation (Fox Video Games license) 414,533 11 1

ATOMIC BOY Internet Arcade: Atomic Boy by Irem (Memtron license) 407,906 17 3

BANK PANIC Internet Arcade: Bank Panic by Sanritsu / Sega 407,752 48 5

BASEBALL CHAMPION Internet Arcade: Champion Baseball by Alpha Denchi Co. 407,538 19 4

BATTLE OF ATLANTIS Internet Arcade: Battle of Atlantis by Comsoft 407,038 90 1

AZTARAC Internet Arcade: Aztarac by Centuri 399,641 6 0

PAPERBOY Internet Arcade: Paperboy by Atari Games 397,551 154 7

BEEZER Internet Arcade: Beezer by Tong Electronic 397,090 3 0

Internet Arcade

Internet Archive TV NEWS : See [+](#)

<https://archive.org/details/tv?q=%22black+lives+matter%22>

INTERNET ARCHIVE [ABOUT](#) [CONTACT](#) [BLOG](#) [PROJECTS](#) [HELP](#) [DONATE](#) [JOBS](#) [VOLUNTEER](#) [PEOPLE](#)

SEARCH "black lives m" [GO](#) [Advanced Search](#)

**TV News Archive**

**Search** "black lives matter" [GO](#)

**25,305 RESULTS**

SORT BY RELEVANCE · VIEWS · TITLE · DATE ARCHIVED · CREATOR

**Program**

- Democracy Now! 768
- News 480
- Key Capitol Hill Hearings 399
- CBS This Morning 342
- World News Now 263
- Hannity 256
- [More ▶](#)

**Creator**

- FOXNEWS (FOX W News) 2,649
- MSNBC(MSNBC) 2,062
- CNN(CNN) 1,782
- CSPAN 1,482
- CSPAN2(CSPAN) 1,349
- FBC(FOX Business) 977
- [More ▶](#)

**Year**

- 2020 2,618
- [2019](#)
- [2018](#)
- [2017](#)
- [2016](#)
- [2015](#)
- [2014](#)
- [2013](#)
- [2012](#)
- [2011](#)
- [2010](#)
- [2009](#)
- [2008](#)
- [2007](#)
- [2006](#)
- [2005](#)
- [2004](#)
- [2003](#)
- [2002](#)
- [2001](#)
- [2000](#)
- [1999](#)
- [1998](#)
- [1997](#)
- [1996](#)
- [1995](#)
- [1994](#)
- [1993](#)
- [1992](#)
- [1991](#)
- [1990](#)
- [1989](#)
- [1988](#)
- [1987](#)
- [1986](#)
- [1985](#)
- [1984](#)
- [1983](#)
- [1982](#)
- [1981](#)
- [1980](#)
- [1979](#)
- [1978](#)
- [1977](#)
- [1976](#)
- [1975](#)
- [1974](#)
- [1973](#)
- [1972](#)
- [1971](#)
- [1970](#)
- [1969](#)
- [1968](#)
- [1967](#)
- [1966](#)
- [1965](#)
- [1964](#)
- [1963](#)
- [1962](#)
- [1961](#)
- [1960](#)
- [1959](#)
- [1958](#)
- [1957](#)
- [1956](#)
- [1955](#)
- [1954](#)
- [1953](#)
- [1952](#)
- [1951](#)
- [1950](#)
- [1949](#)
- [1948](#)
- [1947](#)
- [1946](#)
- [1945](#)
- [1944](#)
- [1943](#)
- [1942](#)
- [1941](#)
- [1940](#)
- [1939](#)
- [1938](#)
- [1937](#)
- [1936](#)
- [1935](#)
- [1934](#)
- [1933](#)
- [1932](#)
- [1931](#)
- [1930](#)
- [1929](#)
- [1928](#)
- [1927](#)
- [1926](#)
- [1925](#)
- [1924](#)
- [1923](#)
- [1922](#)
- [1921](#)
- [1920](#)
- [1919](#)
- [1918](#)
- [1917](#)
- [1916](#)
- [1915](#)
- [1914](#)
- [1913](#)
- [1912](#)
- [1911](#)
- [1910](#)
- [1909](#)
- [1908](#)
- [1907](#)
- [1906](#)
- [1905](#)
- [1904](#)
- [1903](#)
- [1902](#)
- [1901](#)
- [1900](#)
- [1899](#)
- [1898](#)
- [1897](#)
- [1896](#)
- [1895](#)
- [1894](#)
- [1893](#)
- [1892](#)
- [1891](#)
- [1890](#)
- [1889](#)
- [1888](#)
- [1887](#)
- [1886](#)
- [1885](#)
- [1884](#)
- [1883](#)
- [1882](#)
- [1881](#)
- [1880](#)
- [1879](#)
- [1878](#)
- [1877](#)
- [1876](#)
- [1875](#)
- [1874](#)
- [1873](#)
- [1872](#)
- [1871](#)
- [1870](#)
- [1869](#)
- [1868](#)
- [1867](#)
- [1866](#)
- [1865](#)
- [1864](#)
- [1863](#)
- [1862](#)
- [1861](#)
- [1860](#)
- [1859](#)
- [1858](#)
- [1857](#)
- [1856](#)
- [1855](#)
- [1854](#)
- [1853](#)
- [1852](#)
- [1851](#)
- [1850](#)
- [1849](#)
- [1848](#)
- [1847](#)
- [1846](#)
- [1845](#)
- [1844](#)
- [1843](#)
- [1842](#)
- [1841](#)
- [1840](#)
- [1839](#)
- [1838](#)
- [1837](#)
- [1836](#)
- [1835](#)
- [1834](#)
- [1833](#)
- [1832](#)
- [1831](#)
- [1830](#)
- [1829](#)
- [1828](#)
- [1827](#)
- [1826](#)
- [1825](#)
- [1824](#)
- [1823](#)
- [1822](#)
- [1821](#)
- [1820](#)
- [1819](#)
- [1818](#)
- [1817](#)
- [1816](#)
- [1815](#)
- [1814](#)
- [1813](#)
- [1812](#)
- [1811](#)
- [1810](#)
- [1809](#)
- [1808](#)
- [1807](#)
- [1806](#)
- [1805](#)
- [1804](#)
- [1803](#)
- [1802](#)
- [1801](#)
- [1800](#)
- [1799](#)
- [1798](#)
- [1797](#)
- [1796](#)
- [1795](#)
- [1794](#)
- [1793](#)
- [1792](#)
- [1791](#)
- [1790](#)
- [1789](#)
- [1788](#)
- [1787](#)
- [1786](#)
- [1785](#)
- [1784](#)
- [1783](#)
- [1782](#)
- [1781](#)
- [1780](#)
- [1779](#)
- [1778](#)
- [1777](#)
- [1776](#)
- [1775](#)
- [1774](#)
- [1773](#)
- [1772](#)
- [1771](#)
- [1770](#)
- [1769](#)
- [1768](#)
- [1767](#)
- [1766](#)
- [1765](#)
- [1764](#)
- [1763](#)
- [1762](#)
- [1761](#)
- [1760](#)
- [1759](#)
- [1758](#)
- [1757](#)
- [1756](#)
- [1755](#)
- [1754](#)
- [1753](#)
- [1752](#)
- [1751](#)
- [1750](#)
- [1749](#)
- [1748](#)
- [1747](#)
- [1746](#)
- [1745](#)
- [1744](#)
- [1743](#)
- [1742](#)
- [1741](#)
- [1740](#)
- [1739](#)
- [1738](#)
- [1737](#)
- [1736](#)
- [1735](#)
- [1734](#)
- [1733](#)
- [1732](#)
- [1731](#)
- [1730](#)
- [1729](#)
- [1728](#)
- [1727](#)
- [1726](#)
- [1725](#)
- [1724](#)
- [1723](#)
- [1722](#)
- [1721](#)
- [1720](#)
- [1719](#)
- [1718](#)
- [1717](#)
- [1716](#)
- [1715](#)
- [1714](#)
- [1713](#)
- [1712](#)
- [1711](#)
- [1710](#)
- [1709](#)
- [1708](#)
- [1707](#)
- [1706](#)
- [1705](#)
- [1704](#)
- [1703](#)
- [1702](#)
- [1701](#)
- [1700](#)
- [1699](#)
- [1698](#)
- [1697](#)
- [1696](#)
- [1695](#)
- [1694](#)
- [1693](#)
- [1692](#)
- [1691](#)
- [1690](#)
- [1689](#)
- [1688](#)
- [1687](#)
- [1686](#)
- [1685](#)
- [1684](#)
- [1683](#)
- [1682](#)
- [1681](#)
- [1680](#)
- [1679](#)
- [1678](#)
- [1677](#)
- [1676](#)
- [1675](#)
- [1674](#)
- [1673](#)
- [1672](#)
- [1671](#)
- [1670](#)
- [1669](#)
- [1668](#)
- [1667](#)
- [1666](#)
- [1665](#)
- [1664](#)
- [1663](#)
- [1662](#)
- [1661](#)
- [1660](#)
- [1659](#)
- [1658](#)
- [1657](#)
- [1656](#)
- [1655](#)
- [1654](#)
- [1653](#)
- [1652](#)
- [1651](#)
- [1650](#)
- [1649](#)
- [1648](#)
- [1647](#)
- [1646](#)
- [1645](#)
- [1644](#)
- [1643](#)
- [1642](#)
- [1641](#)
- [1640](#)
- [1639](#)
- [1638](#)
- [1637](#)
- [1636](#)
- [1635](#)
- [1634](#)
- [1633](#)
- [1632](#)
- [1631](#)
- [1630](#)
- [1629](#)
- [1628](#)
- [1627](#)
- [1626](#)
- [1625](#)
- [1624](#)
- [1623](#)
- [1622](#)
- [1621](#)
- [1620](#)
- [1619](#)
- [1618](#)
- [1617](#)
- [1616](#)
- [1615](#)
- [1614](#)
- [1613](#)
- [1612](#)
- [1611](#)
- [1610](#)
- [1609](#)
- [1608](#)
- [1607](#)
- [1606](#)
- [1605](#)
- [1604](#)
- [1603](#)
- [1602](#)
- [1601](#)
- [1600](#)
- [1599](#)
- [1598](#)
- [1597](#)
- [1596](#)
- [1595](#)
- [1594](#)
- [1593](#)
- [1592](#)
- [1591](#)
- [1590](#)
- [1589](#)
- [1588](#)
- [1587](#)
- [1586](#)
- [1585](#)
- [1584](#)
- [1583](#)
- [1582](#)
- [1581](#)
- [1580](#)
- [1579](#)
- [1578](#)
- [1577](#)
- [1576](#)
- [1575](#)
- [1574](#)
- [1573](#)
- [1572](#)
- [1571](#)
- [1570](#)
- [1569](#)
- [1568](#)
- [1567](#)
- [1566](#)
- [1565](#)
- [1564](#)
- [1563](#)
- [1562](#)
- [1561](#)
- [1560](#)
- [1559](#)
- [1558](#)
- [1557](#)
- [1556](#)
- [1555](#)
- [1554](#)
- [1553](#)
- [1552](#)
- [1551](#)
- [1550](#)
- [1549](#)
- [1548](#)
- [1547](#)
- [1546](#)
- [1545](#)
- [1544](#)
- [1543](#)
- [1542](#)
- [1541](#)
- [1540](#)
- [1539](#)
- [1538](#)
- [1537](#)
- [1536](#)
- [1535](#)
- [1534](#)
- [1533](#)
- [1532](#)
- [1531](#)
- [1530](#)
- [1529](#)
- [1528](#)
- [1527](#)
- [1526](#)
- [1525](#)
- [1524](#)
- [1523](#)
- [1522](#)
- [1521](#)
- [1520](#)
- [1519](#)
- [1518](#)
- [1517](#)
- [1516](#)
- [1515](#)
- [1514](#)
- [1513](#)
- [1512](#)
- [1511](#)
- [1510](#)
- [1509](#)
- [1508](#)
- [1507](#)
- [1506](#)
- [1505](#)
- [1504](#)
- [1503](#)
- [1502](#)
- [1501](#)
- [1500](#)
- [1499](#)
- [1498](#)
- [1497](#)
- [1496](#)
- [1495](#)
- [1494](#)
- [1493](#)
- [1492](#)
- [1491](#)
- [1490](#)
- [1489](#)
- [1488](#)
- [1487](#)
- [1486](#)
- [1485](#)
- [1484](#)
- [1483](#)
- [1482](#)
- [1481](#)
- [1480](#)
- [1479](#)
- [1478](#)
- [1477](#)
- [1476](#)
- [1475](#)
- [1474](#)
- [1473](#)
- [1472](#)
- [1471](#)
- [1470](#)
- [1469](#)
- [1468](#)
- [1467](#)
- [1466](#)
- [1465](#)
- [1464](#)
- [1463](#)
- [1462](#)
- [1461](#)
- [1460](#)
- [1459](#)
- [1458](#)
- [1457](#)
- [1456](#)
- [1455](#)
- [1454](#)
- [1453](#)
- [1452](#)
- [1451](#)
- [1450](#)
- [1449](#)
- [1448](#)
- [1447](#)
- [1446](#)
- [1445](#)
- [1444](#)
- [1443](#)
- [1442](#)
- [1441](#)
- [1440](#)
- [1439](#)
- [1438](#)
- [1437](#)
- [1436](#)
- [1435](#)
- [1434](#)
- [1433](#)
- [1432](#)
- [1431](#)
- [1430](#)
- [1429](#)
- [1428](#)
- [1427](#)
- [1426](#)
- [1425](#)
- [1424](#)
- [1423](#)
- [1422](#)
- [1421](#)
- [1420](#)
- [1419](#)
- [1418](#)
- [1417](#)
- [1416](#)
- [1415](#)
- [1414](#)
- [1413](#)
- [1412](#)
- [1411](#)
- [1410](#)
- [1409](#)
- [1408](#)
- [1407](#)
- [1406](#)
- [1405](#)
- [1404](#)
- [1403](#)
- [1402](#)
- [1401](#)
- [1400](#)
- [1399](#)
- [1398](#)
- [1397](#)
- [1396](#)
- [1395](#)
- [1394](#)
- [1393](#)
- [1392](#)
- [1391](#)
- [1390](#)
- [1389](#)
- [1388](#)
- [1387](#)
- [1386](#)
- [1385](#)
- [1384](#)
- [1383](#)
- [1382](#)
- [1381](#)
- [1380](#)
- [1379](#)
- [1378](#)
- [1377](#)
- [1376](#)
- [1375](#)
- [1374](#)
- [1373](#)
- [1372](#)
- [1371](#)
- [1370](#)
- [1369](#)
- [1368](#)
- [1367](#)
- [1366](#)
- [1365](#)
- [1364](#)
- [1363](#)
- [1362](#)
- [1361](#)
- [1360](#)
- [1359](#)
- [1358](#)
- [1357](#)
- [1356](#)
- [1355](#)
- [1354](#)
- [1353](#)
- [1352](#)
- [1351](#)
- [1350](#)
- [1349](#)
- [1348](#)
- [1347](#)
- [1346](#)
- [1345](#)
- [1344](#)
- [1343](#)
- [1342](#)
- [1341](#)
- [1340](#)
- [1339](#)
- [1338](#)
- [1337](#)
- [1336](#)
- [1335](#)
- [1334](#)
- [1333](#)
- [1332](#)
- [1331](#)
- [1330](#)
- [1329](#)
- [1328](#)
- [1327](#)
- [1326](#)
- [1325](#)
- [1324](#)
- [1323](#)
- [1322](#)
- [1321](#)
- [1320](#)
- [1319](#)
- [1318](#)
- [1317](#)
- [1316](#)
- [1315](#)
- [1314](#)
- [1313](#)
- [1312](#)
- [1311](#)
- [1310](#)
- [1309](#)
- [1308](#)
- [1307](#)
- [1306](#)
- [1305](#)
- [1304](#)
- [1303](#)
- [1302](#)
- [1301](#)
- [1300](#)
- [1299](#)
- [1298](#)
- [1297](#)
- [1296](#)
- [1295](#)
- [1294](#)
- [1293](#)
- [1292](#)
- [1291](#)
- [1290](#)
- [1289](#)
- [1288](#)
- [1287](#)
- [1286](#)
- [1285](#)
- [1284](#)
- [1283](#)
- [1282](#)
- [1281](#)
- [1280](#)
- [1279](#)
- [1278](#)
- [1277](#)
- [1276](#)
- [1275](#)
- [1274](#)
- [1273](#)
- [1272](#)
- [1271](#)
- [1270](#)
- [1269](#)
- [1268](#)
- [1267](#)
- [1266](#)
- [1265](#)
- [1264](#)
- [1263](#)
- [1262](#)
- [1261](#)
- [1260](#)
- [1259](#)
- [1258](#)
- [1257](#)
- [1256](#)
- [1255](#)
- [1254](#)
- [1253](#)
- [1252](#)
- [1251](#)
- [1250](#)
- [1249](#)
- [1248](#)
- [1247](#)
- [1246](#)
- [1245](#)
- [1244](#)
- [1243](#)
- [1242](#)
- [1241](#)
- [1240](#)
- [1239](#)
- [1238](#)
- [1237](#)
- [1236](#)
- [1235](#)
- [1234](#)
- [1233](#)
- [1232](#)
- [1231](#)
- [1230](#)
- [1229](#)
- [1228](#)
- [1227](#)
- [1226](#)
- [1225](#)
- [1224](#)
- [1223](#)
- [1222](#)
- [1221](#)
- [1220](#)
- [1219](#)
- [1218](#)
- [1217](#)
- [1216](#)
- [1215](#)
- [1214](#)
- [1213](#)
- [1212](#)
- [1211](#)
- [1210](#)
- [1209](#)
- [1208](#)
- [1207](#)
- [1206](#)
- [1205](#)
- [1204](#)
- [1203](#)
- [1202](#)
- [1201](#)
- [1200](#)
- [1199](#)
- [1198](#)
- [1197](#)
- [1196](#)
- [1195](#)
- [1194](#)
- [1193](#)
- [1192](#)
- [1191](#)
- [1190](#)
- [1189](#)
- [1188](#)
- [1187](#)
- [1186](#)
- [1185](#)
- [1184](#)
- [1183](#)
- [1182](#)
- [1181](#)
- [1180](#)
- [1179](#)
- [1178](#)
- <

Internet Archive TV NEWS : Se x + archive.org/services/third-eye.php

INTERNET ARCHIVE ABOUT CONTACT BLOG PROJECTS HELP DONATE VOLUNTEER JOBS PEOPLE SIGN IN UPLOAD Search Paused

## Third Eye: API for Internet Archive TV News chyrons

How we turn TV News chyrons into data

TV cable news channels display [chyrons](#) on the "lower thirds" of screens, to display breaking news and other highlights. Using the [Internet Archive TV News](#), TV architect Tracey Jaquith built the Third Eye to scan the lower parts of the screen and apply OCR, or optical character recognition, to turn the words into text.

Third Eye captures four TV cable news channels: BBC News, CNN, Fox News, and MSNBC. The project launched with [four million chyrons](#) captured in just over [two weeks](#).

TV chyron "lower third" and OCR example:

AFTER WH MEETING, SCHUMER DISHES  
WHEN HE THOUGHT MIC WAS OFF

LIVE

MSNBC

AFTER WH MEETING, SCHUMER DISHES  
WHEN HE THOUGHT MIC WAS OFF

ING, SCHUMER DISHES  
UGHT NIC WAS OFF

[Third Eye: API for Internet Archive TV News chyrons](#)

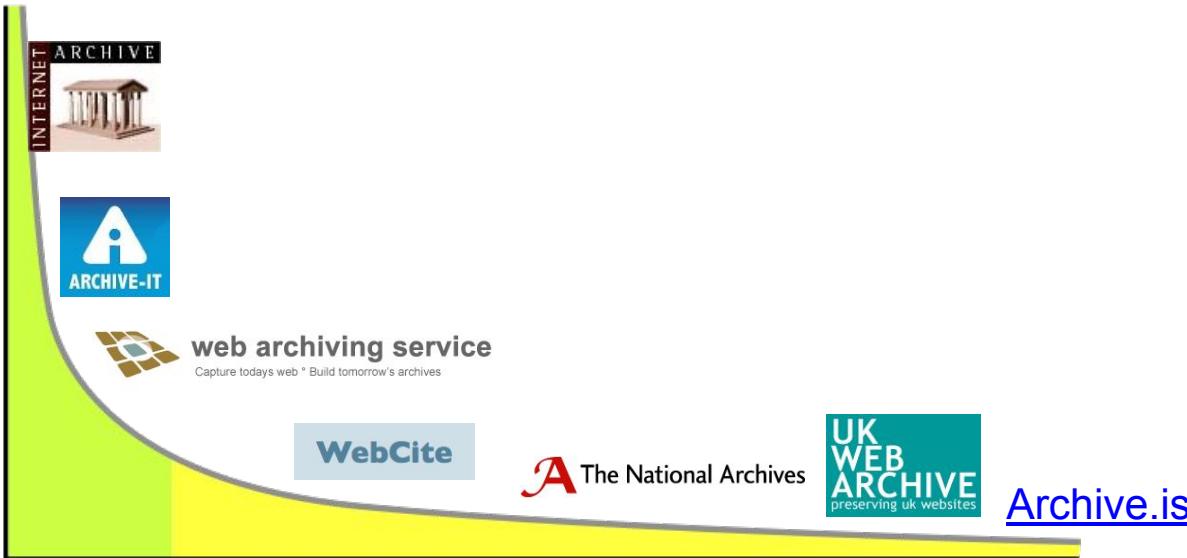
# Who is archiving the Web?

- National libraries & national archives, usually focusing on culturally significant web collections
  - list from the IIPC: [Member Archives \(page archived in 2017\)](#)
- Commercial organizations
  - Hanzo Archives: commercial web archiving tools
    - ref: [Compliance Archiving](#)
  - Nextpoint: archiving service for organizations
    - ref: [Automated Ediscovery Software / How It Works](#)
  - PageFreezer: electronic records management
    - ref: [Pagefreezer Website Archiving Solutions](#)

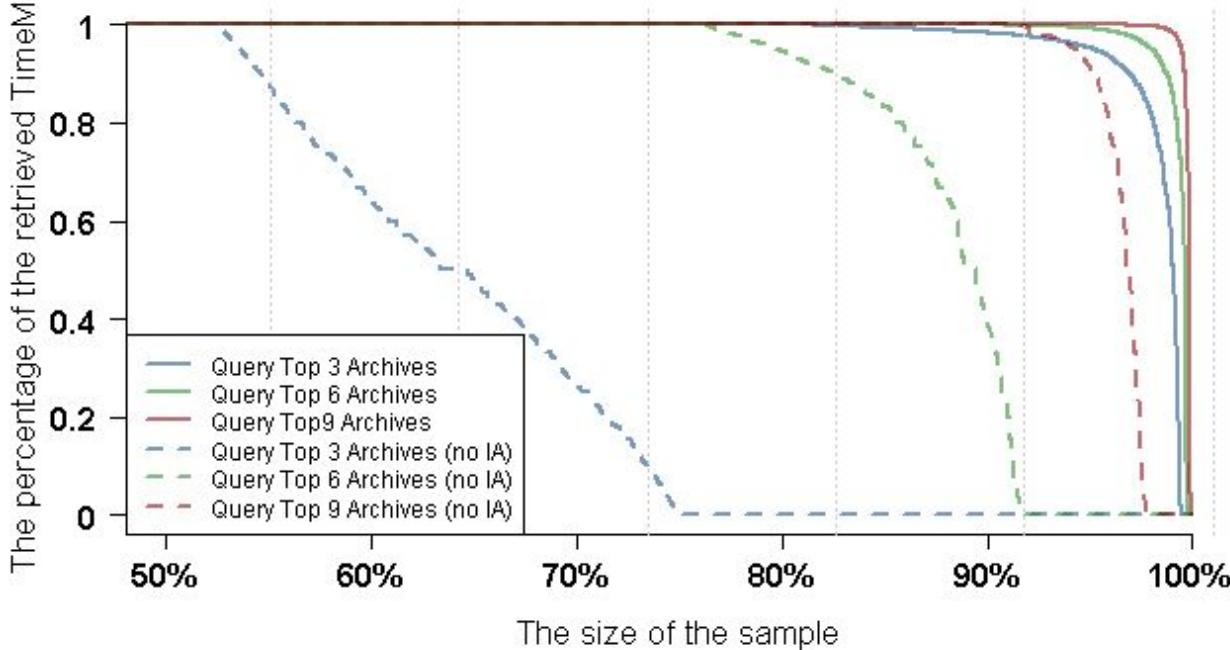
# Other Players

- Free on-demand archiving (page-at-a-time model, not a crawler)
  - [Conifer](#)
    - used to be webrecorder.io
  - [WebCite](#)
    - but no longer accepting new pages
  - [archive.is](#)
  - [freezePage](#)
  - [Perma.cc](#)

# Long Tail of Archives



# Using Only Top-k Archives for URI Lookup Yields Good Results



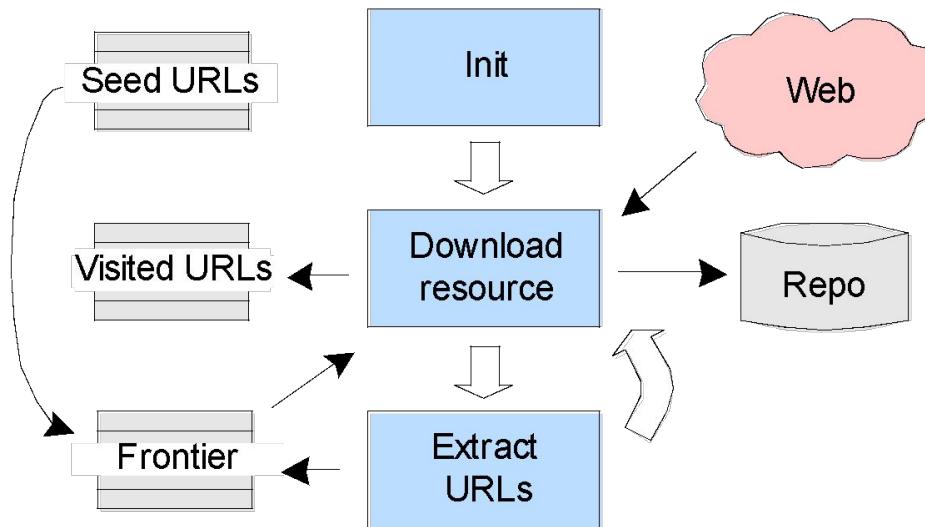
Even when there are 100s of archives, we only need to talk to a few.

see: [Profiling Web Archive Coverage for Top-Level Domain and Content Language](#)

# How is the Web Archived?

# Web Crawler

Web crawlers are used to fetch a page, place all the page's links in a queue, and continue the process for each URL in the queue



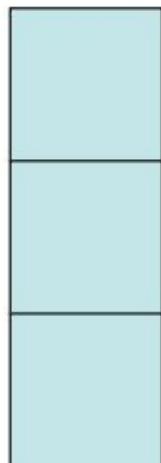
**Figure:** McCown, [Lazy Preservation: Reconstructing Websites from the Web Infrastructure](#), Dissertation, 2007  
See also: [Web crawler \(Wikipedia\)](#)

# Web Crawlers

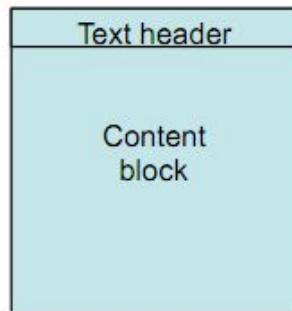
- Wget and HTTrack
  - Simple tools for mirroring a website
  - All crawled URLs are converted into a path and file
    - `http://foo.org/test/` saved as `foo.org/test/index.html`
  - Not designed for large-scale crawling
- Heritrix
  - Built by Internet Archive and Nordic national libraries for larger web archiving tasks
  - Archived content stored in Web ARChive file format ([WARC](#))
  - Uses web interface
  - Can find links in JavaScript, Flash, etc.

# WARC File Organization

W/ARC File



W/ARC Record



*Length, source URI, date, type, ...*

E.g., HTTP response  
headers and *length* bytes  
of HTML, GIF, PDF, ...

.

.

.

Append at will

## Target Websites and Social Media



## WARC ISO 28500 File

```
WARC/1.0
WARC-Type: response
WARC-Record-ID: <urn:uuid:7004e5fd-3f87-f10f-0539-e116845021fe>
WARC-Date: 2011-11-01T18:12:33Z
WARC-Target-URI: http://news.bbc.co.uk/
WARC-Concurrent-To: <urn:uuid:58c7c6ab-458e-008a-cd05-41ccadice188>
X-Hanzo-Page-Id: <bbc.warc>
X-Hanzo-Page-Uri: http://news.bbc.co.uk
X-Remote-Host: 212.58.246.82:80
X-Hanzo-Record-Id: 8750395-c82e-46b4-a251-0f910535a344
WARC-IP-Address: 212.58.246.82
Content-Type: application/http;msgtype=response
Content-Length: 939
WARC-Block-Digest:
sha256:1c5806cbd371f39a25f3bbae6ee30c853b6044423c2294c62
1bf9bb69579f
```

```
HTTP/1.1 301 Moved Permanently
Date: Tue, 01 Nov 2011 18:12:33 GMT
Server: Apache
Set-Cookie: BBC-
UID=44ae0be013563901a7a39e0291faeaa9f756127670a0c1ffd299c6b
4e41f72f60Mozilla/5.0(Firefox/3.6.12;Windows NT 6.1;en-US;rv:12.0) AppleWebKit/534.34(KHTML, like Gecko) Chrome/12.0.742.112 Safari/534.34
X-Original-Content-Length: 234
Keep-Alive: timeout=5, max=693
X-Original-Connection: Keep-Alive
Content-Type: text/html; charset=iso-8859-1
Content-Length: 234
```

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>301 Moved Permanently</title>
</head><body>
<h1>Moved Permanently</h1>
<p>The document has moved <a href="http://www.bbc.co.uk/news/">here</a>.</p>
</body></html>
```

Metadata

Hash

Headers

Original Native Format Content

Image from Hanzo, [ISO 28500 WARC FAQ](#)

# WARC Tools

- warc-tools
  - Python scripts for interacting with WARC files
  - [internetarchive/warctools: Command line tools and libraries for handling and manipulating WARC files \(and HTTP contents\)](#)
- WARCreate & WAIL - *developed at ODU*
  - [WS-DL blog, "WARCreate and WAIL: WARC, Wayback and Heritrix Made Easy"](#)
  - [WS-DL blog, "Electric WAILs and Ham"](#)
  - [WS-DL blog, "Replacing Heritrix with Chrome in WAIL, and the release of node-warc, node-cdxj, and Squidwarc"](#)
- Webrecorder
  - [Webrecorder Tools](#)

Webrecorder | Tools

https://webrecorder.net/tools



# Webrecorder

*Web archiving for all!*

Home | About | **Tools** | Blog | Contributors

The Webrecorder Project has developed many tools to help with web archive capture and replay.

There are **tools for everyone** who are interested in creating or replaying web archives on their own, and more advanced **tools for developers**, which provide tools for software developers to integrate into other workflows and require a bit more technical expertise.

## Tools for Everyone

**ReplayWeb.page**

ReplayWeb.page provides a web archive replay system as a single web site (which also works offline), allowing users to view web archives from anywhere, including local computer or even Google Drive. See the [User guide](#) for more info.

**Webrecorder Desktop**

Webrecorder Desktop is a standalone desktop app, developed as a modification of the Webrecorder (now Conifer) hosted service. It includes support for creating and replaying web archives on your own machine.

**Conifer Hosted Service**

Conifer (formerly <https://webrecorder.io/>) is an integrated hosted service offered by Rhizome, and includes capture, replay and storage capabilities.

## All Webrecorder Tools (Developer and User Tools)

All current Webrecorder tools are listed below. Select one of the categories to further filter this list.

**Take a look at these tools/projects if you are interested in deploying web archiving tools on your, or integrating into other projects.**

# Archive Viewers/Searchers

- [OpenWayback](#)
  - Open source Java implementation of IA's Wayback Machine
  - Search by URL
- [pywb](#)
  - Python implementation of IA's Wayback Machine
- [NutchWAX](#)
  - Full-text search of a web archive
- [Shine](#)
  - [UK WebArchive Shine](#)
  - Web archive search, based on Apache SOLR

# How Much of The Web Is Archived?

# Public Archives, ca. Late 2010 / Early 2011

Three categories of archives

- Internet Archive
- Search engine
- Other archives



WebCite



UK

archiefweb.eu

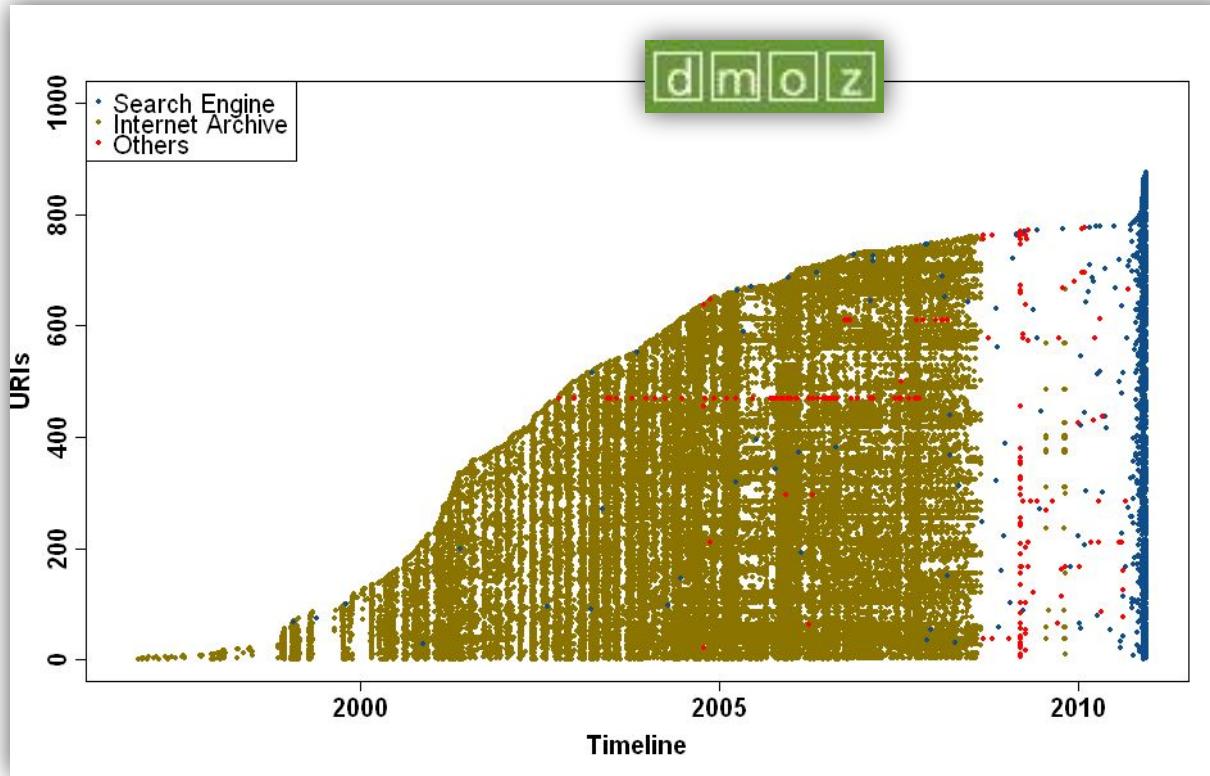


US

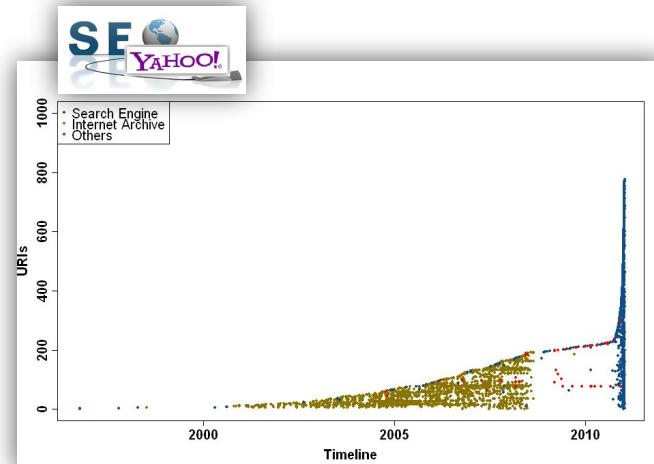
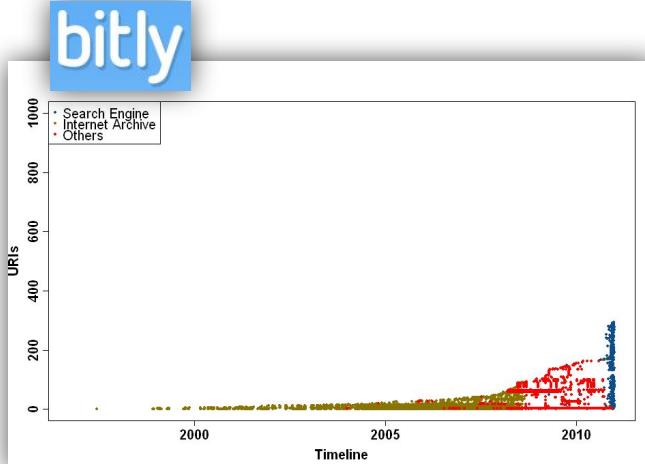
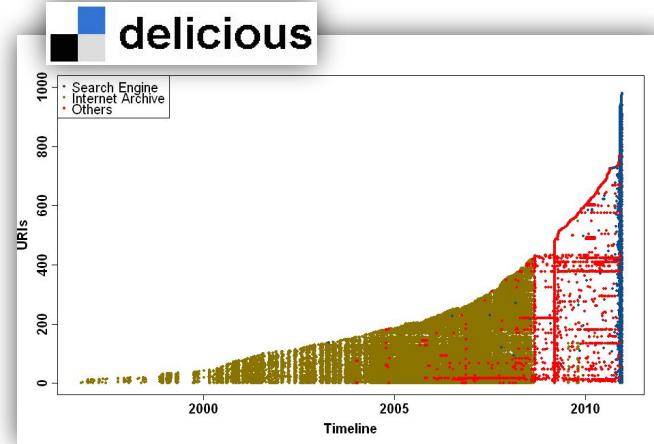
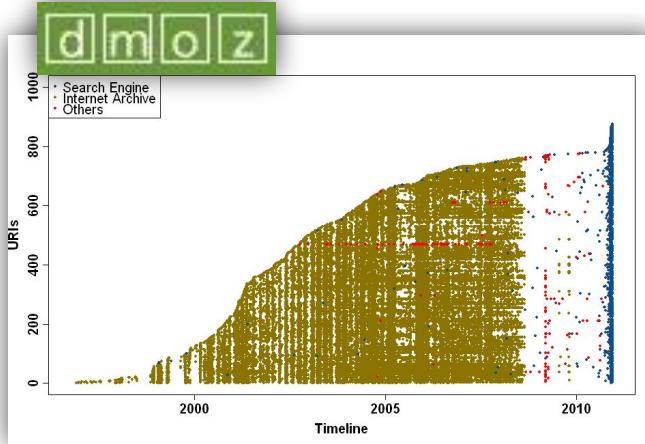


See also: Ainsworth et al., ["How Much of the Web Is Archived?"](#)

# 1000 URIs Ordered by First Observation Date

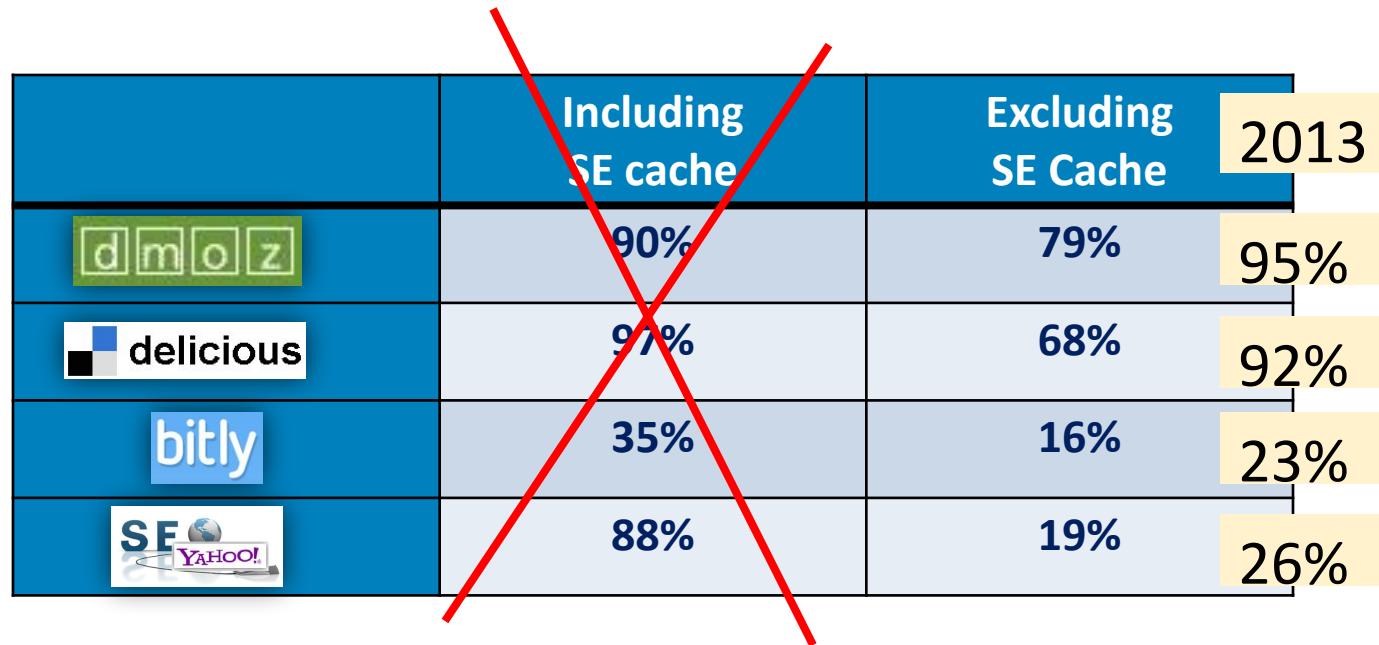


See also: WS-DL blog, "[How Much of the Web Is Archived?](#)"



# How Much of the Web is Archived?

## It Depends on Which Web...



Changes since 2011: no more free SE APIs;  
greatly reduced IA quarantine period; 15+ public web archives

# Web Science: Archiving the Web

## (Part 3 - How Can We Navigate Through Time on the Web?)

CS 432/532

Old Dominion University

*Permission has been granted to use these slides from Frank McCown, Michael L. Nelson, Alexander Nwala, Michele C. Weigle*



This work is licensed under a [Creative Commons](#)  
[Attribution-NonCommercial-ShareAlike 3.0 Unported License](#)

# Memento: Time Travel for the Web



The Memento Team

**Herbert Van de Sompel**

**Michael L. Nelson**

Robert Sanderson

Lyudmila Balakireva

Scott Ainsworth

Harihar Shankar

Memento was partially funded by the  
Library of Congress

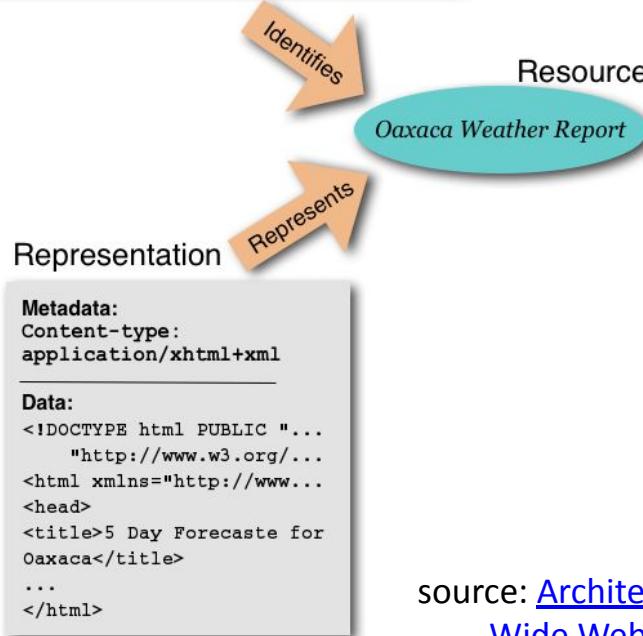
[Memento 101](#) slides

[RFC 7089: HTTP Framework for Time-Based Access to Resource States -- Memento](#), December 2013

# Recall Web Architecture

URI

`http://weather.example.com/oaxaca`



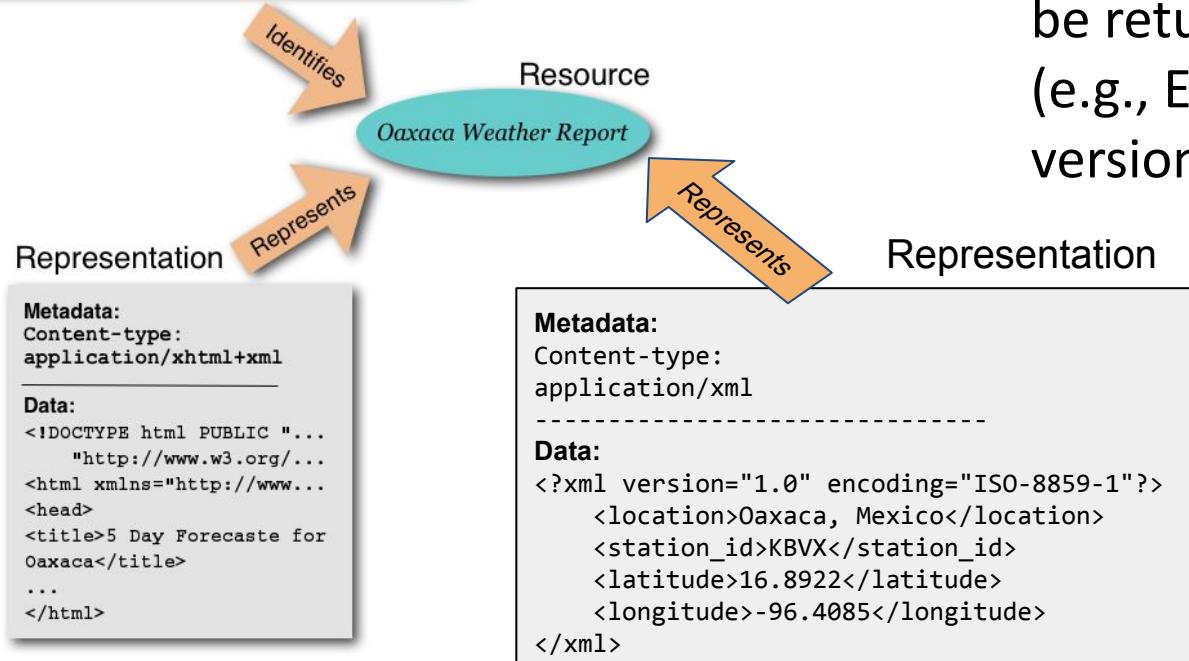
- URIs *identify* Resources
- Representations *represent* Resources
- When URIs are *dereferenced*, they return representations (not resources)

source: [Architecture of the World Wide Web, Volume One](#)

# Content Negotiation

URI

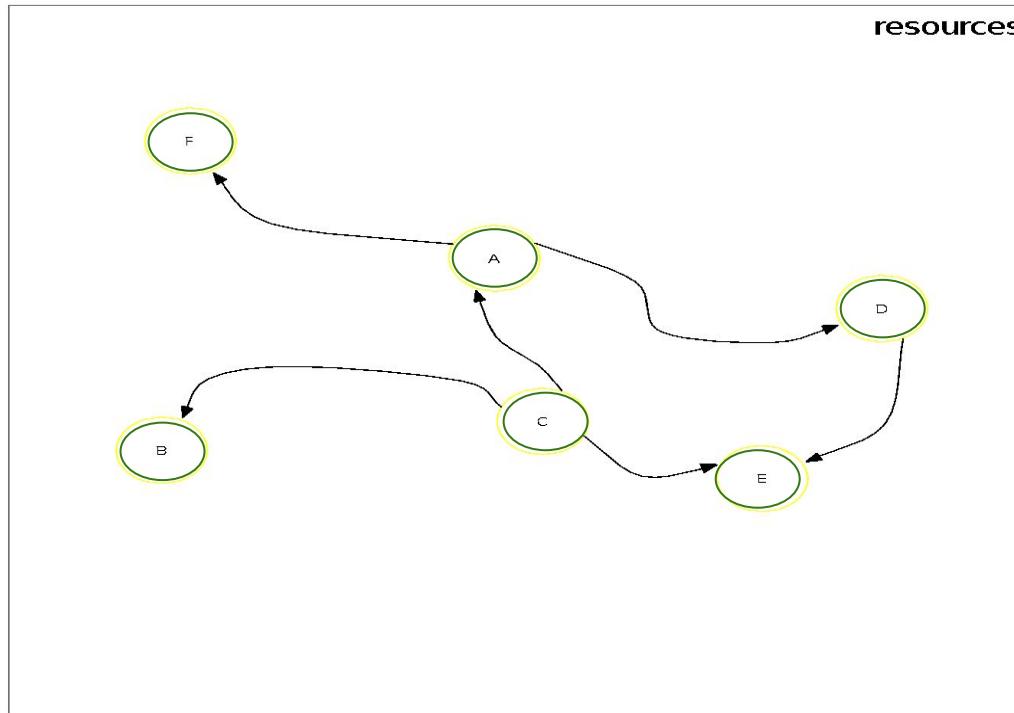
`http://weather.example.com/oaxaca`



- Different representations may be returned for the same URI (e.g., English vs. French version, HTML vs. XML)

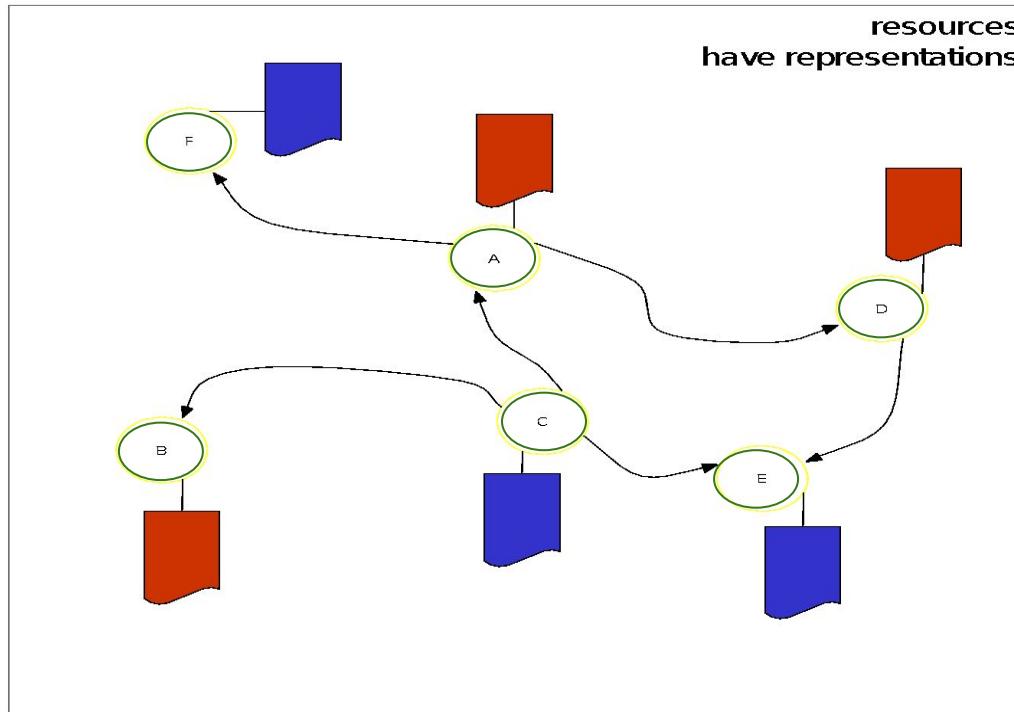
source: [Architecture of the World Wide Web, Volume One](#)

# Resources



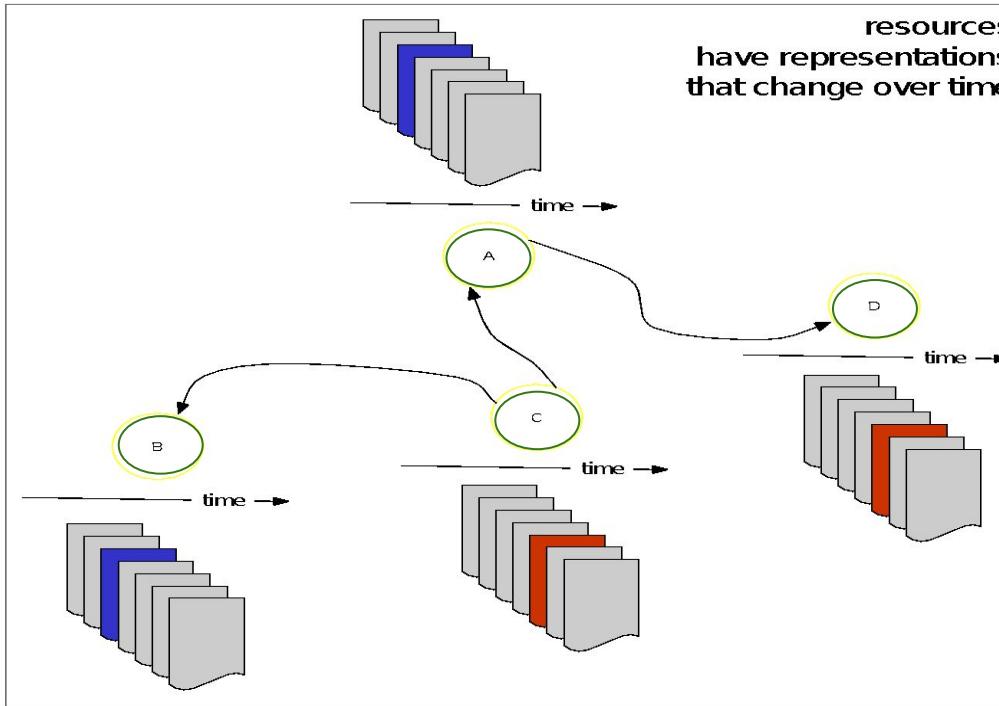
source: [Memento 101](#) slides

# Resources have Representations



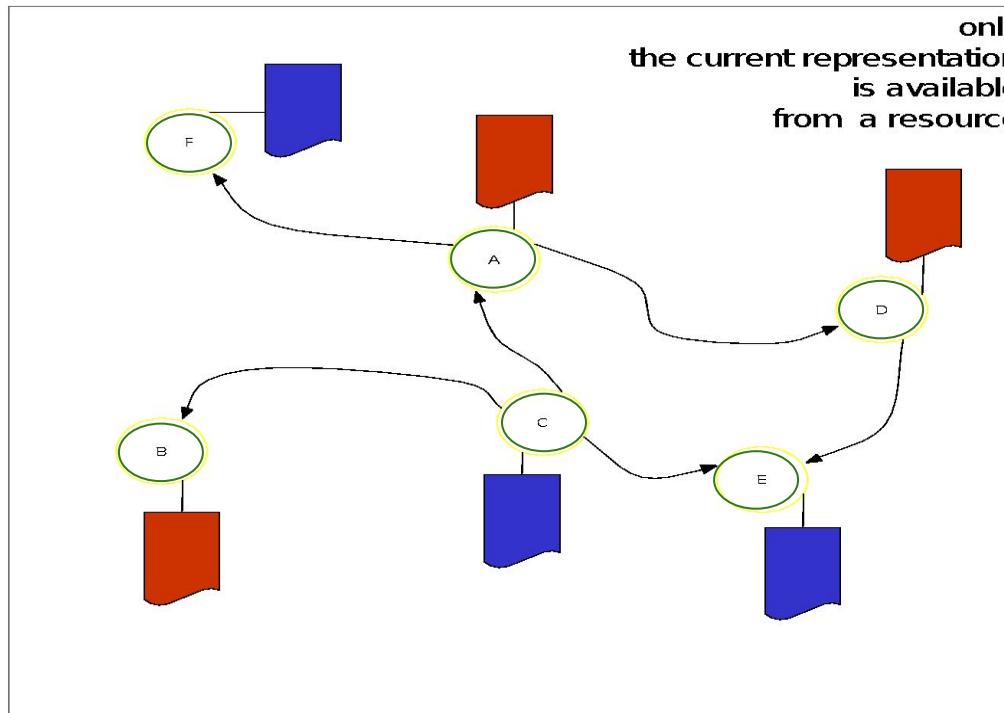
source: [Memento 101](#) slides

# Resources have Representations that Change over Time



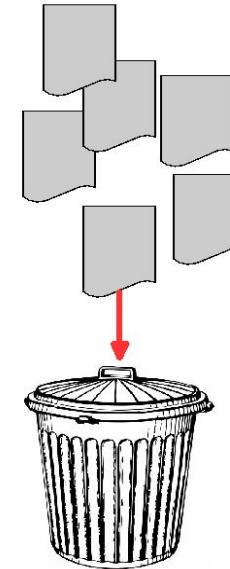
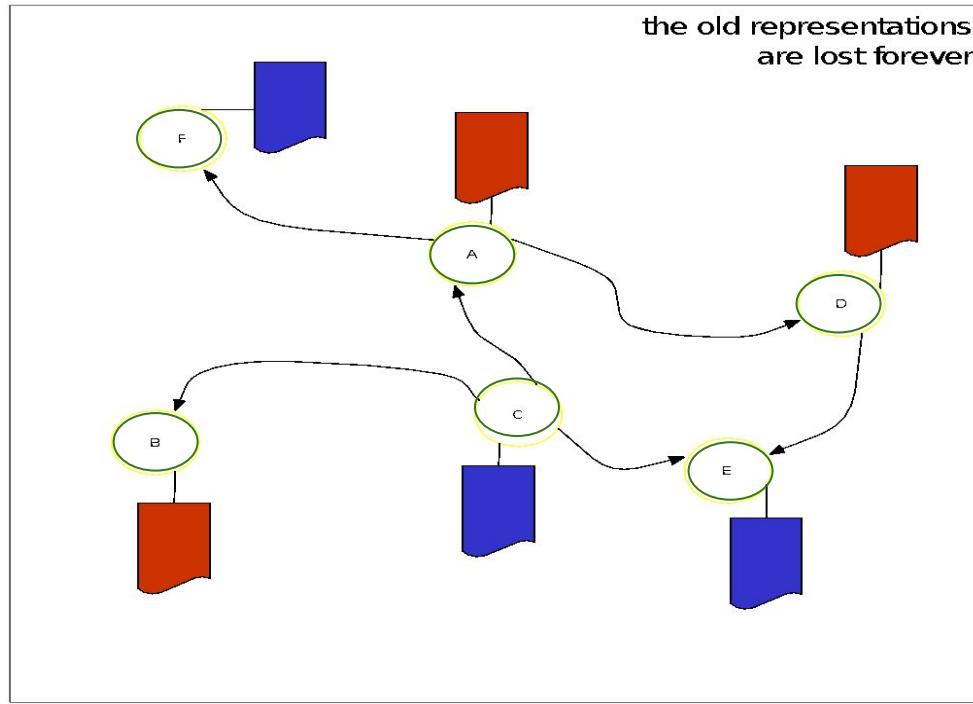
source: [Memento 101](#) slides

# Only the Current Representation is Available from a Resource



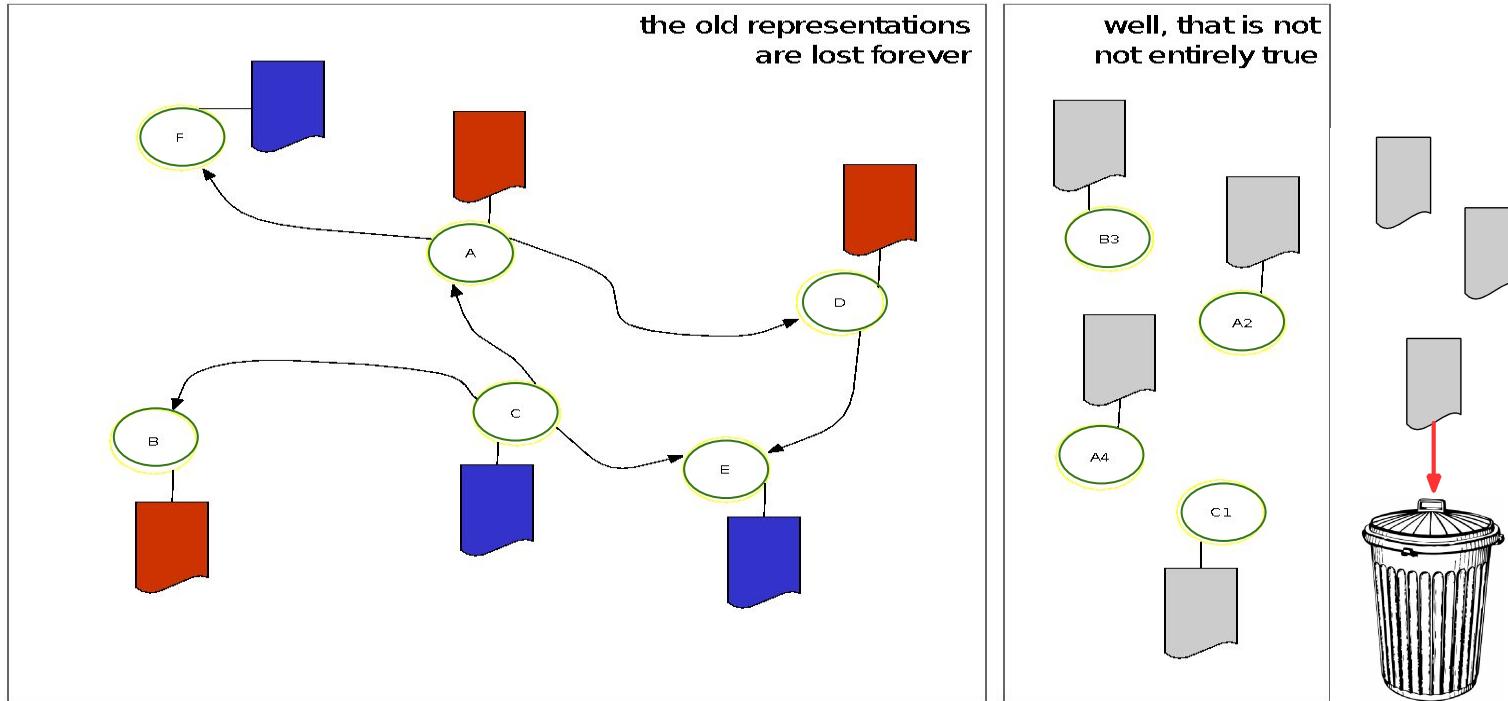
source: [Memento 101](#) slides

# Old Representations are Lost Forever



source: [Memento 101](#) slides

# Archived Resources Exist



source: [Memento 101](#) slides

Sep 11 2001, 20:36:10 UTC

Dec 20 2001, 4:51:00 UTC

# Archived Resources

**CNN.com SPECIAL REPORT**

SEARCH GO Updated: 5:42 p.m. EDT -- 11 September 2001

**AMERICA UNDER ATTACK**  
CNN EXCLUSIVE

**COMPLETE COVERAGE**

- [World Trade Center towers collapse after hit](#)
- [Crash destroys part of Pentagon](#)
- [White House, Capitol evacuated](#)
- [American, United confirm losing planes](#)
- [Bush: 'We'll hunt them down' | Statement](#)
- [U.S. officials: More attacks can't be ruled out](#)
- [Attacks strike financial markets](#)

**U.S. SCENE**

- [FAA grounds all U.S. flights until noon Wednesday](#)
- [U.S. military on 'high alert'](#)
- [Major sports events postponed](#)

**WORLD SCENE**

- [Taliban statement: 'Afghanistan feels your pain'](#)
- [World in shock | Evacuations amid chaos](#)
- [Middle East, Asia condemn attacks](#)

**RESOURCES AND RESCUE**

**FULL STORY »**

**SPECIAL REPORT »**

**VIDEO** Watch video of [second plane crash at World Trade Center](#)  
[More audio/video >>](#)

**PHOTO GALLERY** Aftermath of the attacks in New York and Washington  
[Map of attacks](#)

**CHRONOLOGY** At 8:45 a.m. EDT, day of disaster begins  
[More audio/video >>](#)

**EXTRA INFO** [World Trade Center, Pentagon, 767](#)

**EDITIONS**  
CNN.com Asia  
CNN.com Europe  
set your edition

**CNN TV** show transcripts  
CNN Headline News  
CNN International  
askCNN

**ABOUT US**

The Free Encyclopedia

September 11 attacks

From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by The Cunctator (talk | contribs) at 04:51, 20 December 2001. It may differ significantly from the current revision.  
(diff) ← Previous revision | Current revision (diff) | Newer revision → (diff)

In Memoriam, September 11, 2001

On the morning of September 11, 2001, what might well be the most devastating terrorist attack in the history of the world occurred concurrently in New York City, Washington, D.C. and near Pittsburgh. Four passenger jets were hijacked and then deliberately crashed into the World Trade Center and the Pentagon. Both towers of the World Trade Center subsequently collapsed, and part of the Pentagon was destroyed in the ensuing fire. Casualties are expected to be in the thousands: 265 on the planes; about 3000 people (early estimates ranged as high as 6500 people), including hundreds of firefighters who had rushed in, at the World Trade Center; and 125 at the Pentagon.

Some passengers on the doomed flights were able to make phone calls reporting on events on board. They reported that there were more than one hijacker on each plane and that they took control of the planes using box-cutter knives. It appears that the passengers on the fourth jet tried to overpower the hijackers and that the plane crashed in a sparsely populated area as a result, thereby missing its intended target, which may have been the White House.

The attack had immediate and deep global political effects and economic effects and inspired an international outpouring of memorials and services. The massive undertaking of rescue and recovery, and of providing assistance to the survivors and victims, is ongoing. There will be a great need for donations for a long time.

Though no group has claimed responsibility, the U.S. government immediately launched a full-scale response, stating its intentions to go to war against those responsible. On October 7, a coalition led by the United States launched an attack in Afghanistan. See 2001 U.S. Attack on Afghanistan.

Following the attack, the United States has been on heightened alert for new terrorist attacks. In late September, cases of anthrax started breaking out; although, no connection to the September 11 attack has been found. See 2001 anthrax attack.

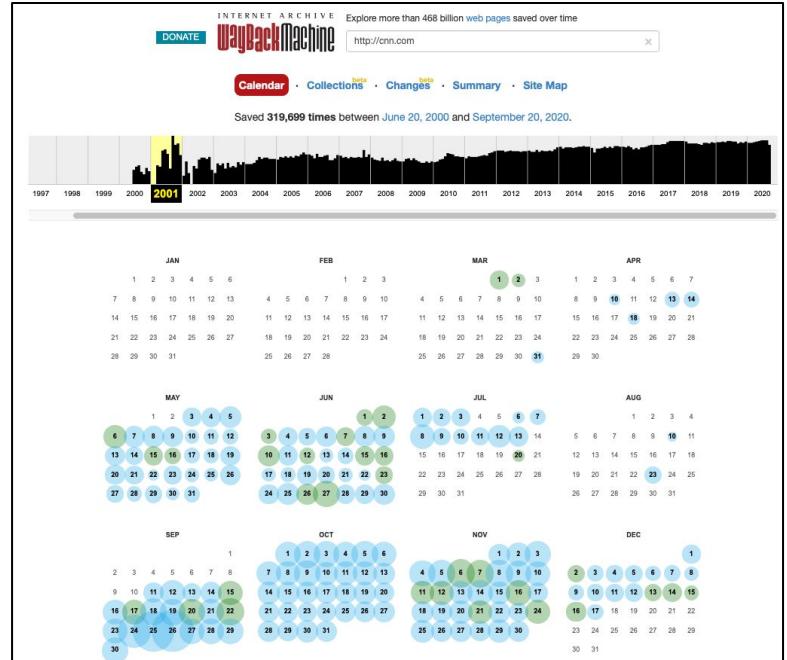
<http://web.archive.org/web/20010911203610/http://www.cnn.com/>  
archived resource for <http://cnn.com>

source: [Memento 101](#) slides

[http://en.wikipedia.org/w/index.php?title=September\\_11\\_attacks&oldid=282333](http://en.wikipedia.org/w/index.php?title=September_11_attacks&oldid=282333)  
archived resource for  
[http://en.wikipedia.org/wiki/September\\_11\\_attacks](http://en.wikipedia.org/wiki/September_11_attacks)

# Finding Archived Resources

The screenshot shows the Wayback Machine homepage. At the top, there's a navigation bar with links for ABOUT, BLOG, PROJECTS, HELP, DONATE, CONTACT, JOBS, VOLUNTEER, and PEOPLE. A search bar is also at the top. Below the header, there's a banner for "INTERNET ARCHIVE Explore more than 468 billion web pages saved over time". A "DONATE" button is visible. The main content area features a grid of thumbnail images representing different archived websites. On the left, there's a "Tools" section with links for "Wayback Machine Availability API", "Chrome Extension", "Firefox Add-on", "Safari Extension", "iOS app", and "Android app". In the center, there's a "Subscription Service" section with a "Save Page Now" button and a form to enter a URL. At the bottom, there's a note about the Wayback Machine being an initiative of the Internet Archive and terms of use.



Go to <http://web.archive.org/> and search  
<http://cnn.com>

On [http://web.archive.org/web/\\*/http://cnn.com](http://web.archive.org/web/*/http://cnn.com), select  
desired datetime

source: [Memento 101 slides \(Wayback Machine images updated Sep 2020\)](#)

# Finding Archived Resources

WIKIPEDIA  
The Free Encyclopedia

navigation

- Main page
- Contents
- Featured content
- Current events
- Random article

search

interaction

- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Cite this page

languages

- Alemanisch
- العربية
- Azərbaycan
- Беларуская

article discussion view source history Try Beta Log In / create account

## September 11 attacks

From Wikipedia, the free encyclopedia

"9/11" redirects here. For the date, see [September 11](#) or [November 9](#). For other uses, see [911 \(disambiguation\)](#).

The **September 11 attacks** (often referred to as **September 11th** or **9/11**) were a series of coordinated suicide attacks by Al-Qaeda upon the United States on September 11, 2001. On that morning, 19 Al-Qaeda terrorists hijacked four commercial passenger jet airliners.<sup>[1][2]</sup> The hijackers intentionally crashed two of the airliners into the Twin Towers of the **World Trade Center** in New York City, killing everyone on board and many others working in the buildings. Both buildings collapsed within two hours, destroying nearby buildings and damaging others. The hijackers crashed a third airplane into the **Pentagon** in Arlington, Virginia, just outside of Washington, D.C. The fourth plane crashed into a field near Shanksville in rural Pennsylvania, after some of its passengers and flight crew attempted to retake control of the plane, which the hijackers had redirected toward Washington, D.C. There were no survivors from any of the flights.

In total 2,993 people, including the hijackers, died in the attacks.<sup>[3]</sup> The overwhelming majority of casualties were civilians, including nationals of over 90 countries. In addition, the death of at least one person from lung disease was ruled by a medical examiner to be a result of exposure to dust from the World Trade Center's collapse.<sup>[4]</sup>

The United States responded to the attacks by launching a "War on Terrorism", invading Afghanistan to depose the Taliban, who had harbored al-Qaeda terrorists, and enacting the **USA PATRIOT Act**. Many other countries also strengthened their anti-terrorism legislation and expanded law enforcement powers. Some American stock exchanges stayed closed for the rest of the week following the attack, and posted enormous



Twin towers of the World Trade Center burning.

**Location:** New York City; Arlington County, Virginia; and near Shanksville, Pennsylvania.

**Date:** Tuesday, September 11, 2001  
8:46 am – 10:28 am (UTC-4)

**Attack type:** Aircraft hijacking, Mass murder, Suicide attack

**Death(s):** 2,993 (including 19 hijackers)

**Injured:** 6,000+

**Perpetrator(s):** al-Qaeda led by Osama bin Laden, see also Responsibility and Hijackers.



WIKIPEDIA  
The Free Encyclopedia

navigation

- Main page
- Contents
- Featured content
- Current events
- Random article

search

interaction

- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox

- What links here
- Related changes
- Upload file
- Special pages

article discussion view source history Try Beta Log In / create account

## Revision history of September 11 attacks

From Wikipedia, the free encyclopedia

View logs for this page

Browse history

From year (and earlier): From month (and earlier): all Tag filter: Go

For any version listed below, click on its date to view it.

For more help, see [Help:Page history](#) and [Help:Edit summary](#).

External tools: [Revision history statistics](#) • [Revision history search](#) • [Number of watchers](#) • [Page view statistics](#)

(cur) = difference from current version, (prev) = difference from preceding version,  
m = minor edit, → = section edit, ← = automatic edit summary  
(latest | earliest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)  
Compare selected revisions

- (cur) (prev) 18:20, 9 November 2009 InnerParty (talk | contribs) (133,903 bytes) (→Conspiracy theories: adding 9/11 opinion polls)
- (cur) (prev) 23:01, 8 November 2009 InnerParty (talk | contribs) (133,650 bytes) (→9/11 Commission: extrapolating)
- (cur) (prev) 22:27, 8 November 2009 WordyGirl90 (talk | contribs) (133,608 bytes) (→Planning of the attacks: wikilink needed, methinks)
- (cur) (prev) 22:20, 8 November 2009 WordyGirl90 (talk | contribs) (133,604 bytes) (→Attackers and their motivation: rmv excess wikilinking)
- (cur) (prev) 02:09, 7 November 2009 DaleEastman (talk | contribs) (133,612 bytes)
- (cur) (prev) 12:07, 6 November 2009 LawBot (talk | contribs) m (133,541 bytes) (Bot: Removed unnecessary and invalid <span>'s.)
- (cur) (prev) 14:32, 3 November 2009 Ohnoitsjamie (talk | contribs) (133,548 bytes) (restored sourced material from cbsnews and the Guardian)
- (cur) (prev) 11:25, 3 November 2009 InnerParty (talk | contribs) m (132,453 bytes) (Undid revision 323549441 by The Original Wildbear (talk) I have no idea what are you talking about...)
- (cur) (prev) 21:36, 2 November 2009 The Original Wildbear (talk | contribs) (133,548 bytes) (Undid revision 323502267 by InnerParty) Exactly; it IS outlandish. We should represent history as it is; not put lipstick on a pig)

Go to

[http://en.wikipedia.org/wiki/September\\_11\\_attacks](http://en.wikipedia.org/wiki/September_11_attacks)  
and click History

Browse History

source: [Memento 101](#) slides

Dec 20 2001, 4:51:00 UTC

current

# Navigating Archived Resources

Try Beta Log in / create account

September 11 attacks

From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by The Cunctator (talk | contribs) at 04:51, 20 December 2001. It may differ significantly from the current revision.

(diff) ← Previous revision | Current revision (diff) | Newer revision → (diff)

In Memoriam, September 11, 2001

On the morning of September 11, 2001, what might well be the most devastating terrorist attack in the history of the world occurred concurrently in New York City, Washington, D.C. and near Pittsburgh. Four passenger jets were of the World Trade Center, one of which was hijacked by terrorists and crashed into the Pentagon. Both towers were destroyed in the ensuing fire. Casualties ranged as high as 6000 people, including hundreds of firefighters who had rushed in, at the World Trade Center; and 125 at the Pentagon.

## Pentagon

Some passengers on the doomed flights were able to make phone calls reporting on events on board. They reported that there were more than one hijacker on each plane and that they took control of the planes using box-cutter knives. It appears that the passengers on the fourth jet tried to overpower the hijackers and that the plane crashed in a sparsely populated area as a result, thereby missing its intended target, which may have been the White House.

The attack had immediate and deep global political effects and economic effects and inspired an international outpouring of memorials and services. The massive undertaking of rescue and recovery, and of providing assistance to the survivors and victims, is ongoing. There will be a great need for donations for a long time.

Though no group has claimed responsibility, the U.S. government immediately launched a full-scale response, stating its intentions to go to war against those responsible. On October 7, a coalition led by the United States launched an attack in Afghanistan. See 2001 U.S. Attack on Afghanistan.

Following the attack, the United States has been on heightened alert for new terrorist attacks. In late September, cases of anthrax started breaking out; although, no connection to the September 11 attack has been found. See 2001 anthrax attack.

Try Beta Log in / create account

The Pentagon

From Wikipedia, the free encyclopedia

Coordinates 38°52'15.56"N 77°3'21.46"W

This article is about the United States military building. For the shape, see [Pentagon](#).

The Pentagon is the headquarters of the United States Department of Defense, located in Arlington, Virginia. As a symbol of the U.S. military, "the Pentagon" is often used metonymically to refer to the Department of Defense rather than the building itself.

Designed by the American architect George Bergstrom (1876 – 1955), and built by Philadelphia, Pennsylvania, general contractor John McShain, the building was dedicated on January 15, 1943, after ground was broken for construction on September 11, 1941. General Brethon Somervell provided the major motive power behind the project.<sup>[1]</sup> Colonel Leslie Groves and Major Clarence Renshaw were responsible for overseeing the project for the Army.

The Pentagon is the world's largest office building by floor area<sup>[2][3]</sup>, by a margin of about 35,580 square metres (383,000 sq ft). Approximately 23,000 military and civilian employees<sup>[4]</sup> and about 3,000 non-defense support personnel work in the Pentagon. It has five sides, five floors above ground (plus two basement levels), and five ring corridors per floor with a total of 17.5 miles (28.2 km)<sup>[5]</sup> of corridors. The Pentagon includes a five-acre (20,000 m<sup>2</sup>) central plaza, which is shaped like a pentagon and informally known as "ground zero", a nickname originating during the Cold War and based on the presumption that the Soviet Union would target one or more nuclear missiles at this central location.

On September 11, 2001, hijacked American Airlines Flight 77 was crashed into the western side of the

interaction

- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Cite this page

languages

- العربية
- Asturianu
- Български
- Bosanski
- Български
- Català
- Čeština

Type Headquarters

Coordinates 38°52'15.56"N 77°3'21.46"W

Built 1943 (broke ground on September 11, 1941)

Height Five floors above ground and Two floors below ground

In use 1943–present

Southwest view of the Pentagon with the Potomac River and Washington Monument in background.





[http://en.wikipedia.org/w/index.php?title=September\\_11\\_attacks&oldid=282333](http://en.wikipedia.org/w/index.php?title=September_11_attacks&oldid=282333)  
archived resource for [http://en.wikipedia.org/wiki/September\\_11\\_attacks](http://en.wikipedia.org/wiki/September_11_attacks)

[http://en.wikipedia.org/wiki/The\\_Pentagon](http://en.wikipedia.org/wiki/The_Pentagon)

source: [Memento 101](#) slides

Sep 11 2001, 20:36:10 UTC

Sep 11 2001, 21:38:55 UTC

# Navigating Archived Resources

**CNN.com SPECIAL REPORT**

Updated: 5:42 p.m. EDT -- 11 September 2001

**AMERICA UNDER ATTACK**

**CNN EXCLUSIVE**

**SPACE**

**Terrorists strike New York, D.C.**

At 8:45 a.m. EDT, the first of two airliners crashed into the World Trade Center, opening a horrifying and apparently coordinated terrorist attack on the United States, which saw the collapse of the two 110-story towers into surrounding Manhattan streets and a later attack on the Pentagon.

**FULL STORY »**

**SPECIAL REPORT »**

**VIDEO**

Watch video of second plane crash at World Trade Center

**PHOTO GALLERY**

Aftermath of the attacks in New York and Washington

**CHRONOLOGY**

At 8:45 a.m. EDT, day of disaster begins

**EXTRA INFO**

World Trade Center, Pentagon, 767  
Map of attacks

**COMPLETE COVERAGE**

- [World Trade Center towers collapse after hit](#)
- [Crash destroys part of Pentagon](#)
- [White House, Capitol evacuated](#)
- [American, United confirm losing planes](#)
- [Bush: 'We'll hunt them down! I Statement'](#)
- [U.S. officials: More attacks can't be ruled out](#)
- [Attacks strike financial markets](#)

**U.S. SCENE**

- [FAA grounds all U.S. flights until noon Wednesday](#)
- [U.S. military on 'high alert'](#)
- [Major sports events postponed](#)

**WORLD SCENE**

- [Taliban statement: 'Afghanistan feels your pain'](#)
- [World in shock | Evacuations amid chaos](#)
- [Middle East, Asia condemn attacks](#)

**RESOURCES AND RESCUE**

[Giuliani: 1,500 'walking wounded' in New York](#)

[Emergency contact information](#)

**SEARCH GO**

**MAIN PAGE**

**WORLD**

**U.S.**

**WEATHER**

**BUSINESS**

**SPACE**

**HEALTH**

**ENTERTAINMENT**

**TRAVEL**

**EDUCATION**

**CAREER**

**IN-DEPTH**

**QUICK NAVS**

**MULTIMEDIA**

**E-MAIL SERVICES**

**CNN ON PDA**

**ABOUT US**

**CNN TV**

**BEST OF CNN**

**CNN Headline News**

**CNN International**

**askCNN**

**EDITIONS**

**CNN.com Asia**

**CNN.com Europe**

**Get Your Edition**

**CNN.com / SPACE**

Updated: 04:56 PM EDT 20:56 GMT -- 11 Sep 2001

**Cool star chills stellar theories**

- [Cool star chills stellar theories](#)
- [Sleuths try to unmask new spy satellites](#)
- [Probe limps to close encounter with comet](#)
- [Report: Plane tracking satellites can be jammed](#)

**Cool star chills stellar theories**

A star besieged by unbelievably hot waves of nuclear explosions is ten times cooler than expected, forcing astronomers to rethink their ideas about stellar physics.

**FULL STORY »**

**ALPHA EXTRA**

Space correspondent Miles O'Brien on the space station tour

**SCENES FROM SPACE**

A NASA telescope is marking a milestone

**RESOURCES:**

**SPACE AGENCY NEWS**

The latest news releases from worldwide space agencies.

**SPACE CALENDARS**

Calendars of upcoming launches and other events from worldwide space agencies.

**HIT THE LINKS**

**INTERACTIVE LIBRARY**

Your window to the universe - launch video, 360-degree photo tours and more.

**NASA WEBCAM**

Live pictures of NASA projects.

**SPACE PROBES**

A guide to active unmanned missions.

**LAUNCH IT**

**PROBE LIST**

**MORE NEWS**

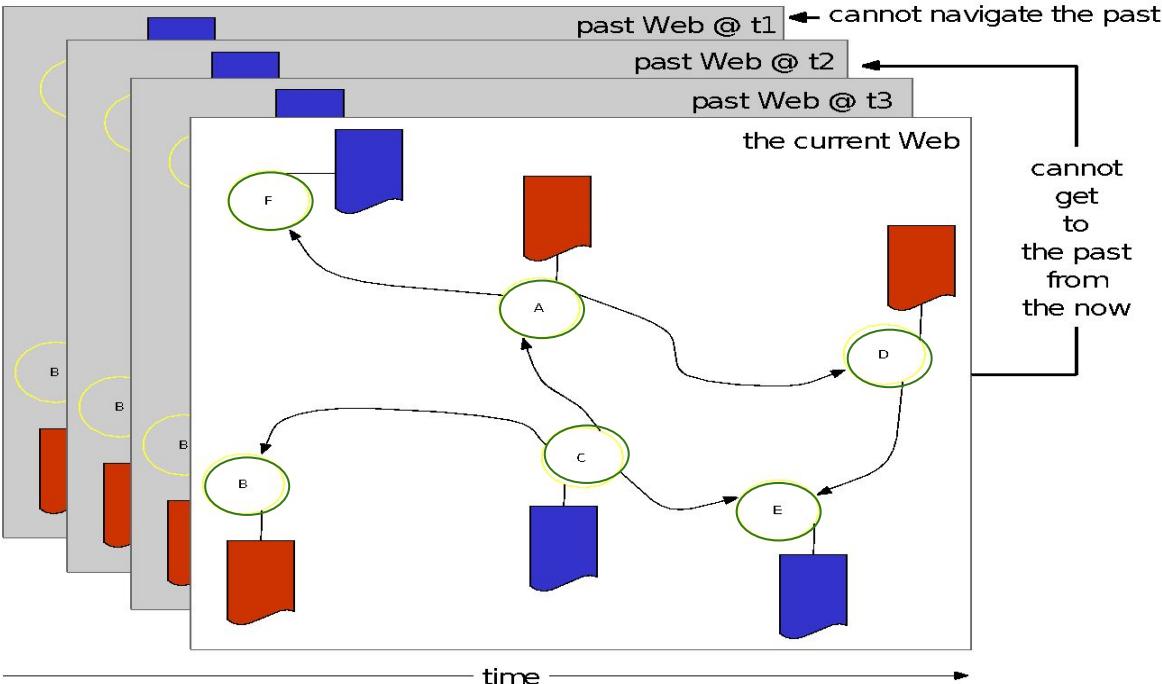
[CNN.com archive of 'Antennae' pillar](#)

<http://web.archive.org/web/20010911203610/http://www.cnn.com/>  
archived resource for <http://cnn.com>

<http://web.archive.org/web/20010911213855/www.cnn.com/TECH/space/>

source: [Memento 101](#) slides

# By Default, Current and Past Web are Not Integrated



- Current and Past Web based on same technology.
- But, going from Current to Past Web is a matter of (manual) discovery.
- Memento makes going from Current to Past Web a (HTTP) protocol matter.
- Memento integrates the Current and Past Web.

source: [Memento 101](#) slides

# The Past Links to the Present...

The screenshot shows a web browser window with the title "WebCite query result". The address bar contains the URL <http://webcitation.org/query?id=1257028234035091>, which is highlighted with a red box. Below the address bar, the "Requested Date" is set to 1996, and the "Cached" date is 2009-10-31 18:30:35. A large red box highlights the "Showing WebCite for URL: http://www.digitalpreservation.gov/" message and the "Permalink to this cache" link. The main content area displays the "Digital Preservation" page from the Library of Congress. The URL in the address bar is also displayed as a large white box on the right side of the image. A blue callout box on the right contains the text: "explicit HTML link; no HTTP links; opaque URI".

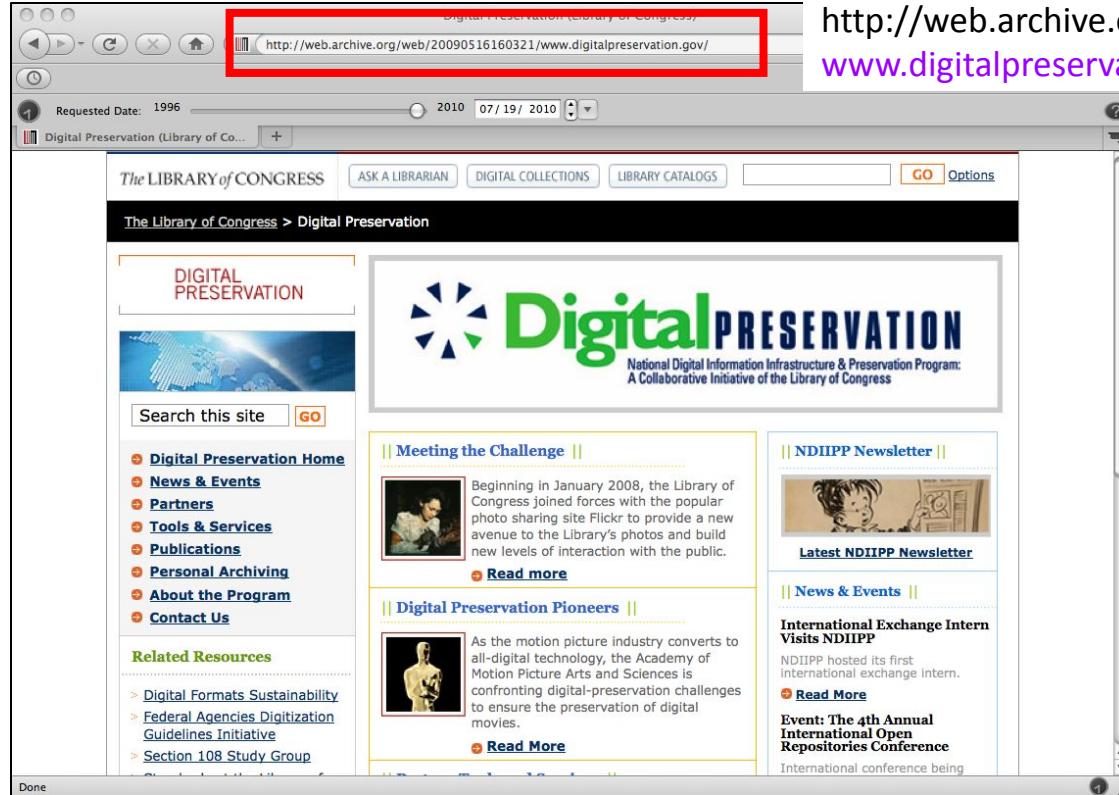
http://webcitation.org/query?id=1257028234035091

Showing WebCite for URL: <http://www.digitalpreservation.gov/> [Permalink to this cache]

Cached: 2009-10-31 18:30:35

explicit HTML link;  
no HTTP links;  
opaque URI

# The Past Links to the Present...



[http://web.archive.org/web/20090516160321/  
www.digitalpreservation.gov/](http://web.archive.org/web/20090516160321/www.digitalpreservation.gov/)

no HTML links;  
no HTTP links;  
implicit from URI

# But the Present Does Not Link to the Past

Digital Preservation (Library of Congress) http://www.digitalpreservation.gov/ Requested Date: 1996 2010 07/19/2010 Digital Preservation (Library of Congress)

The Library of Congress > Digital Preservation

DIGITAL PRESERVATION

Search this site GO

DIGITAL PRESERVATION National Digital Information Infrastructure & Preservation Program: A Collaborative Initiative of the Library of Congress

On the Leading Edge || The Digital Lives project reports that libraries need to be more aggressive in helping the general public save their personal digital possessions. Read this Article »

News & Events || New LC podcast on preserving digital video  
Audiovisual archivist Linda Tadic offers practical solutions for preserving digital video. Read this Article »

Meeting the Challenge || The Metadata for Digital Content group is working to ensure that consistent, rich metadata are available for innovative Library of Congress web services. Read this Article »

Future for Our Digital Memory || The NCCD issued a new report, A Future for Our Digital Memory. Read this Article »

Digital Preservation Pioneers || Julie Sweetkind-Singer says the growth of geospatial technology in our daily lives is just beginning and it presents unique challenges for librarians and digital preservationists. Read this Article »

Library Perspectives on Scientific Data Stewardship || A National Research Council Board of Research Data and Information symposium shared some perspectives on the topic. Read this Article »

Data Conservancy || Recap of Sayeed Choudhury's talk about Data Conservancy at the

Done

no hints in HTML,  
HTTP, or URI

```
% telnet www.digitalpreservation.gov 80
Trying 140.147.249.7...
Connected to www.digitalpreservation.gov.
Escape character is '^['.
HEAD / HTTP/1.1
Host: www.digitalpreservation.gov
Connection: close

HTTP/1.1 200 OK
Date: Mon, 19 Jul 2010 21:36:00 GMT
Server: Apache
Accept-Ranges: bytes
Connection: close
Content-Type: text/html

Connection closed by foreign host.
```

# Linking the Past and the Present

- Codify existing methods to create linkage from the past to the present
  - easy: an archived version knows for which URI it is an archived version
- Create a linkage from the present to the past
  - hard: solve with a level of indirection from present to past

# Memento Terminology

- URI-R: Original Resource Identifier



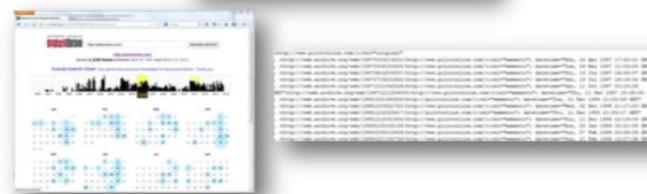
Page on the live web

- URI-M: memento Identifier



Archived version of a page

- URI-T: TimeMap Identifier



List of archived pages

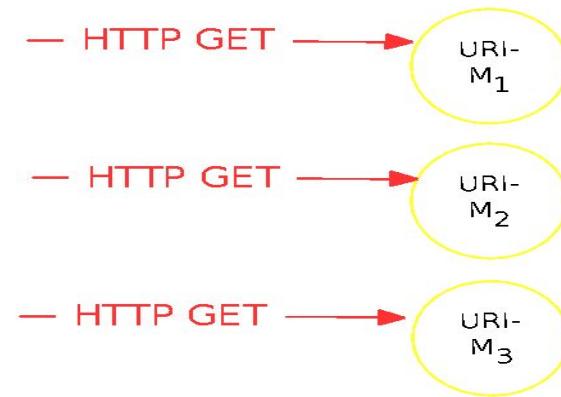
source: Brunelle, “[Scripts in a Frame: A Two-Tiered Approach for Archiving Deferred Representations](#)”, Dissertation Defense, February 5, 2016.

# The Web without a Time Dimension



Original  
resource

# The Web without a Time Dimension



Archived  
resources

# The Web without a Time Dimension

Need to use a different URI to access archived versions of a resource and its current version



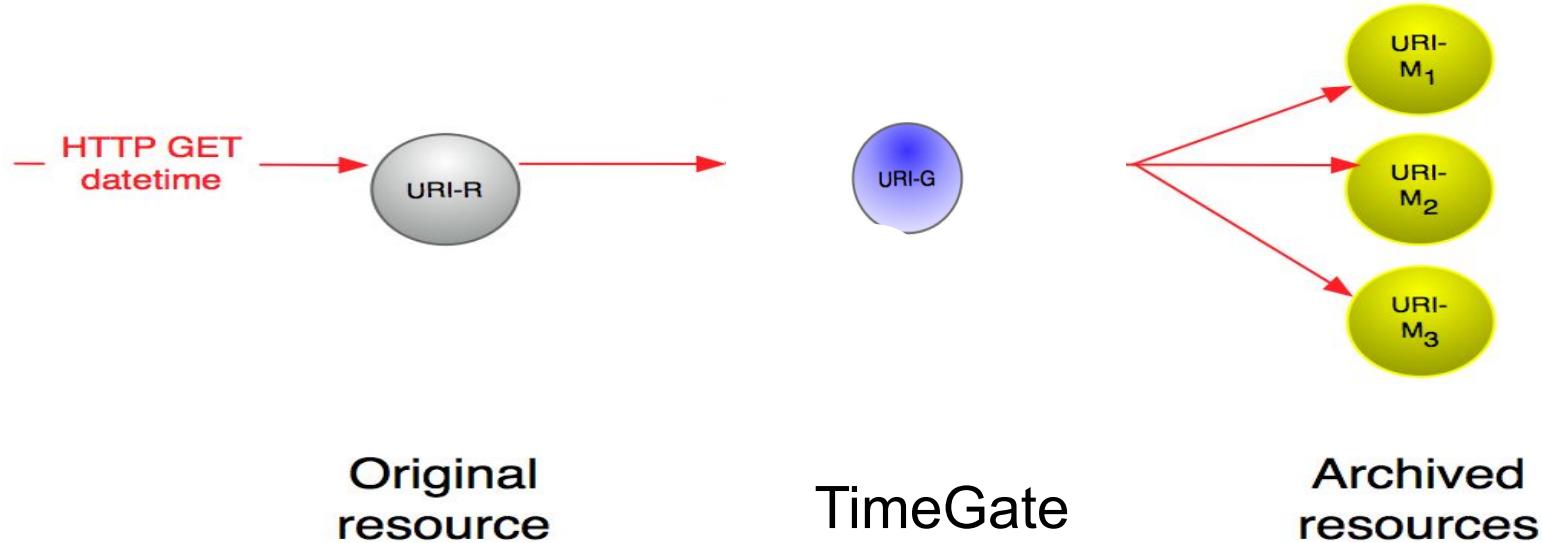
# The Web with Time Dimension added by Memento

In Memento: use URI of the current version to access archived versions, but qualify it with datetime

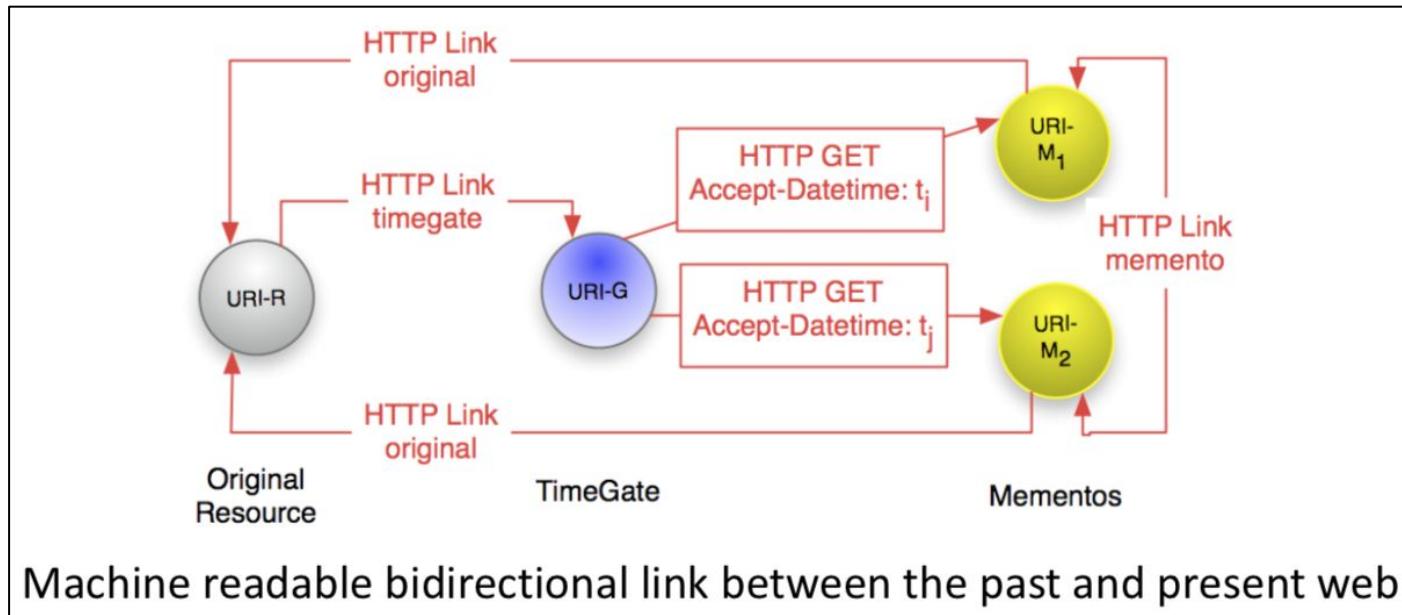


# The Web with Time Dimension added by Memento

... and arrive at an archived version via level of indirection

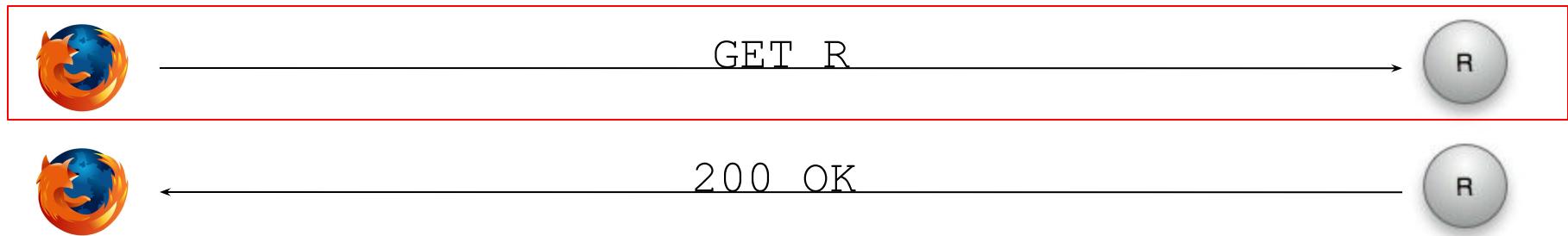


# Memento Framework

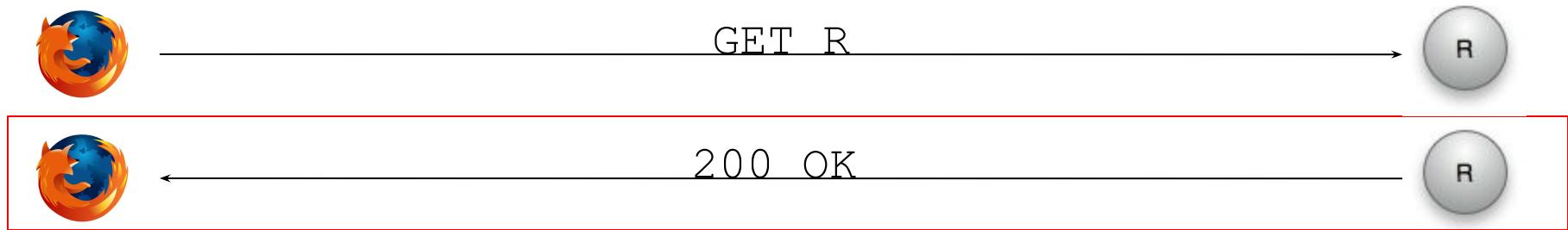


source: Brunelle, “[Scripts in a Frame: A Two-Tiered Approach for Archiving Deferred Representations](#)”, Dissertation Defense, February 5, 2016.

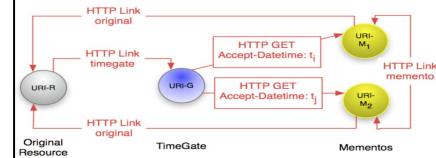
# Normal HTTP Flow



# Normal HTTP Flow



# Memento HTTP Flow



HEAD R, Accept-Datetime



200, Link→G



GET G, Accept-Datetime



302→M, Vary, Link→R, B, M



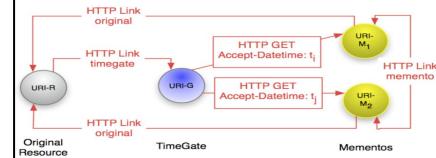
GET M, Accept-Datetime



200, Memento-Datetime, Link→R, B, M



# Memento HTTP Flow



HEAD R, Accept-Datetime



200, Link→G



GET G, Accept-Datetime



302→M, Vary, Link→R, B, M



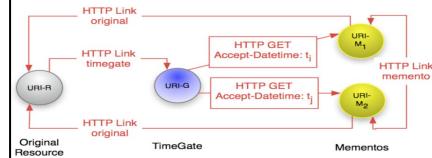
GET M, Accept-Datetime



200, Memento-Datetime, Link→R, B, M



# Memento HTTP Flow



HEAD R, Accept-Datetime



200, Link→G



GET G, Accept-Datetime



302→M, Vary, Link→R, B, M



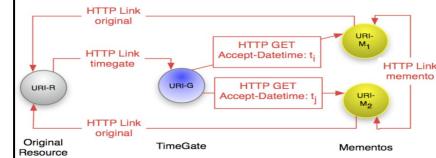
GET M, Accept-Datetime



200, Memento-Datetime, Link→R, B, M



# Memento HTTP Flow



HEAD R, Accept-Datetime



200, Link→G



GET G, Accept-Datetime



302→M, Vary, Link→R, B, M



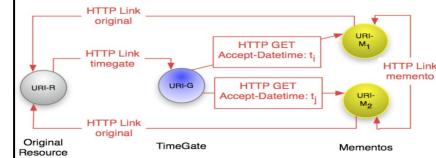
GET M, Accept-Datetime



200, Memento-Datetime, Link→R, B, M



# Memento HTTP Flow



HEAD R, Accept-Datetime



200, Link→G



GET G, Accept-Datetime



302→M, Vary, Link→R, B, M



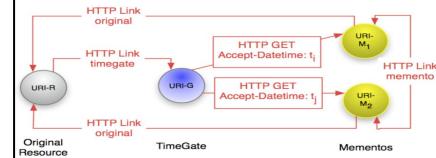
GET M, Accept-Datetime



200, Memento-Datetime, Link→R, B, M



# Memento HTTP Flow



HEAD R, Accept-Datetime



200, Link→G



GET G, Accept-Datetime



302→M, Vary, Link→R, B, M



GET M, Accept-Datetime



200, Memento-Datetime, Link→R, B, M



# Hints from the live web server

```
HEAD /demo-special/index.php/Daenerys_Targaryen HTTP/1.1
User-Agent: curl/7.19.7 (x86_64-redhat-linux-gnu) libcurl/7.19.7 NSS/3.14.3.0 zlib/1.2.3 libidn/1.18 libssh2/1.4.2
Host: localhost:8080
Accept: /*
Accept-Datetime: Thu, 01 Jan 2012 00:00:00 GMT

HTTP/1.1 200 OK
Date: Mon, 14 Apr 2014 22:35:17 GMT
Server: Apache/2.2.15 (Red Hat)
X-Powered-By: PHP/5.3.3
X-Content-Type-Options: nosniff
Link: <http://ws-dl-10.cs.odu.edu/demo-special/index.php/Daenerys_Targaryen>; rel="original latest-version",
<http://ws-dl-10.cs.odu.edu/demo-special/index.php/Special:TimeGate/Daenerys_Targaryen>; rel="timegate",
<http://ws-dl-10.cs.odu.edu/demo-special/index.php/Special:TimeMap/Daenerys_Targaryen>; rel="timemap";
type="application/link-format"
Content-language: en
Vary: Accept-Encoding,Accept-Datetime,Cookie
Cache-Control: s-maxage=18000, must-revalidate, max-age=0
Last-Modified: Tue, 08 Apr 2014 00:20:01 GMT
Connection: close
Content-Type: text/html; charset=UTF-8
```

WS-DL blog, "[TimeGate Design Options For MediaWiki](#)"

# Aggregator TimeMap - Link Format

```
$ curl -i http://memgator.cs.odu.edu/timemap/link/http://www.cs.odu.edu/~mln/
HTTP/1.1 200 OK
Access-Control-Allow-Origin: *
Access-Control-Expose-Headers: Link, Location, X-Memento-Count, X-Generator
Content-Type: application/link-format
Date: Sun, 02 Feb 2020 20:10:43 GMT
X-Generator: MemGator:1.0-rc7
X-Memento-Count: 633
Transfer-Encoding: chunked

<http://www.cs.odu.edu/~mln/>; rel="original",
<https://memgator.cs.odu.edu/timemap/link/http://www.cs.odu.edu/~mln/>; rel="self"; type="application/link-format",
<http://web.archive.org/web/20020526053005/http://www.cs.odu.edu:80/~mln/>; rel="first memento"; datetime="Sun, 26 May 2002
05:30:05 GMT",
<http://web.archive.org/web/20020528083321/http://www.cs.odu.edu:80/~mln/>; rel="memento"; datetime="Tue, 28 May 2002
08:33:21 GMT",
<http://web.archive.org/web/20020720051448/http://www.cs.odu.edu:80/~mln>; rel="memento"; datetime="Sat, 20 Jul 2002
05:14:48 GMT",
<http://web.archive.org/web/20020724230216/http://www.cs.odu.edu:80/~mln/>; rel="memento"; datetime="Wed, 24 Jul 2002
23:02:16 GMT",
<http://web.archive.org/web/20020811035033/http://www.cs.odu.edu:80/~mln>; rel="memento"; datetime="Sun, 11 Aug 2002
03:50:33 GMT",
<http://web.archive.org/web/20021005054529/http://www.cs.odu.edu:80/~mln/>; rel="memento"; datetime="Sat, 05 Oct 2002
05:45:29 GMT",
...
```

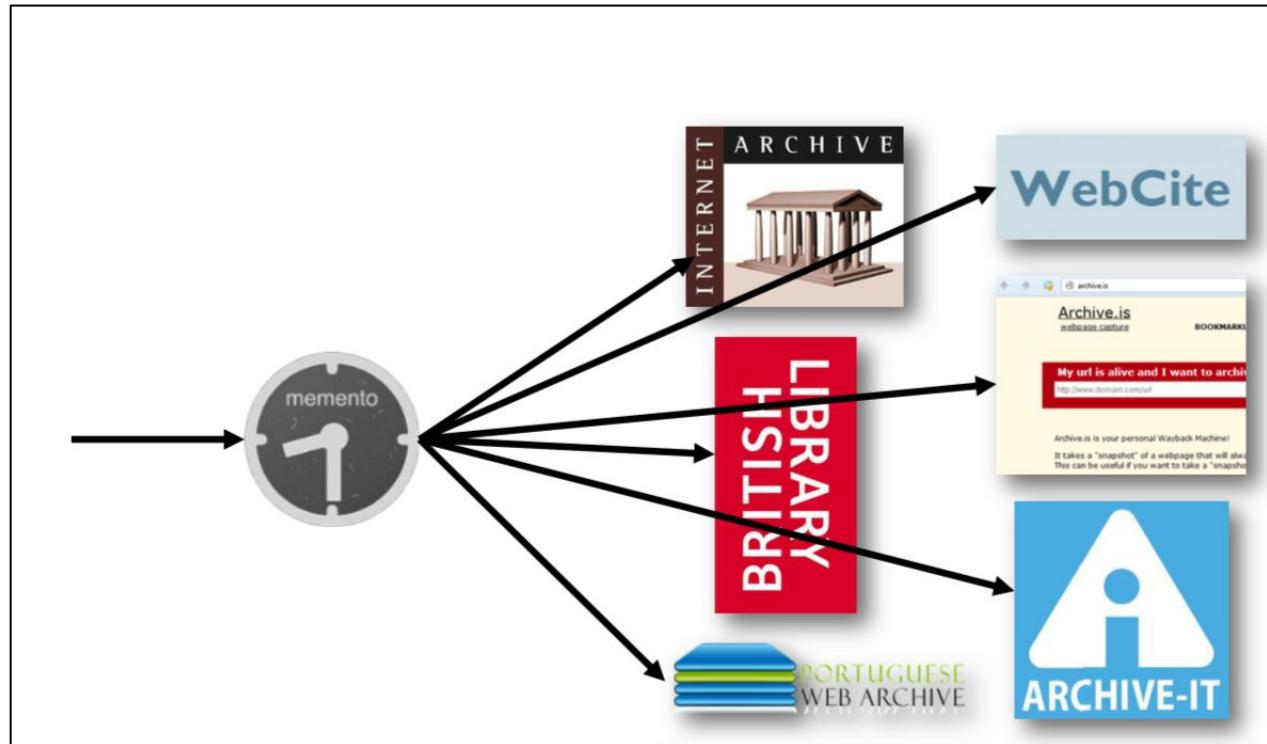
[ODU MemGator](#)

# Aggregator TimeMap - JSON Format

```
$ curl -i http://memgator.cs.odu.edu/timemap/json/http://www.cs.odu.edu/~mln/
HTTP/1.1 200 OK
Access-Control-Allow-Origin: *
Access-Control-Expose-Headers: Link, Location, X-Memento-Count, X-Generator
Content-Type: application/json
Date: Sun, 02 Feb 2020 20:09:25 GMT
X-Generator: MemGator:1.0-rc7
X-Memento-Count: 633
Transfer-Encoding: chunked

{
  "original_uri": "http://www.cs.odu.edu/~mln/",
  "self": "https://memgator.cs.odu.edu/timemap/json/http://www.cs.odu.edu/~mln/",
  "mementos": [
    {
      "list": [
        {
          "datetime": "2002-05-26T05:30:05Z",
          "uri": "http://web.archive.org/web/20020526053005/http://www.cs.odu.edu:80/~mln/"
        },
        {
          "datetime": "2002-05-28T08:33:21Z",
          "uri": "http://web.archive.org/web/20020528083321/http://www.cs.odu.edu:80/~mln/"
        },
        {
          "datetime": "2002-07-20T05:14:48Z",
          "uri": "http://web.archive.org/web/20020720051448/http://www.cs.odu.edu:80/~mln"
        },
        ...
      ]
    }
  ]
}
```

# More Than Just Internet Archive



source: Brunelle, "[Scripts in a Frame: A Two-Tiered Approach for Archiving Deferred Representations](#)", Dissertation Defense, February 5, 2016.

# More Than Just Internet Archive

```
$ curl -i http://memgator.cs.odu.edu/timemap/link/http://www.cs.odu.edu/~mln/ | grep -v web.archive.org
Server: nginx/1.11.3
Date: Thu, 02 Feb 2017 05:46:35 GMT
Content-Type: application/link-format
Transfer-Encoding: chunked
Connection: keep-alive
Access-Control-Allow-Origin: *
Access-Control-Expose-Headers: Link, Location, X-Memento-Count, X-Generator
X-Generator: MemGator:1.0-rc7
X-Memento-Count: 338

<http://www.cs.odu.edu/~mln/; rel="original",
<http://memgator.cs.odu.edu/timemap/link/http://www.cs.odu.edu/~mln/; rel="self"; type="application/link-format",
<http://archive.is/20130618185742/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Tue, 18 Jun 2013 18:57:42 GMT",
<http://archive.is/20130621194047/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Fri, 21 Jun 2013 19:40:47 GMT",
<http://archive.is/20130723172543/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Tue, 23 Jul 2013 17:25:43 GMT",
<http://archive.is/20130723173951/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Tue, 23 Jul 2013 17:39:51 GMT",
<http://archive.is/20130728010336/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Sun, 28 Jul 2013 01:03:36 GMT",
<http://archive.is/20131217200455/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Tue, 17 Dec 2013 20:04:55 GMT",
<http://archive.is/20141015185128/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Wed, 15 Oct 2014 18:51:28 GMT",
<http://archive.is/20141020194524/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Mon, 20 Oct 2014 19:45:24 GMT",
<http://archive.is/20141112185338/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Wed, 12 Nov 2014 18:53:38 GMT",
<http://archive.is/20141120172147/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Thu, 20 Nov 2014 17:21:47 GMT",
<http://archive.is/20141207184200/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Sun, 07 Dec 2014 18:42:00 GMT",
<http://archive.is/20141207184737/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Sun, 07 Dec 2014 18:47:37 GMT",
<http://archive.is/20150205233109/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Thu, 05 Feb 2015 23:31:09 GMT",
<http://archive.is/20150514160142/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Thu, 14 May 2015 16:01:42 GMT",
<http://archive.is/20150519034827/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Tue, 19 May 2015 03:48:27 GMT",
<http://archive.is/20150519041853/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Tue, 19 May 2015 04:18:53 GMT",
<http://archive.is/20150707000027/http://www.cs.odu.edu/~mln/; rel="memento"; datetime="Tue, 07 Jul 2015 00:00:27 GMT",
<http://wayback.archive-it.org/all/20150921151807/http://www.cs.odu.edu/%7Emln/; rel="memento";
datetime="Mon, 21 Sep 2015 15:18:07 GMT",
<http://wayback.archive-it.org/all/20150921164637/http://www.cs.odu.edu/~mln/; rel="memento";
datetime="Mon, 21 Sep 2015 16:46:37 GMT",
...
...
```

# Aggregator TimeGate

```
$ curl -I -H "Accept-datetime: Mon, 22 Jul 2013 00:00:00 GMT" \
      http://memgator.cs.odu.edu/timegate/http://www.cs.odu.edu/~mln/
HTTP/1.1 302 Found
Server: nginx/1.11.3
Date: Thu, 02 Feb 2017 05:49:13 GMT
Content-Type: text/plain; charset=utf-8
Connection: keep-alive
Access-Control-Allow-Origin: *
Access-Control-Expose-Headers: Link, Location, X-Memento-Count, X-Generator
Link: <http://www.cs.odu.edu/~mln/>; rel="original",
      <http://web.archive.org/web/20020526053005/http://www.cs.odu.edu/~mln/>;
      rel="first memento"; datetime="Sun, 26 May 2002 05:30:05 GMT",
      <http://web.archive.org/web/20130702115017/http://www.cs.odu.edu/~mln/>;
      rel="prev memento"; datetime="Tue, 02 Jul 2013 11:50:17 GMT",
      <http://archive.is/20130723172543/http://www.cs.odu.edu/~mln/>; rel="memento";
      datetime="Tue, 23 Jul 2013 17:25:43 GMT",
      <http://web.archive.org/web/20130730042519/http://www.cs.odu.edu/~mln/>;
      rel="next memento"; datetime="Tue, 30 Jul 2013 04:25:19 GMT",
      <http://archive.is/20170202053722/http://www.cs.odu.edu/~mln/>; rel="last memento";
      datetime="Thu, 02 Feb 2017 05:37:22 GMT",
      <http://memgator.cs.odu.edu/timemap/link/http://www.cs.odu.edu/~mln/>; rel="timemap";
      type="application/link-format",
      <http://memgator.cs.odu.edu/timemap/json/http://www.cs.odu.edu/~mln/>;
      rel="timemap"; type="application/json",
      <http://memgator.cs.odu.edu/timemap/cdxj/http://www.cs.odu.edu/~mln/>;
      rel="timemap"; type="application/cdxj+ors",
      <http://memgator.cs.odu.edu/timegate/http://www.cs.odu.edu/~mln/>; rel="timegate"
Location: http://archive.is/20130723172543/http://www.cs.odu.edu/~mln/
Vary: accept-datetime
X-Generator: MemGator:1.0-rc7
```

# Reminder that 302s can have entities...

```
$ curl -H "Accept-datetime: Mon, 22 Jul 2013 00:00:00 GMT" \
http://memgator.cs.odu.edu/timegate/http://www.cs.odu.edu/~mln/
<a href="http://archive.is/20130723172543/http://www.cs.odu.edu/~mln/">Found</a>.
```

```
$ curl -i -H "Accept-datetime: Mon, 22 Jul 2013 00:00:00 GMT" \
http://memgator.cs.odu.edu/timegate/http://www.cs.odu.edu/~mln/
```

HTTP/1.1 302 Found

Server: nginx/1.11.3

Date: Thu, 02 Feb 2017 05:54:20 GMT

Content-Type: text/html; charset=utf-8

Content-Length: 83

Connection: keep-alive

...

Location: http://archive.md/20130723172543/http://www.cs.odu.edu/~mln/

Vary: accept-datetime

X-Generator: MemGator:1.0-rc7

```
<a href="http://archive.md/20130723172543/http://www.cs.odu.edu/~mln/">Found</a>.
```

# Web Science: Archiving the Web

## (Part 4 - User Interfaces for Navigating Through Time on the Web)

CS 432/532

Old Dominion University

*Permission has been granted to use these slides from Frank McCown, Michael L. Nelson, Alexander Nwala, Michele C. Weigle*

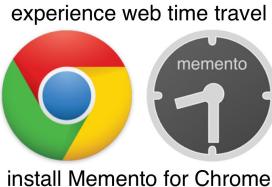


This work is licensed under a [Creative Commons](#)  
[Attribution-NonCommercial-ShareAlike 3.0 Unported License](#)

# Tools

## Browsing Past Web

- Memento for Chrome
- Mink (Chrome)
- Wayback Machine (Firefox)



## On-Demand Archiving

- IA Save Page Now
- archive.is/archive.today
- Conifer



# Time Travel with Memento for Chrome



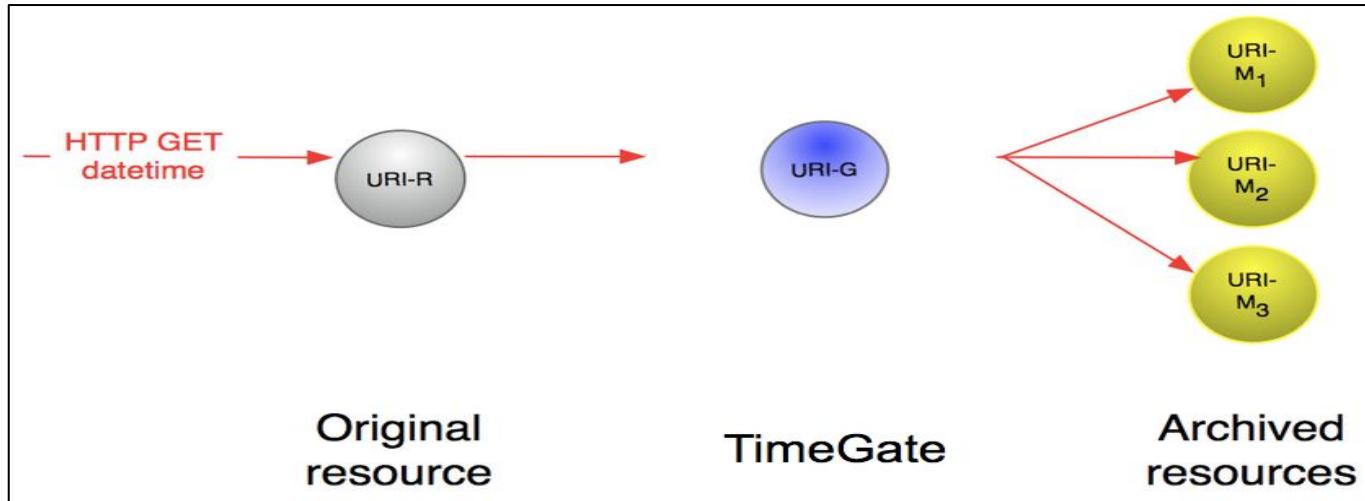
[Memento Time Travel \(Google Chrome store\)](#)

[WS-DL blog, "Right-Click to the Past -- Memento for Chrome"](#)

# Memento for Chrome



- Uses TimeGate for negotiation
- Can request a link at a specific datetime



# Time Travel: Read Like It's 2004

D-Lib Magazine  
September 2004

Volume 10 Number 9  
ISSN 1082-9873

## Rethinking Scholarly Communication

### Building the System that Scholars Deserve

**Herbert Van de Sompel**  
Los Alamos National Laboratory, Research Library  
<herbervt@lanl.gov>

**Sandy Payette**  
Cornell University, Computing and Information Science  
<spayette@cs.cornell.edu>

**John Erickson**  
Hewlett-Packard Laboratories, Digital Media Systems Lab  
<john.erickson@hp.com>

**Carl Lagoze**  
Cornell University, Computing and Information Science  
<lagoze@cs.cornell.edu>

**Simeon Warner**  
Cornell University, Computing and Information Science  
<simeon@cs.cornell.edu>

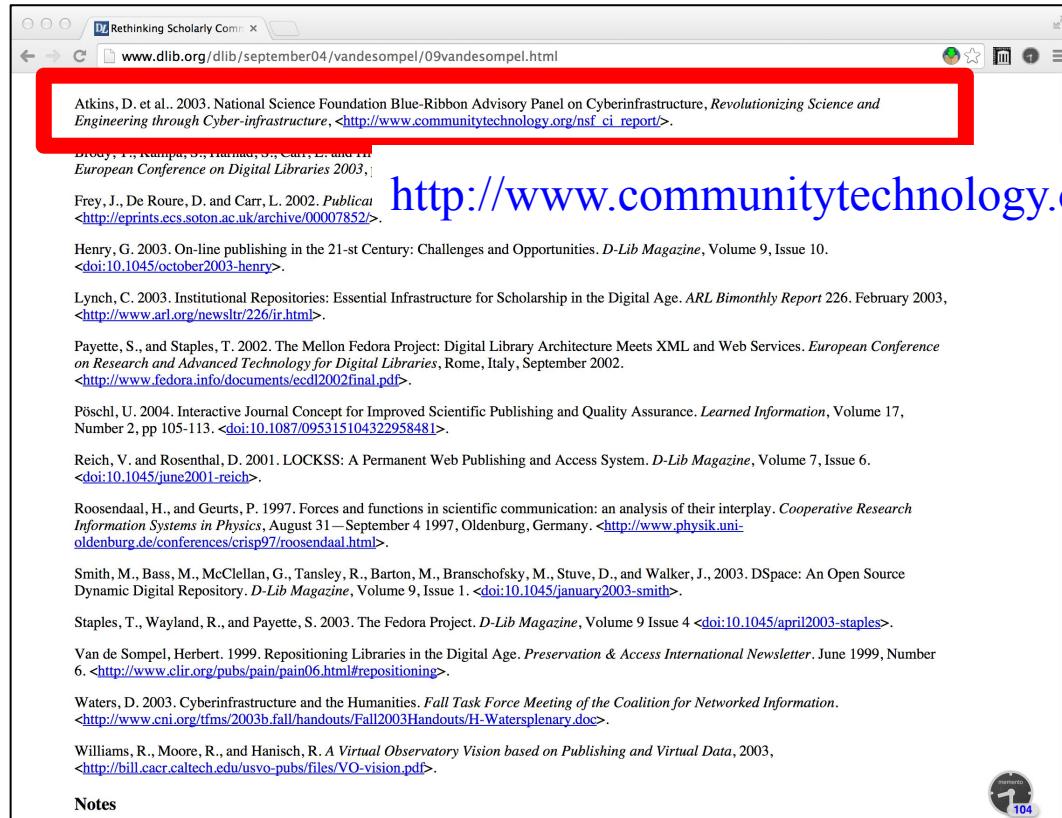
*(This Opinion piece presents the opinions of the author. It does not necessarily reflect the views of D-Lib Magazine, its publisher, the Corporation for National Research Initiatives, or its sponsor.)*

### Introduction

There is growing dissatisfaction with the established scholarly communication system. This dissatisfaction is the result of a variety of factors including rapidly rising subscription prices, concerns about copyright, latency between results and their actual publication, and restrictions on what can be published and how it can be disseminated. The result is a global debate on how to remedy the system's deficiencies, and that debate

webpage: [Rethinking Scholarly Communication: Building the System that Scholars Deserve](#)

# Time Travel: Scroll To The Bottom, First Reference



Atkins, D. et al.. 2003. National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Science and Engineering through Cyber-infrastructure*, <[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

Brody, T., Kumpu, S., Hurniak, S., Carr, L. and Trifunovic, S. 2003. Publishing in the Digital Library Age. In: European Conference on Digital Libraries 2003, 1-4 July 2003, Paris, France.

Frey, J., De Roure, D. and Carr, L. 2002. PublicaL. In: <http://eprints.ecs.soton.ac.uk/archive/00007852/>.

Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 10. doi:10.1045/october2003-henry.

Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003, doi:10.1045/february2003-lynch.

Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *European Conference on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. doi:10.1045/september2002-payette.

Pöschl, U. 2004. Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. *Learned Information*, Volume 17, Number 2, pp 105-113. doi:10.1087/095315104322958481.

Reich, V. and Rosenthal, D. 2001. LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, Volume 7, Issue 6. doi:10.1045/june2001-reich.

Roosendaal, H., and Geurts, P. 1997. Forces and functions in scientific communication: an analysis of their interplay. *Cooperative Research Information Systems in Physics*, August 31 – September 4 1997, Oldenburg, Germany. <<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>>.

Smith, M., Bass, M., McClellan, G., Tansley, R., Barton, M., Branschofsky, M., Stuve, D., and Walker, J., 2003. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, Volume 9, Issue 1. doi:10.1045/january2003-smith.

Staples, T., Wayland, R., and Payette, S. 2003. The Fedora Project. *D-Lib Magazine*, Volume 9 Issue 4 doi:10.1045/april2003-staples.

Van de Sompel, Herbert. 1999. Repositioning Libraries in the Digital Age. *Preservation & Access International Newsletter*. June 1999, Number 6. <<http://www.clir.org/pubs/pain/pain06.html#repositioning>>.

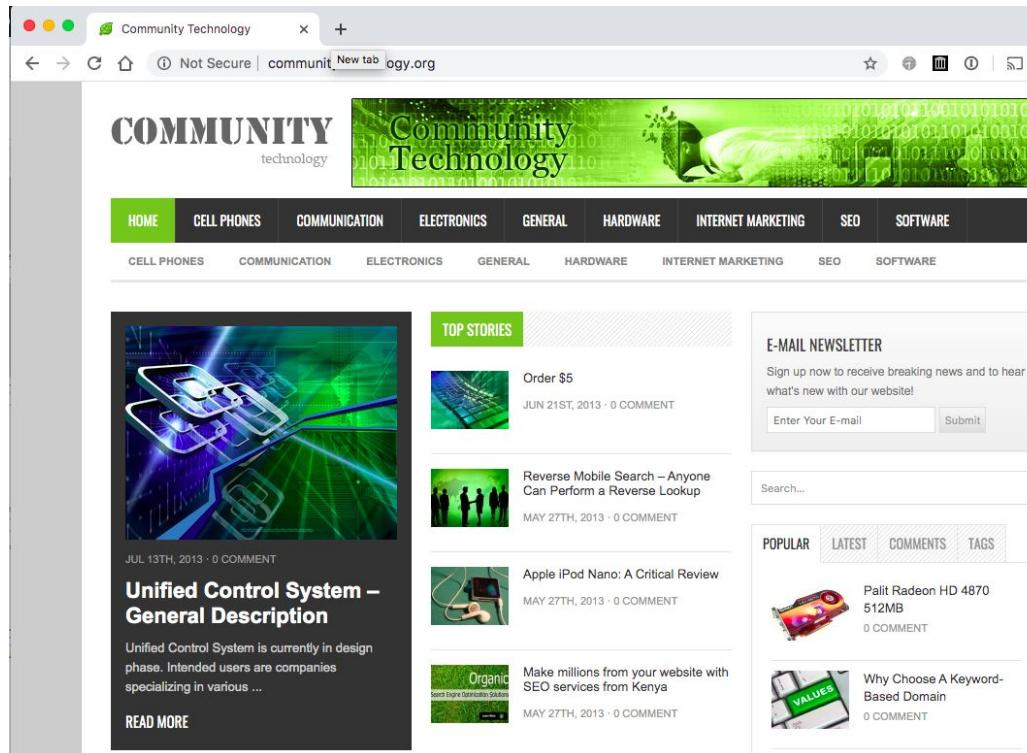
Waters, D. 2003. Cyberinfrastructure and the Humanities. *Fall Task Force Meeting of the Coalition for Networked Information*. <<http://www.cni.org/tfms/2003b.fall/handouts/Fall2003Handouts/H-Watersplenary.doc>>.

Williams, R., Moore, R., and Hanisch, R. A Virtual Observatory Vision based on Publishing and Virtual Data, 2003, doi:10.1045/september2003-williams.

Notes



# Time Travel: Link in 2020 Redirects To Main Site



URI: [http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/) -> <http://www.communitytechnology.org/>

# Time Travel: Set For September 15, 2004

Rethinking Scholarly Communication  
www.dlib.org/dlib/september04/vandesompel/09vandesompel.html

Atkins, D. et al.. 2003. National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Engineering through Cyber-infrastructure*, <[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

Brody, T., Kampa, S., Harnad, S., Carr, L. and Hitchcock, S. 2003. Digitometric Services for Open Archives Environment European Conference on Digital Libraries 2003, pages pp. 207-220, Trondheim, Norway. <<http://eprints.ecs.soton.ac.uk/120782/>>.

Frey, J., De Roure, D. and Carr, L. 2002. Publication at Source: Scientific Communication from a Publication Web to a D-Lib System. <<http://eprints.ecs.soton.ac.uk/archive/00007852/>>.

Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 1. <<doi:10.1045/october2003-henry>>.

Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* <<http://www.arl.org/newsltr/226/tr.htm>>.

Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. <<http://www.fedora.info/documents/ecdl2002/final.pdf>>.

Pöschl, U. 2004. Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. *Learned Information Number 2*, pp 105-113. <<doi:10.1087/095315104322958481>>.

Reich, V. and Rosenthal, D. 2001. LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, Volume 7, Issue 6. <<doi:10.1045/june2001-reich>>.

Roosendaal, H., and Geurts, P. 1997. Forces and functions in scientific communication: an analysis of their interplay. *Cooperative Research Information Systems in Physics*, August 31 – September 4 1997, Oldenburg, Germany. <<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.htm>>.

Smith, M., Bass, M., McClellan, G., Tansley, R., Barton, M., Branschofsky, M., Stuve, D., and Walker, J., 2003. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, Volume 9, Issue 1. <<doi:10.1045/january2003-smith>>.

Staples, T., Wayland, R., and Payette, S. 2003. The Fedora Project. *D-Lib Magazine*, Volume 9 Issue 4 <<doi:10.1045/april2003-staples>>.

Van de Sompel, Herbert. 1999. Repositioning Libraries in the Digital Age. *Preservation & Access International Newsletter*. June 1999, Number 6. <<http://www.clir.org/pubs/pain/pain06.html#repositioning>>.

Waters, D. 2003. Cyberinfrastructure and the Humanities. *Fall Task Force Meeting of the Coalition for Networked Information*. <<http://www.cni.org/tfms/2003b.fall/handouts/Fall2003Handouts/H-Watersplenary.doc>>.

Williams, R., Moore, R., and Hanisch, R. A Virtual Observatory Vision based on Publishing and Virtual Data, 2003, <<http://bill.caer.caltech.edu/usvo-pubs/files/VO-vision.pdf>>.

Notes

Date: 2004-09-15 Time: 22:48:57

Sep 2004

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

Select a date for time travel (in GMT). Right click on the page or on links to travel through time.

Set Cancel

Date: 2004-09-15 Time: 22:48:57

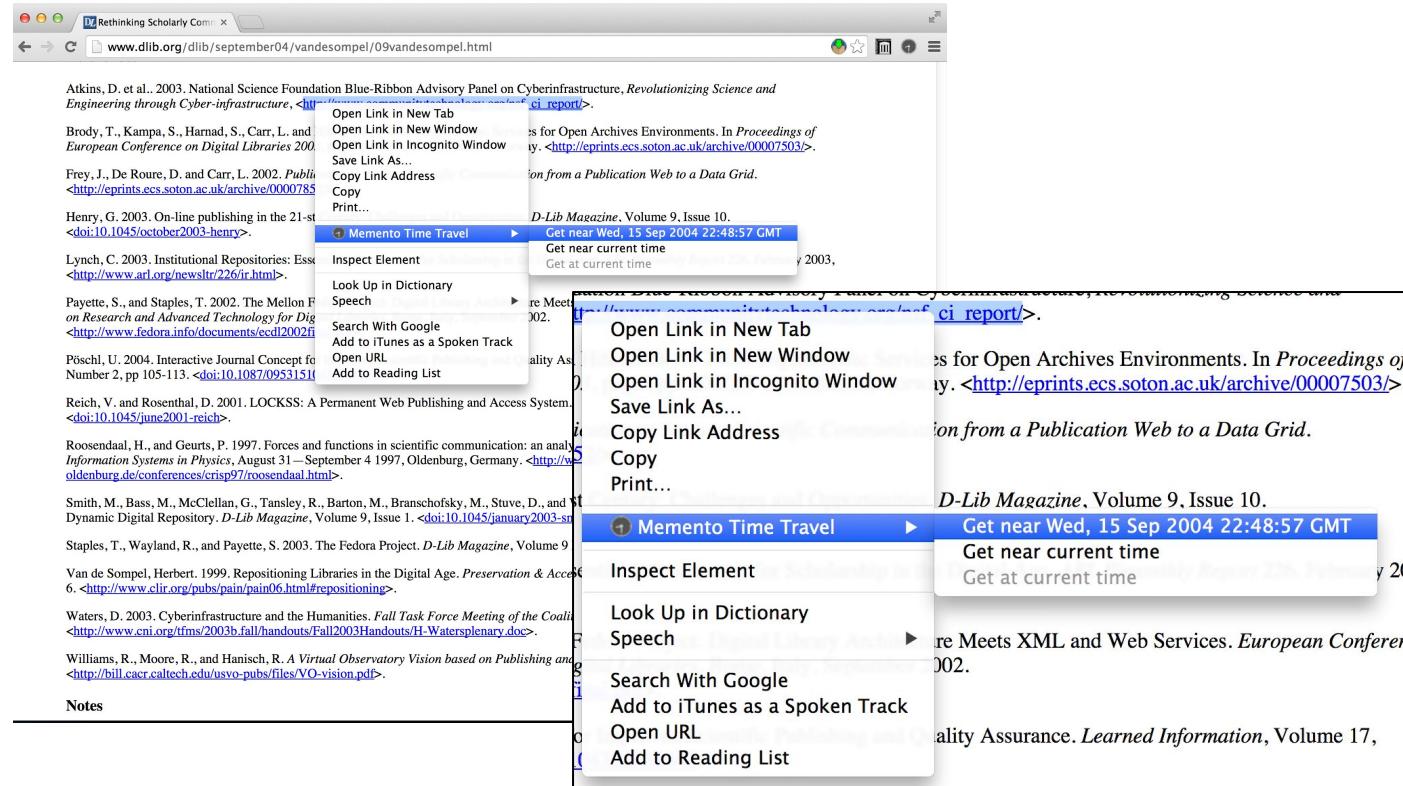
Sep 2004

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

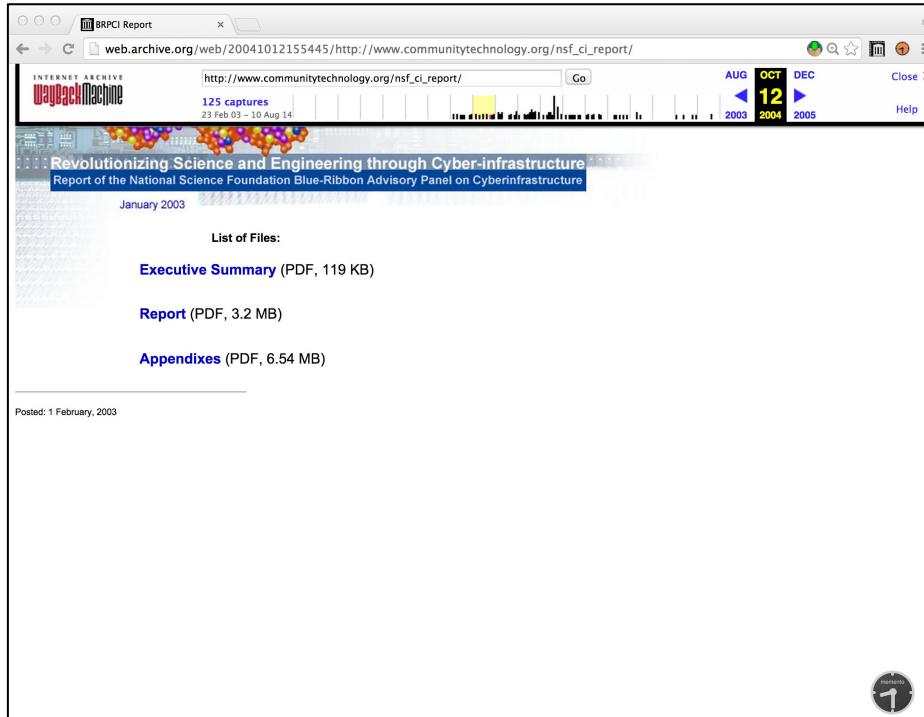
Select a date for time travel (in GMT). Right click on the page or on links to travel through time.

Set Cancel

# Time Travel: Right Click On The Link



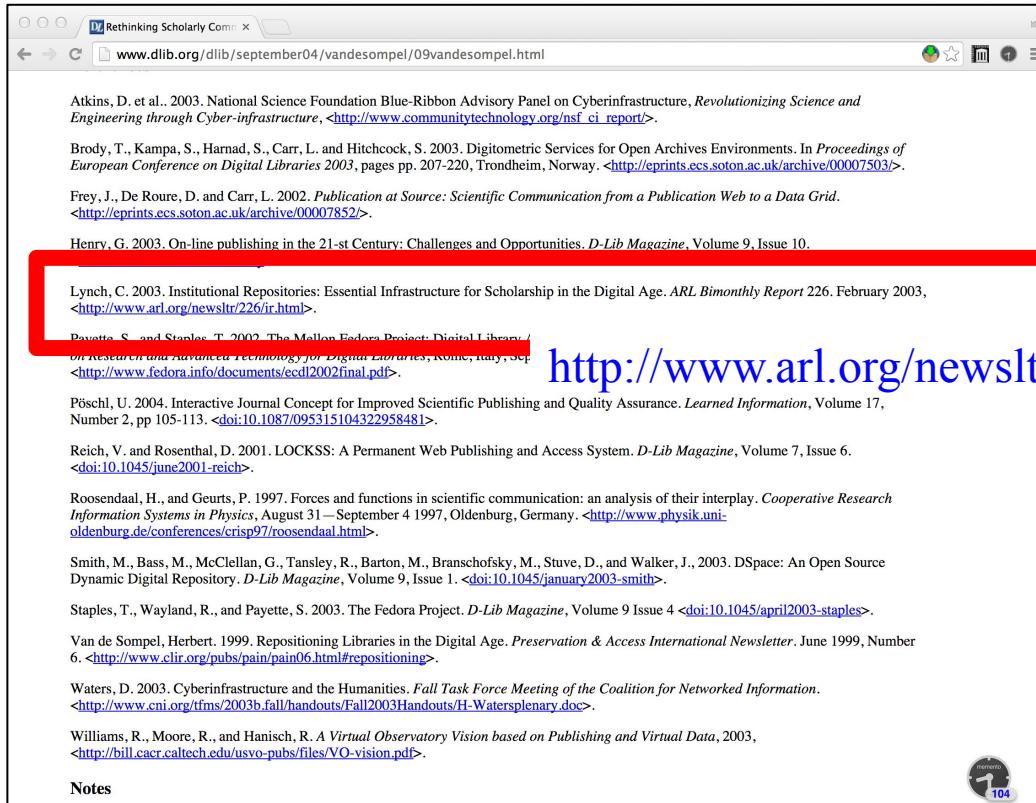
# Time Travel: The Report @ 2004-10-12



URI-M: [https://web.archive.org/web/20041012155445/http://www.communitytechnology.org/nsf\\_ci\\_report/](https://web.archive.org/web/20041012155445/http://www.communitytechnology.org/nsf_ci_report/)

List of URI-Ms: [http://timetravel.mementoweb.org/list/20040915093256/http://www.communitytechnology.org/nsf\\_ci\\_report/](http://timetravel.mementoweb.org/list/20040915093256/http://www.communitytechnology.org/nsf_ci_report/)

# Time Travel: Fifth Reference



Atkins, D. et al.. 2003. National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Science and Engineering through Cyber-infrastructure*, <[http://www.communitytechnology.org/nsf\\_ci\\_report](http://www.communitytechnology.org/nsf_ci_report)>.

Brody, T., Kampa, S., Harnad, S., Carr, L. and Hitchcock, S. 2003. Digitometric Services for Open Archives Environments. In *Proceedings of European Conference on Digital Libraries 2003*, pages pp. 207-220, Trondheim, Norway. <<http://eprints.ecs.soton.ac.uk/archive/00007503>>.

Frey, J., De Roure, D. and Carr, L. 2002. *Publication at Source: Scientific Communication from a Publication Web to a Data Grid*. <<http://eprints.ecs.soton.ac.uk/archive/00007852>>.

Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 10.

Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003, <<http://www.arl.org/newsltr/226/ir.html>>.

Payette, S. and Staples, T. 2002. The Mellon Fedora Project: Digital Library / on Research and Advanced Technology for Digital Libraries, Rome, Italy, Sep <<http://www.fedora.info/documents/ecl2002final.pdf>>.

Pöschl, U. 2004. Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. *Learned Information*, Volume 17, Number 2, pp 105-113. <<doi:10.1087/09531510432958481>>.

Reich, V. and Rosenthal, D. 2001. LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, Volume 7, Issue 6. <<doi:10.1045/june2001-reich>>.

Roosendaal, H., and Geurts, P. 1997. Forces and functions in scientific communication: an analysis of their interplay. *Cooperative Research Information Systems in Physics*, August 31 – September 4 1997, Oldenburg, Germany. <<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>>.

Smith, M., Bass, M., McClellan, G., Tansley, R., Barton, M., Branschofsky, M., Stuve, D., and Walker, J., 2003. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, Volume 9, Issue 1. <<doi:10.1045/january2003-smith>>.

Staples, T., Wayland, R., and Payette, S. 2003. The Fedora Project. *D-Lib Magazine*, Volume 9 Issue 4 <<doi:10.1045/april2003-staples>>.

Van de Sompel, Herbert. 1999. Repositioning Libraries in the Digital Age. *Preservation & Access International Newsletter*. June 1999, Number 6. <<http://www.clir.org/pubs/pain/06.html#repositioning>>.

Waters, D. 2003. Cyberinfrastructure and the Humanities. *Fall Task Force Meeting of the Coalition for Networked Information*. <<http://www.cni.org/tfms/2003b/fall/handouts/Fall2003Handouts/H-Watersplenary.doc>>.

Williams, R., Moore, R., and Hanisch, R. A *Virtual Observatory Vision based on Publishing and Virtual Data*, 2003, <<http://bill.cacr.caltech.edu/usvo-pubs/files/VO-vision.pdf>>.

**Notes**

104

# Time Travel: Current Page Is 404

The screenshot shows a web browser window with a blue header bar. The title bar says "Page not found - Association". The URL in the address bar is "arl.org/newsltr/226/ir.html". The main content area has a dark blue background with white text. At the top left is the "ASSOCIATION OF RESEARCH LIBRARIES" logo. On the right side of the header is a "Member Login" link. Below the header is a navigation bar with links: WHO WE ARE, OUR PRIORITIES, ARL ACADEMY, NEWS, RESOURCES, EVENTS, JOBS, and a magnifying glass icon for search. The main content starts with "Home / Error 404". Below that is a large bold heading "We're sorry, this page doesn't exist". Underneath the heading, there is a paragraph of text: "The page you requested could not be found and may no longer exist on our new website. Use search to find related information or the navigation bar to explore the site content." To the right of this text is a sidebar titled "Tweets by @ARLnews". It shows two tweets: one from "Association of Research Libraries" (@ARLnews) and one from "Library of Congress" (@librarycongress). Below the sidebar is a small grid of five thumbnail images.

Home / Error 404

## We're sorry, this page doesn't exist

The page you requested could not be found and may no longer exist on our new website. Use search to find related information or the navigation bar to explore the site content.

Are you looking for one of these topics?

**Advocacy & Public Policy**  
Copyright & intellectual property, codes of best practices in fair use, net neutrality, privacy, accessibility, civil liberties, freedom of expression, information access

**Data & Analytics**  
ARL Statistics Survey, ARL Annual Salary Survey, ClimateQUAL, LibQUAL+, Research Library Impact Framework, statistical trends in ARL libraries

**Diversity, Equity & Inclusion**  
Kaleidoscope Program, Leadership and Career Development Program, Fellowship for Digital and Inclusive Excellence, ARL Annual Leadership Symposium, ARL/SAA Mosaic Program

**Scholars & Scholarship**  
TOME open monograph initiative, scholarly publishing, open scholarship, scholarly communities, digital scholarship, research data, archives and special collections, SHARE

**ARL Academy**  
ARL Leadership Fellows Program, ARL Digital Scholarship Institute, Library Management Skills Institutes, Julia C. Blixrud scholarship and lecture, coaching & mentoring pilot, On

Tweets by @ARLnews

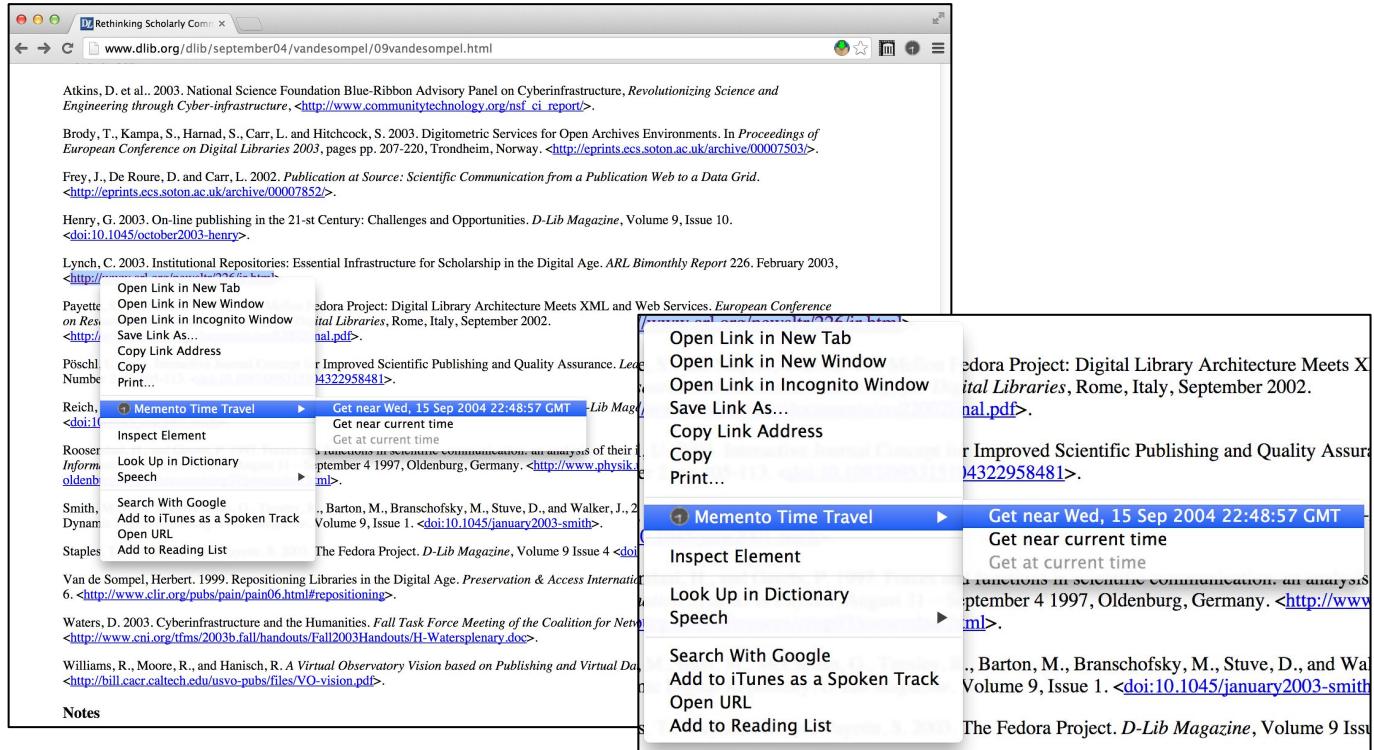
Association of Research Libraries Retweeted

Library of Congress @librarycongress

MUSIC: Summertime concerts from the Library: Every Wednesday at noon ET through the summer, we'll share new concerts in our Homegrown series from the American Folklife Center. DETAILS: loc.gov/item/prn-20-04...



# Time Travel: Right Click On The Link



# Time Travel: Page As Of 2004-10-12

The screenshot shows a browser window with the Wayback Machine interface at the top. The URL is <http://web.archive.org/web/20041012054551/http://www.arl.org/newsltr/226/ir.html>. The Wayback Machine logo and "107 Captures" are visible. A date navigation bar shows "AUG 12 NOV 2003 2004 2005". Below the interface, the ARL Bimonthly Report 226 page is displayed. The title is "ARL Bimonthly Report 226" and the date is "February 2003". The main article is titled "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" by Clifford A. Lynch. The text discusses the development of institutional repositories as a new strategy for scholarly communication, mentioning the fall of 2002, online storage costs, and the role of DSpace. It also mentions the leadership of MIT and the Hewlett Packard Corporation.

**Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age**  
by Clifford A. Lynch, Executive Director, Coalition for Networked Information

**Introduction**

In the fall of 2002, something extraordinary occurred in the continuing networked information revolution, shifting the dynamic among individually driven innovation, institutional progress, and the evolution of disciplinary scholarly practices. The development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content, and also scaling up beyond ad-hoc alliances, partnerships, and support arrangements with a few select faculty pioneers exploring more transformative new uses of the digital medium.

Many technology trends and development efforts came together to make this strategy possible. Online storage costs have dropped significantly; repositories are now affordable. Standards like the open archives metadata harvesting protocol are now in place; some progress has also been made on the standards for the underlying metadata itself. The thinking about digital preservation over the past five years has advanced to the point where the needs are widely recognized and well defined, the technical approaches at least superficially mapped out and the need for action is now clear. The development of free, publicly accessible journal article collections in disciplines such as high-energy physics has demonstrated ways in which the network can change scholarly communication by altering dissemination and access patterns; separately, the development of a series of extraordinary digital works had at least suggested the potential of creative authorship specifically for the digital medium to transform the presentation and transmission of scholarship.

The leadership of the Massachusetts Institute of Technology (MIT) in the development and deployment of the DSpace institutional repository system (<http://www.dspace.org/>), created in collaboration with the Hewlett Packard Corporation, has been a model pointing the way forward for many other universities. In 2003, with funding from The Andrew W. Mellon Foundation and other sources, MIT's DSpace was scheduled to be replicated at a number of additional institutions around the world; the software has also been released publicly under an open source arrangement, greatly lowering the cost and development barriers to implementing repositories for all institutions. The MIT software is

URI-M: <https://web.archive.org/web/20041012054551/http://www.arl.org:80/newsltr/226/ir.html>

List of URI-Ms: <http://timetravel.mementoweb.org/list/20040915054551/http://www.arl.org/newsltr/226/ir.html>

# Time Travel with Mink



*Developed while Mat was a PhD student at ODU*



[Mink at Google Chrome Store](#)

# Mink



- Uses TimeMap (list of mementos)
- Must first have a valid memento to visit other mementos directly

```
<http://www.cs.odu.edu/~mln/>; rel="original",
<https://memgator.cs.odu.edu/timemap/link/http://www.cs.odu.edu/~mln/>; rel="self"; type="application/link-format",
<http://web.archive.org/web/20020526053005/http://www.cs.odu.edu:80/~mln/>; rel="first memento"; datetime="Sun, 26 May 2002 05:30:05 GMT",
<http://web.archive.org/web/20020528083321/http://www.cs.odu.edu:80/~mln/>; rel="memento"; datetime="Tue, 28 May 2002 08:33:21 GMT",
<http://web.archive.org/web/20020720051448/http://www.cs.odu.edu:80/~mln>; rel="memento"; datetime="Sat, 20 Jul 2002 05:14:48 GMT",
<http://web.archive.org/web/20020724230216/http://www.cs.odu.edu:80/~mln/>; rel="memento"; datetime="Wed, 24 Jul 2002 23:02:16 GMT",
<http://web.archive.org/web/20020811035033/http://www.cs.odu.edu:80/~mln>; rel="memento"; datetime="Sun, 11 Aug 2002 03:50:33 GMT",
<http://web.archive.org/web/20021005054529/http://www.cs.odu.edu:80/~mln/>; rel="memento"; datetime="Sat, 05 Oct 2002 05:45:29 GMT",
```

# Mink: Read Like It's 2004

The screenshot shows a web browser window with the title bar "Rethinking Scholarly Comm x". The URL in the address bar is "www.dlib.org/dlib/september04/vandesompel/09vandesompel.html". The page content is as follows:

**D-Lib Magazine**  
September 2004

Volume 10 Number 9  
ISSN 1082-9873

**Rethinking Scholarly Communication**

**Building the System that Scholars Deserve**

**Herbert Van de Sompel**  
Los Alamos National Laboratory, Research Library  
<herbertv@lanl.gov>

**Sandy Payette**  
Cornell University, Computing and Information Science  
<payette@cs.cornell.edu>

**John Erickson**  
Hewlett-Packard Laboratories, Digital Media Systems Lab  
<john.erickson@hp.com>

**Carl Lagoze**  
Cornell University, Computing and Information Science  
<lagoze@cs.cornell.edu>

**Simeon Warner**  
Cornell University, Computing and Information Science  
<simeon@cs.cornell.edu>

*(This Opinion piece presents the opinions of the author. It does not necessarily reflect the views of D-Lib Magazine, its publisher, the Corporation for National Research Initiatives, or its sponsor.)*

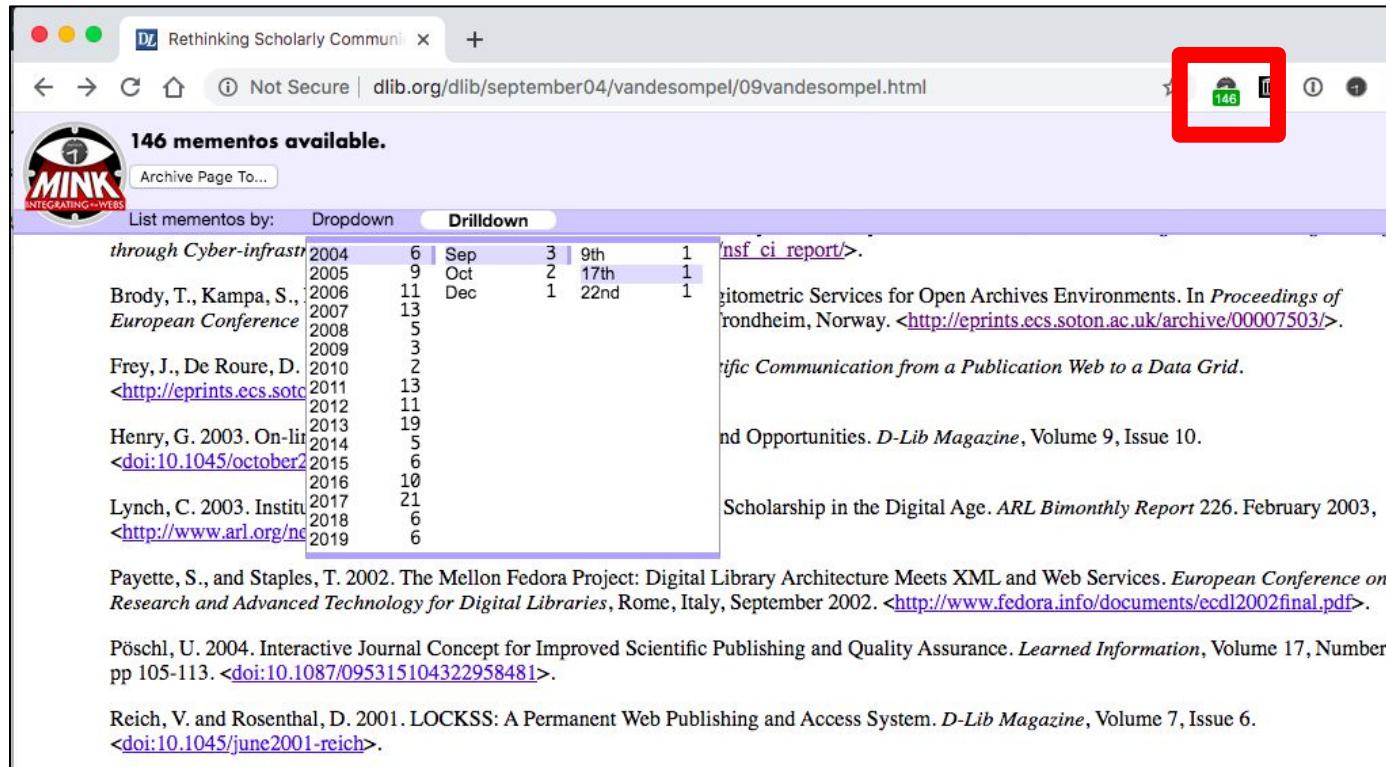
**Introduction**

There is growing dissatisfaction with the established scholarly communication system. This dissatisfaction is the result of a variety of factors including rapidly rising subscription prices, concerns about copyright, latency between results and their actual publication, and restrictions on what can be published and how it can be disseminated. The result is a global debate on how to remedy the system's deficiencies, and that debate

monograph 1 104

webpage: [Rethinking Scholarly Communication: Building the System that Scholars Deserve](http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html)

# Mink: Choose memento to view



146 mementos available.

List mementos by: Dropdown Drilldown

through Cyber-infrastr	2004	6	Sep	3	9th	1
Brody, T., Kampa, S.,	2005	9	Oct	2	17th	1
<i>European Conference</i>	2006	11	Dec	1	22nd	1
	2007	13				
	2008	5				
	2009	3				
Frey, J., De Roure, D.	2010	2				
< <a href="http://eprints.ecs.soton.ac.uk">http://eprints.ecs.soton.ac.uk</a> >	2011	13				
	2012	11				
	2013	19				
Henry, G. 2003. On-line	2014	5				
< <a href="doi:10.1045/october2003-henry">doi:10.1045/october2003-henry</a> >	2015	6				
	2016	10				
Lynch, C. 2003. Institu	2017	21				
< <a href="http://www.arl.org/ne">http://www.arl.org/ne</a> >	2018	6				
	2019	6				

'nsf\_ci\_report/':

gitometric Services for Open Archives Environments. In *Proceedings of Trondheim, Norway*. <<http://eprints.ecs.soton.ac.uk/archive/00007503/>>.

ific Communication from a Publication Web to a Data Grid.

nd Opportunities. *D-Lib Magazine*, Volume 9, Issue 10.

Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003,

Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *European Conference on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. <<http://www.fedora.info/documents/ecdl2002final.pdf>>.

Pöschl, U. 2004. Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. *Learned Information*, Volume 17, Number 2 pp 105-113. <<doi:10.1087/095315104322958481>>.

Reich, V. and Rosenthal, D. 2001. LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, Volume 7, Issue 6. <<doi:10.1045/june2001-reich>>.

# Mink: Now, links are to mementos

The screenshot shows a web browser window with the title "Rethinking Scholarly Communication" and the URL "https://web.archive.org/web/20040909135231/http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html". The browser interface includes standard controls like back, forward, and search, along with a toolbar and a status bar indicating "134 captures /epri". The main content area displays a timeline from "9 Sep 2004 - 13 May 2020" with a vertical bar chart showing activity levels. A specific capture from "2004 SEP 09" is highlighted. Below the timeline, several academic articles are listed:

- Frey, J., De Roure, D. and Carr, L. 2002. *Publication at Source: Scientific Communication from a Publication Web to a Data Grid.* <<http://eprints.ecs.soton.ac.uk/archive/00007852/>>.
- Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 10. <<doi:10.1045/october2003-henry>>.
- Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003, <<http://www.arl.org/newsltr/226/ir.html>>.
- Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *European Conference on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. <<http://www.fedora.info/documents/ecdl2002final.pdf>>.

A red box highlights the URL at the bottom of the page: "https://web.archive.org/web/20040909135231/http://www.arl.org/newsltr/226/ir.html | 222058181".

# Mink: Click on Fifth Reference

The screenshot shows a web browser window with two tabs: "Rethinking Scholarly Commun..." and "ARL 226: Institutional Repository". The main content area displays the "Wayback Machine" interface for the URL <http://www.arl.org/newsltr/226/ir.html>. The interface includes a toolbar with icons for search, refresh, and download, and a status bar indicating "Paused". Below the toolbar, it shows "136 captures" between "29 Mar 2003 - 28 Nov 2019". A timeline at the top right shows months from AUG to NOV, with "OCT" highlighted in yellow and "12" in red. The date "2003 2004 2005" is also visible. The page content itself features the ARL logo (purple letters over a grid background) and the title "ARL Bimonthly Report 226" followed by "February 2003". The main article is titled "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" by Clifford A. Lynch, Executive Director, Coalition for Networked Information. The introduction discusses the development of institutional repositories and their impact on the scholarly communication landscape.

**ARL Bimonthly Report 226**  
February 2003

**Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age**  
by Clifford A. Lynch, Executive Director, Coalition for Networked Information

**Introduction**  
In the fall of 2002, something extraordinary occurred in the continuing networked information revolution, shifting the dynamic among individually driven innovation, institutional progress, and the evolution of disciplinary scholarly practices. The development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content, and also scaling up beyond ad-hoc alliances, partnerships, and support arrangements with a few select faculty pioneers exploring more transformative new uses of the digital medium.

Many technology trends and development efforts came together to make this strategy possible. Online storage costs have dropped significantly; repositories are now affordable. Standards like the open archives metadata harvesting protocol are now in place; some progress has also been made on the standards for the underlying metadata itself. The thinking about digital preservation over the past five years has advanced to the point where the needs are widely recognized and well defined, the technical approaches at least superficially mapped out, and the need for action is now clear. The development of free, publicly accessible journal article collections in disciplines such as high-energy physics has demonstrated ways in which the network can change scholarly communication by altering dissemination and access patterns; separately, the development of a series of extraordinary digital works had at least suggested the potential of creative

# Wayback Machine



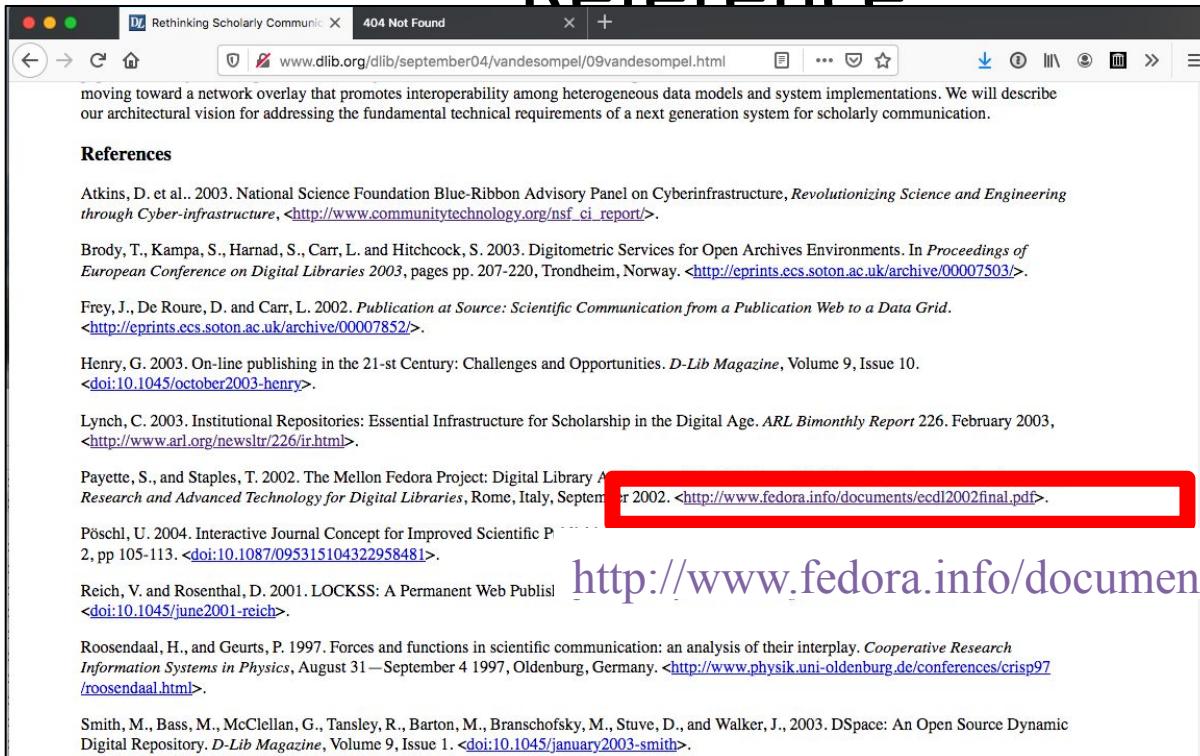
## Wayback Machine by [Internet Archive, Mark Graham](#)

Detects dead pages, 404s, DNS failures & a range of other web breakdowns, offering to show archived versions via the Internet Archive's Wayback Machine. In addition you can archive web pages, and see their most recent & first archives.



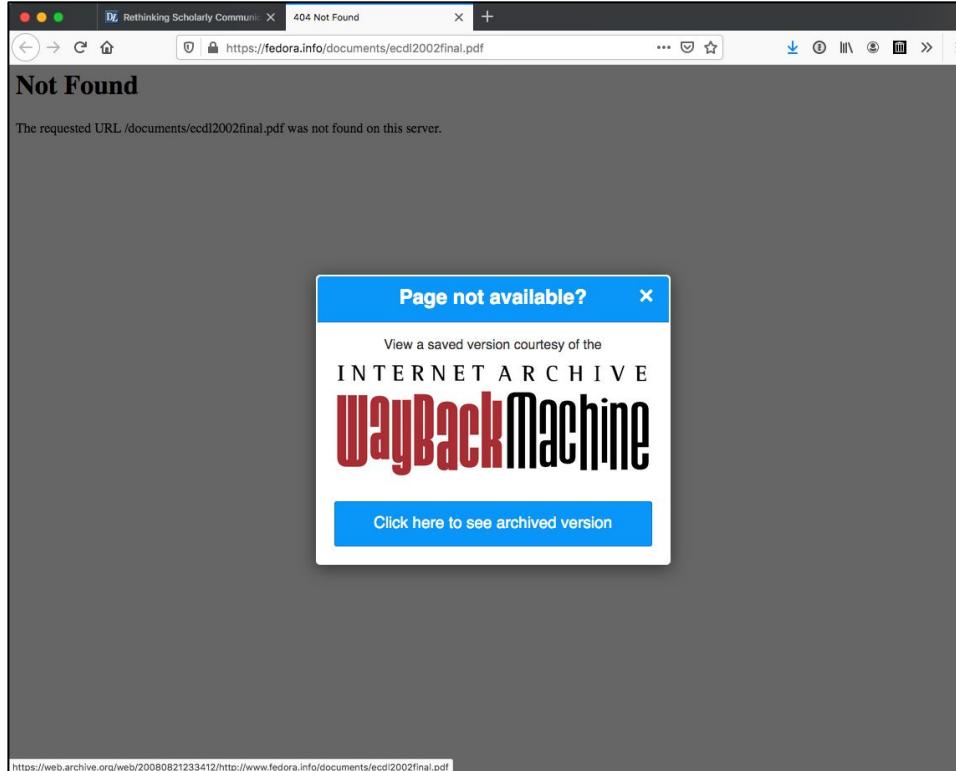
[Wayback Machine Firefox Extension](#)

# Wayback Machine: Click on Sixth Reference



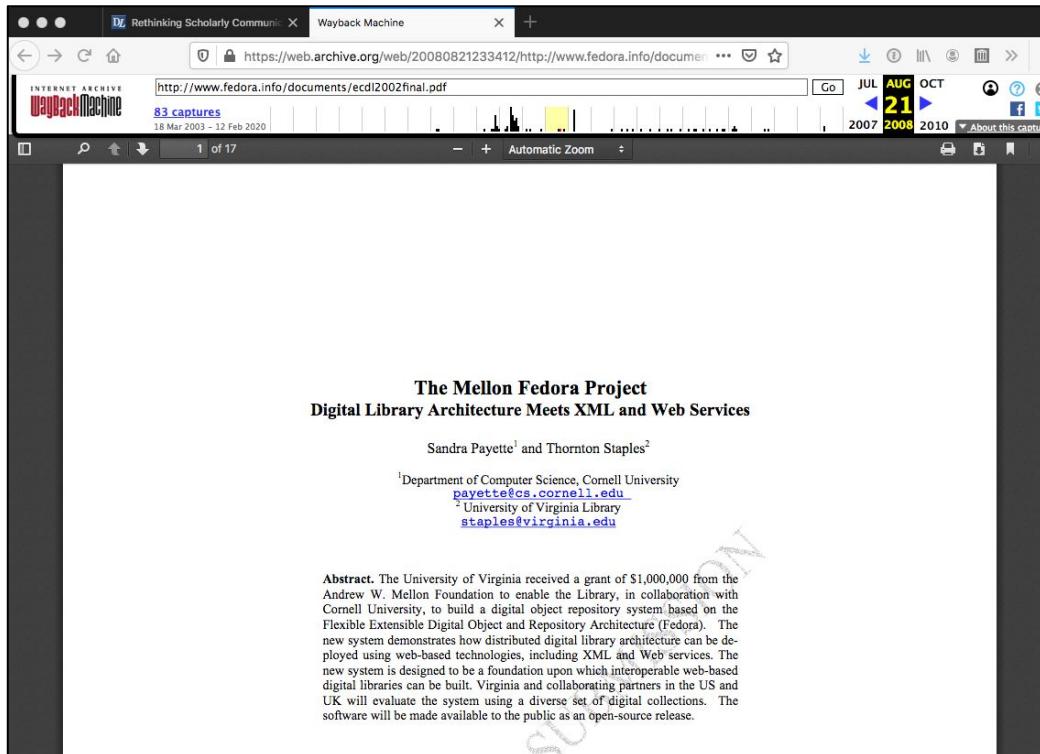
webpage: [Rethinking Scholarly Communication: Building the System that Scholars Deserve](#)

# Wayback Machine: Returns 404



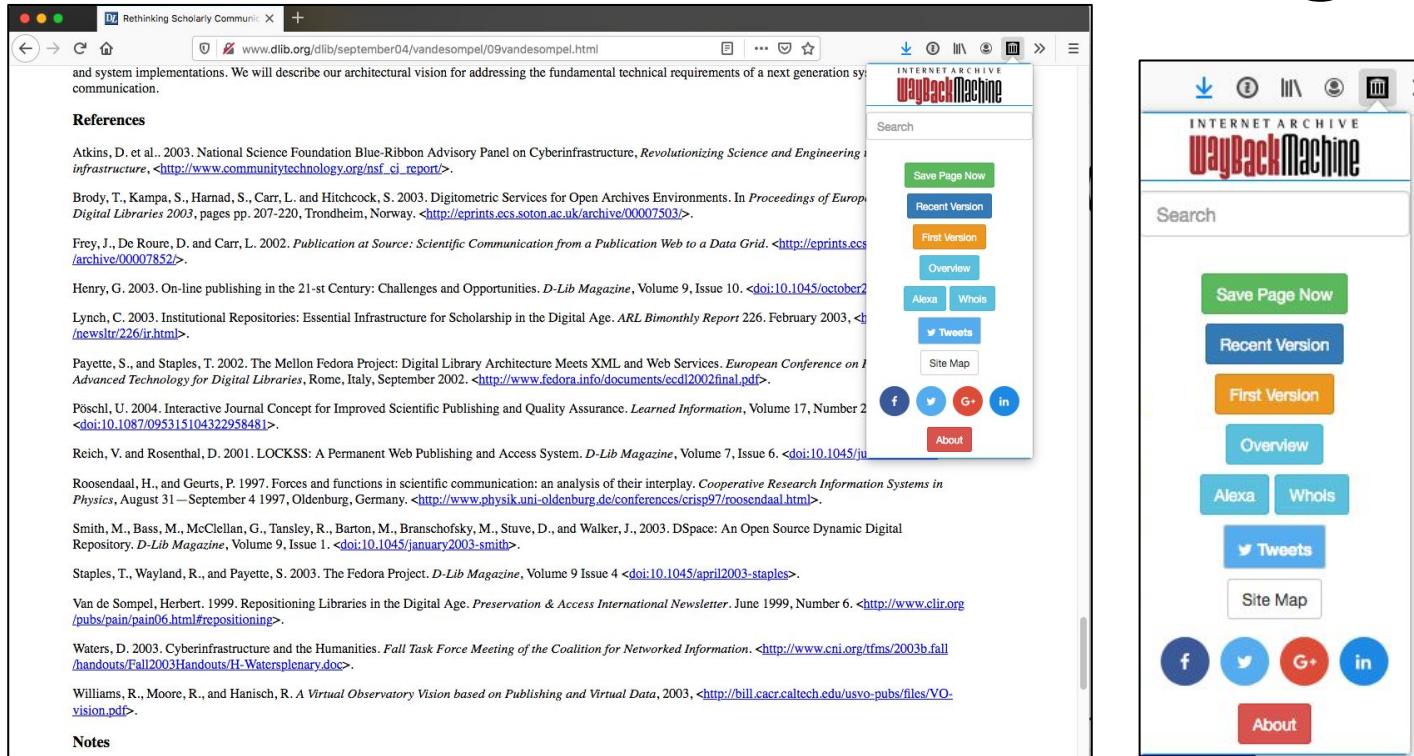
URI: <https://fedora.info/documents/ecdl2002final.pdf>

# Wayback Machine: See Archived Version



archived URI: <https://web.archive.org/web/20080821233412/http://www.fedora.info/documents/ecdl2002final.pdf>

# Wayback Machine: Use on Live Page



The image shows a screenshot of a web browser window. The main content area displays a live webpage from [www.dlib.org/dlib/september04/vandesompel/09vandesompel.html](http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html). The page contains text about system implementations and architectural visions for scholarly communication. Below this is a section titled "References" listing various academic papers. On the right side of the browser window, a sidebar for the "INTERNET ARCHIVE Wayback Machine" is visible. This sidebar includes a search bar, several buttons for interacting with the page (Save Page Now, Recent Version, First Version, Overview, Alexa, Whois, Tweets, Site Map), and social media sharing icons for Facebook, Twitter, Google+, and LinkedIn. At the bottom of the sidebar is a red "About" button.

INTERNET ARCHIVE  
Wayback Machine

Search

Save Page Now

Recent Version

First Version

Overview

Alexa Whois

Tweets

Site Map

About

f t G+ in

webpage: [Rethinking Scholarly Communication: Building the System that Scholars Deserve](http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html)

# Wayback Machine: Choose First Version

The screenshot shows a web browser window displaying a Wayback Machine archive of a page from September 2004. The title bar reads "Rethinking Scholarly Communication" and "Rethinking Scholarly Communication". The Wayback Machine logo is visible at the top left. A timeline at the top right shows dates from 2003 to 2005, with "SEP 09" highlighted in yellow. Below the timeline, a bar chart indicates "134 captures" between "9 Sep 2004 - 13 May 2020". The main content area is titled "References" and lists numerous academic papers and reports. Some entries include DOIs and URLs. The Wayback Machine interface includes standard browser controls like back, forward, and search.

References

Atkins, D. et al.. 2003. National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Science and Engineering through Cyber-infrastructure*, <[http://www.communitytechnology.org/nsf\\_ci\\_report](http://www.communitytechnology.org/nsf_ci_report)>.

Brody, T., Kampa, S., Hamad, S., Carr, L. and Hitchcock, S. 2003. Digitometric Services for Open Archives Environments. In *Proceedings of European Conference on Digital Libraries 2003*, pages pp. 207-220, Trondheim, Norway. <<http://eprints.ecs.soton.ac.uk/archive/00007503>>.

Frey, J., De Roure, D. and Carr, L. 2002. *Publication at Source: Scientific Communication from a Publication Web to a Data Grid*. <<http://eprints.ecs.soton.ac.uk/archive/00007852>>.

Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 10. <[doi:10.1045/october2003-henry](https://doi.org/10.1045/october2003-henry)>.

Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003, <<http://www.arl.org/newslet/226/ir.html>>.

Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *European Conference on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. <<http://www.fedora.info/documents/ecd12002final.pdf>>.

Pöschl, U. 2004. Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. *Learned Information*, Volume 17, Number 2, pp 105-113. <[doi:10.1087/095315104322958481](https://doi.org/10.1087/095315104322958481)>.

Reich, V. and Rosenthal, D. 2001. LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, Volume 7, Issue 6. <[doi:10.1045/june2001-reich](https://doi.org/10.1045/june2001-reich)>.

Rosendaal, H., and Geurts, P. 1997. Forces and functions in scientific communication: an analysis of their interplay. *Cooperative Research Information Systems in Physics*, August 31 – September 4 1997, Oldenburg, Germany. <<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>>.

Smith, M., Bass, M., McClellan, G., Tansley, R., Barton, M., Branschofsky, M., Stuve, D., and Walker, J., 2003. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, Volume 9, Issue 1. <[doi:10.1045/january2003-smith](https://doi.org/10.1045/january2003-smith)>.

Staples, T., Wayland, R., and Payette, S. 2003. The Fedora Project. *D-Lib Magazine*, Volume 9 Issue 4 <[doi:10.1045/april2003-staples](https://doi.org/10.1045/april2003-staples)>.

Van de Sompel, Herbert. 1999. Repositioning Libraries in the Digital Age. *Preservation & Access International Newsletter*. June 1999, Number 6. <<http://www.clir.org/pubs/pain06.html#repositioning>>.

Waters, D. 2003. Cyberinfrastructure and the Humanities. *Fall Task Force Meeting of the Coalition for Networked Information*. <<http://www.cni.org/tfms/2003b.fall/handouts/Fall2003Handouts/H-Watersplenary.doc>>.

Williams, R., Moore, R., and Hanisch, R. *A Virtual Observatory Vision based on Publishing and Virtual Data*, 2003, <<http://bill.cacr.caltech.edu/usvo-pubs/files/VO-vision.pdf>>.

archived URI:

<https://web.archive.org/web/20040909135231/http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>

# Wayback Machine: Links are to mementos

The screenshot shows a web browser window with the title "Rethinking Scholarly Communication" and the URL "https://web.archive.org/web/20040909135231/http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html". The browser interface includes standard navigation buttons, a search bar, and a toolbar. A timeline at the top right indicates the capture date is from September 2004. The main content area displays several academic articles:

- Frey, J., De Roure, D. and Carr, L. 2002. *Publication at Source: Scientific Communication from a Publication Web to a Data Grid.* <<http://eprints.ecs.soton.ac.uk/archive/00007852/>>.
- Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 10. <<doi:10.1045/october2003-henry>>.
- Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003, <<http://www.arl.org/newsltr/226/ir.html>>.
- Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *European Conference on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. <<http://www.fedora.info/documents/ecdl2002final.pdf>>.

At the bottom of the page, there is a redacted URL: "https://web.archive.org/web/20040909135231/http://www.arl.org/newsltr/226/ir.html 232059491".

archived URI: <https://web.archive.org/web/20040909135231/http://www.arl.org/newsltr/226/ir.html>

# Wayback Machine: Click Fifth Reference

The screenshot shows a browser window with the Wayback Machine interface. The address bar displays the URL <https://web.archive.org/web/20040810212907/http://www.arl.org/newsltr/226/ir.html>. Above the browser window, the Wayback Machine logo is visible. The main content area shows the ARL Bimonthly Report 226 from February 2003. The page features the ARL logo at the top, followed by the title "ARL Bimonthly Report 226" and the date "February 2003". Below the title, the article title is "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" by Clifford A. Lynch, Executive Director, Coalition for Networked Information. The text discusses the development of institutional repositories and their role in the digital age. It mentions the DSpace institutional repository system developed at MIT and its impact on scholarly communication. The text also refers to a workshop jointly sponsored by ARL, CNI, and SPARC in Washington, D.C., at the DSpace launch celebration at MIT.

**Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age**  
by Clifford A. Lynch, Executive Director, Coalition for Networked Information

**Introduction**  
In the fall of 2002, something extraordinary occurred in the continuing networked information revolution, shifting the dynamic among individually driven innovation, institutional progress, and the evolution of disciplinary scholarly practices. The development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content, and also scaling up beyond ad-hoc alliances, partnerships, and support arrangements with a few select faculty pioneers exploring more transformative new uses of the digital medium.

Many technology trends and development efforts came together to make this strategy possible. Online storage costs have dropped significantly; repositories are now affordable. Standards like the open archives metadata harvesting protocol are now in place; some progress has also been made on the standards for the underlying metadata itself. The thinking about digital preservation over the past five years has advanced to the point where the needs are widely recognized and well defined, the technical approaches at least superficially mapped out, and the need for action is now clear. The development of free, publicly accessible journal article collections in disciplines such as high-energy physics has demonstrated ways in which the network can change scholarly communication by altering dissemination and access patterns; separately, the development of a series of extraordinary digital works had at least suggested the potential of creative authorship specifically for the digital medium to transform the presentation and transmission of scholarship.

The leadership of the Massachusetts Institute of Technology (MIT) in the development and deployment of the DSpace institutional repository system (<http://www.dspace.org/>), created in collaboration with the Hewlett Packard Corporation, has been a model pointing the way forward for many other universities. In 2003, with funding from The Andrew W. Mellon Foundation and other sources, MIT's DSpace is scheduled to be replicated at a number of additional institutions around the world; the software has also been released publicly under an open source arrangement, greatly lowering the cost and development barriers to implementing repositories for all institutions. The MIT software is not the only option available, although I believe it is the most general-purpose; for example, there is software from the University of Southampton in the U.K. (<http://www.eprints.org/>) designed more specifically for institutional or disciplinary repositories of papers, as opposed to arbitrary digital materials.

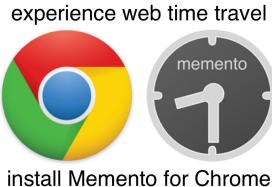
Over the past few months, I have had a number of opportunities to speak about the roles and significance of institutional repositories as a strategy for supporting the use of networked information to advance scholarship, notably at a workshop jointly sponsored by ARL, CNI, and SPARC in Washington, D.C., at the DSpace launch celebration at MIT, and at the University of Tennessee and the University of British Columbia. While video recordings of some of these

archived URI: <https://web.archive.org/web/20040909135231/http://www.arl.org/newsltr/226/ir.html>

# Tools

## Browsing Past Web

- Memento for Chrome
- Mink (Chrome)
- Wayback Machine (Firefox)



## On-Demand Archiving

- IA Save Page Now
- archive.is/archive.today
- Conifer



# IA Save Page Now (SPN)

The screenshot shows the Wayback Machine homepage. At the top, there's a toolbar with various icons. Below it, the Internet Archive logo and the Wayback Machine logo are displayed. A search bar contains the URL <http://>. To the right of the search bar are buttons for "BROWSE HISTORY" and "DONATE". A blue button encourages users to "Please try our Beta Wayback Machine - Now with Site Search!". Below this, a banner states "Explore more than 279 billion web pages saved over time" and shows a grid of thumbnail images of saved web pages.

**Tools**

- Wayback Machine Availability API
- Build your own tools.
- WordPress Broken Link Checker
- Banish broken links from your blog.
- 404 Handler for Webmasters
- Help users get where they were going.

**Subscription Service**

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit [Archive-It to build and browse the collections.](#)

**Save Page Now**

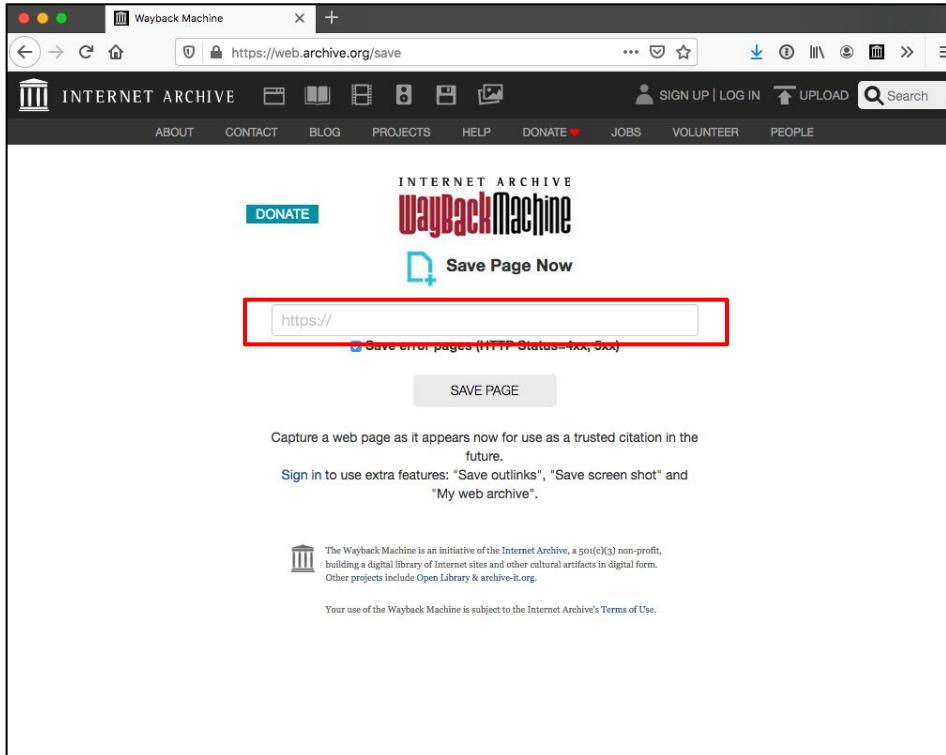
Capture a web page as it appears now for use as a trusted citation in the future.  
Only available for sites that allow crawlers.

A red arrow points from the text "If You See Something, Save Something" below to the "Save Page Now" section. A red circle highlights the "Save Page Now" button and its surrounding text.

"If You See Something, Save Something"

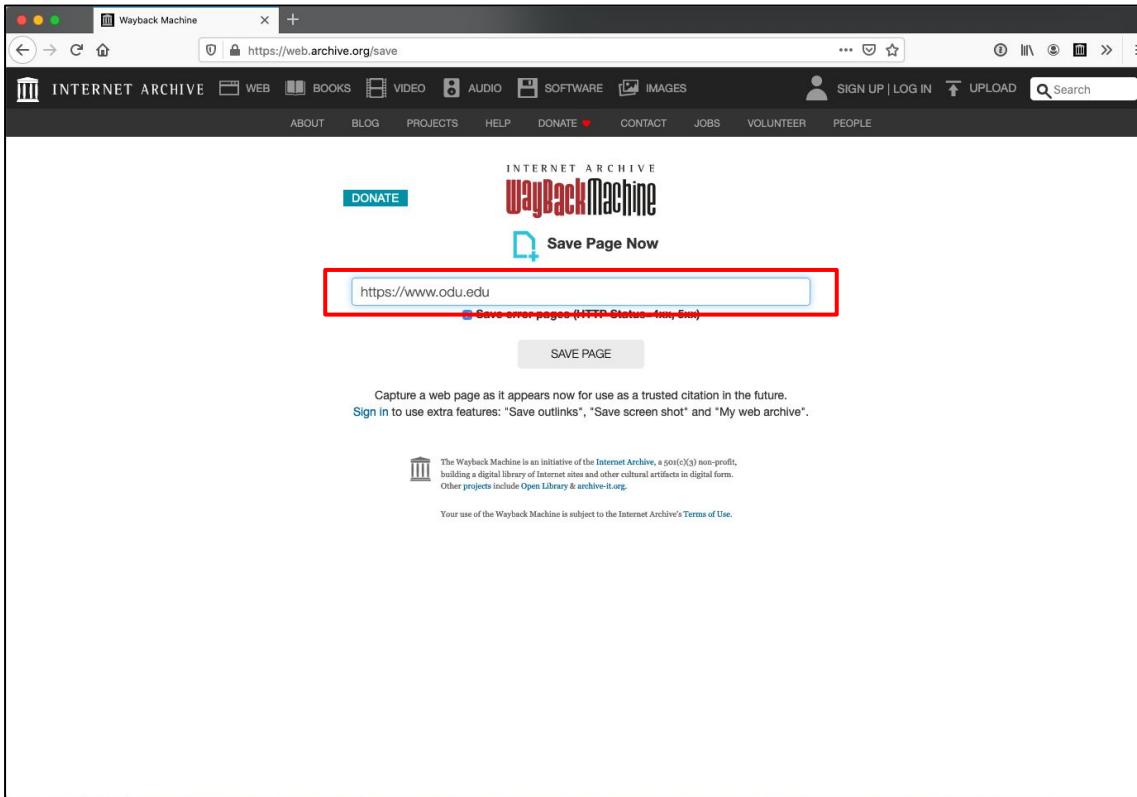
"The Wayback Machine's Save Page Now is New and Improved"

# Save Page Now (direct link)



[Save Page Now](#)

# SPN: Save ODU Homepage



# SPN: ODU Homepage Saved

The screenshot shows the Wayback Machine interface. At the top, there's a browser-like header with tabs, a URL bar showing <https://web.archive.org/save/https://www.odu.edu>, and a search bar. Below this is the Internet Archive navigation bar with links for WEB, BOOKS, VIDEO, AUDIO, SOFTWARE, and IMAGES, along with options to SIGN UP, LOG IN, UPLOAD, and a SEARCH bar.

The main content area features the Wayback Machine logo and a "DONATE" button. A green banner at the top says "Saving page https://www.odu.edu" followed by a "Done!" button. Below this, a message states "A snapshot was captured. Visit page: /web/20200914171203/https://www.odu.edu/".

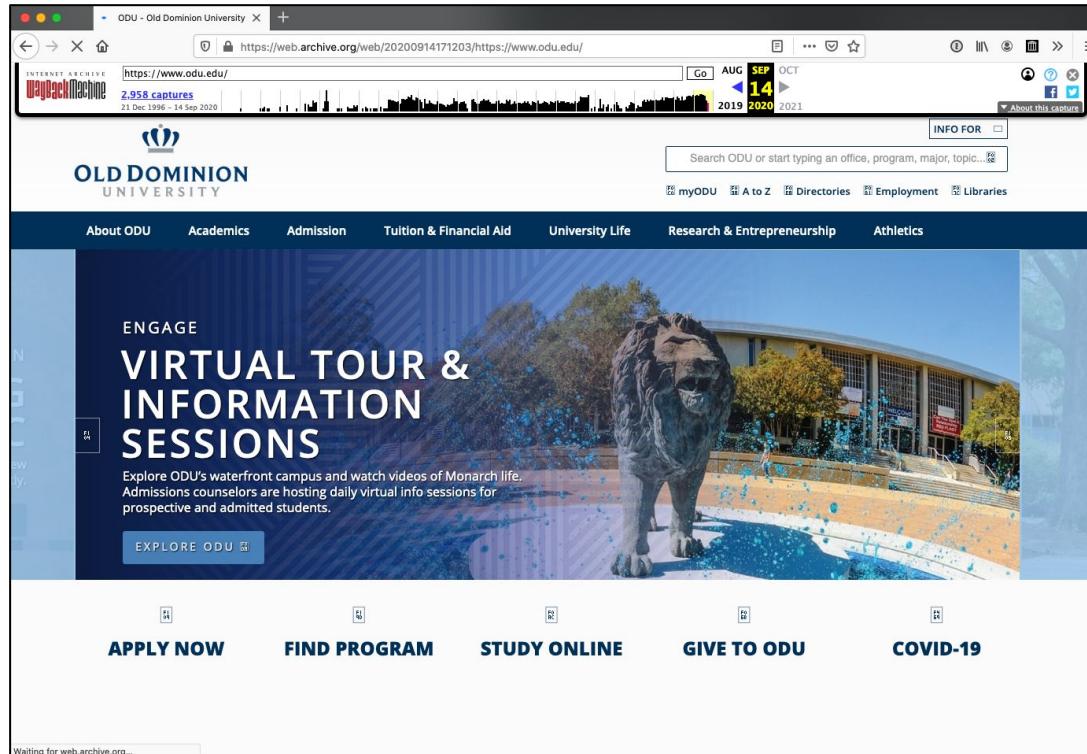
On the left, a list of URLs for the captured page is shown:

```
https://www.odu.edu/
https://kit.fontawesome.com/7e8bd7f817.js
https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/js/bootstrap.min.js
https://fonts.googleapis.com/css?family=Open+Sans:400,600,800,700|Open+Sans+Condensed:300
https://code.jquery.com/jquery-3.2.1.min.js
https://www.odu.edu/etc/designs/odu/clientlibs/libs/slick.min.css
https://www.odu.edu/etc/designs/odu/clientlibs/libs/jquery-migrate-3.0.0.min.js
https://www.odu.edu/etc/designs/odu/fonts/microgramma/MyFontsWebfontsKit.css
https://www.odu.edu/etc/designs/odu/clientlibs/libs/fotorama.min.css
https://hello.myfonts.net/count/37987b
```

A progress bar indicates "Downloaded elements: 95". Below the list is a "Return to Save Page Now" link. At the bottom, there's a note about the Wayback Machine being an initiative of the Internet Archive, a non-profit building a digital library of Internet sites and other cultural artifacts in digital form. It also mentions "Other projects include Open Library & archive-it.org." and the "Terms of Use".

The right side of the screenshot shows a similar interface for the same page, with a list of URLs identical to the one on the left, indicating a successful capture. The "Downloaded elements: 95" message is also present here.

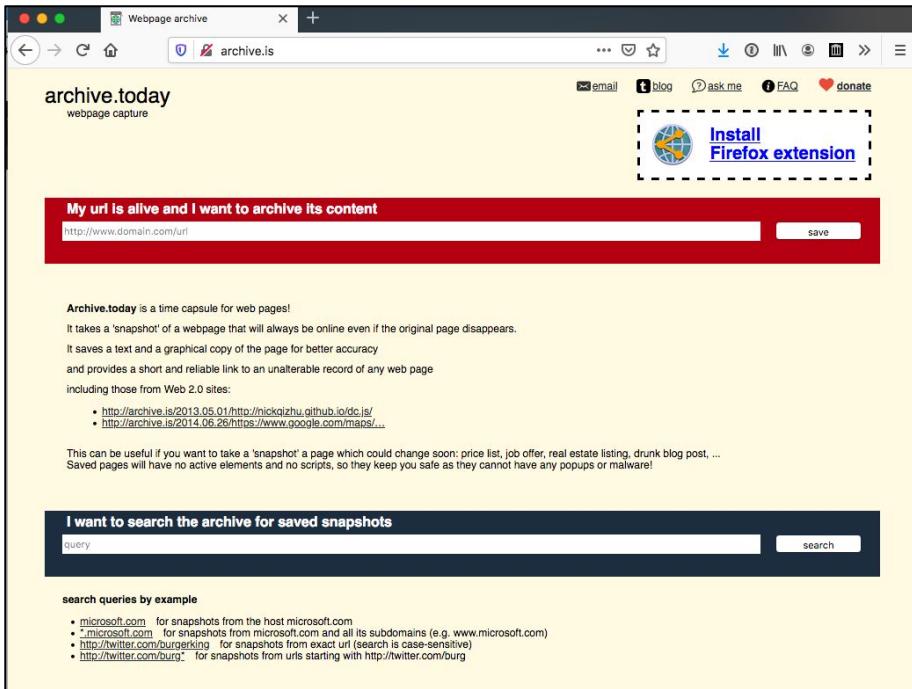
# SPN: Replay ODU Homepage



archived URI: <https://web.archive.org/web/20200914171203/https://www.odu.edu/>

# archive.is

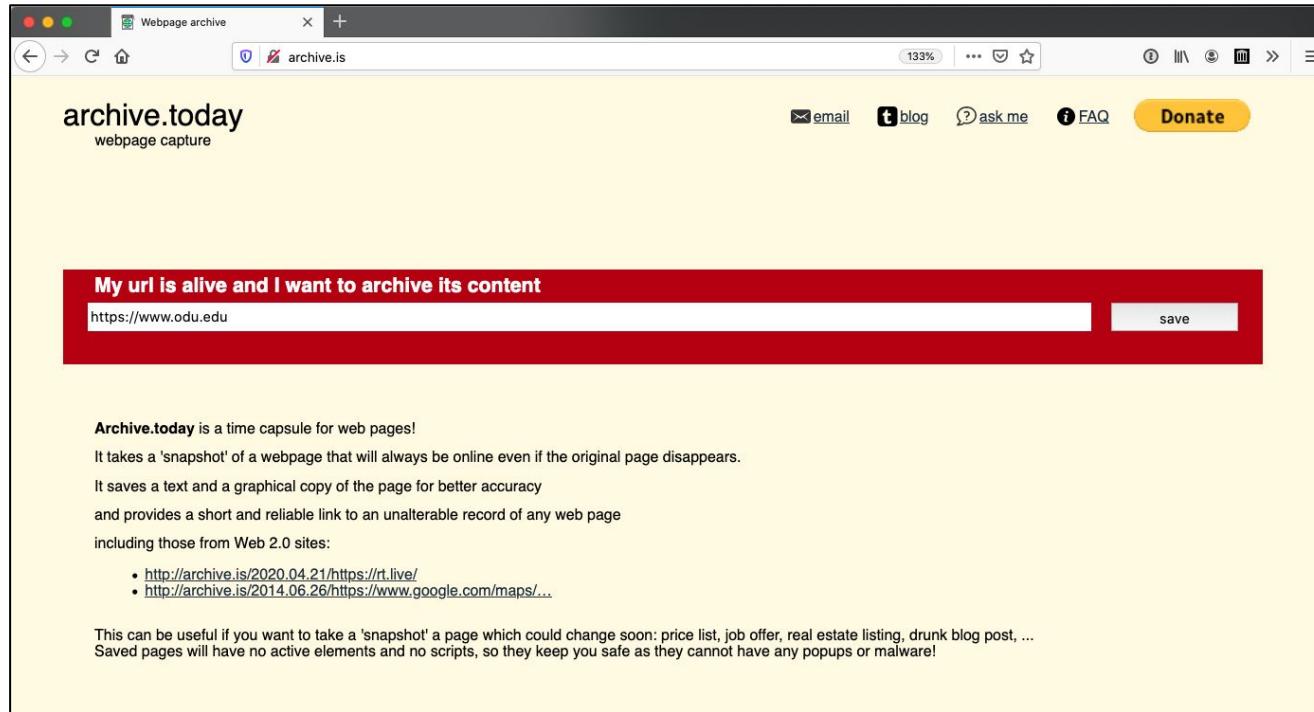
archive.today  
webpage capture



- On-demand web archiving
- Does not render or save JavaScript
- Also saves screenshot image
- Also known as archive.today

<http://archive.is/>

# archive.is: Save ODU Homepage



<http://archive.is/>

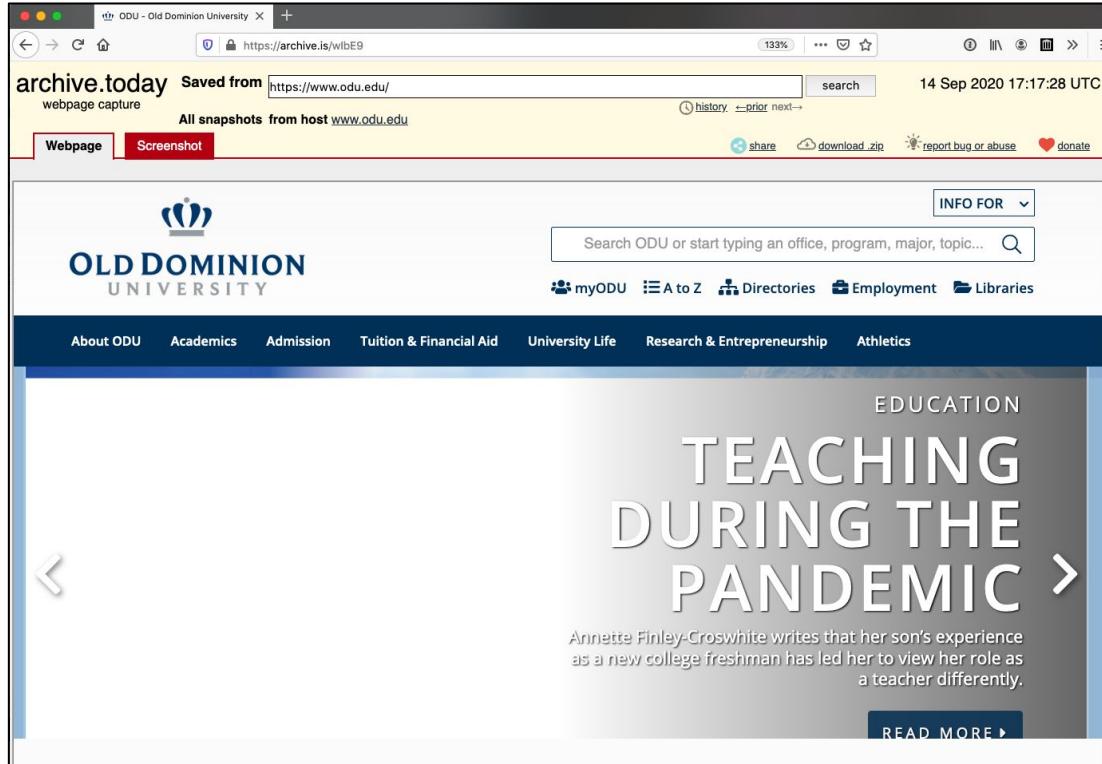
# archive.is: Saving ODU Homepage

The screenshot shows a web browser window with the URL <https://archive.is/wip/wlbE9>. The page title is "Loading...". Below the title is a table of file details:

status	type	size	url
200	text/html	23759	GET https://www.odu.edu/
200	application/javascript	30125	GET https://code.jquery.com/jquery-3.2.1.min.js
200	text/css	913	GET https://www.odu.edu/etc/designs/odu/fonts/microgramma/MyFontsWebfontsKit.css
200	text/css	477	GET https://fonts.googleapis.com/css?family=Open+Sans:400,600,800,700 Open+Sans+Condensed:300
200	text/javascript	2112	GET https://kit.fontawesome.com/7e8bd7f817.js
200	application/javascript	344	GET https://www.odu.edu/etc/designs/odu/clientlibs/libs/jquery-migrate-3.0.0.min.js
200	text/javascript	13105	GET https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/js/bootstrap.min.js
200	text/css	528	GET https://www.odu.edu/etc/designs/odu/clientlibs/libs/slick.min.css
200	text/css	138714	GET https://www.odu.edu/etc/designs/odu/clientlibs/min.css
200	text/css	4458	GET https://www.odu.edu/etc/designs/odu/clientlibs/libs/fotorama.min.css
200	text/css	20879	GET https://www.odu.edu/etc/designs/odu/clientlibs/homepage.min.css
200	application/javascript	15387	GET https://www.odu.edu/etc/designs/odu/clientlibs/libs/slick.min.js
200	application/javascript	105603	GET https://www.odu.edu/etc/designs/odu/clientlibs/min.js
200	application/javascript	22684	GET https://www.odu.edu/etc/designs/odu/clientlibs/libs/fotorama.min.js
200	application/javascript	12367	GET https://www.odu.edu/etc/designs/odu/clientlibs/homepage.min.js
200	text/css	17	GET https://www.odu.edu/etc/designs/odu/clientlibs/min.css
200	image/jpeg	453882	GET https://www.odu.edu/news/2020/9/virtual_engagement/_jcr_content/slideshow.image.jpg_2160.jpg/1592828835933.jpg
200	image/jpeg	494066	GET https://www.odu.edu/news/2020/9/admissions/_jcr_content/slideshow.image.jpg_2160.jpg/1599840032272.jpg
200	image/jpeg	322380	GET https://www.odu.edu/news/2020/9/health_science/_jcr_content/slideshow.image.jpg_2160.jpg/1600088615337.jpg
200	image/jpeg	394547	GET https://www.odu.edu/news/2020/9/athletics/_jcr_content/slideshow.image.jpg_2160.jpg/1600088322869.jpg
200	image/jpeg	320245	GET https://www.odu.edu/news/2020/9/education/_jcr_content/slideshow.image.jpg_2160.jpg/1600088118637.jpg
200	image/jpeg	132310	GET https://www.odu.edu/news/2020/4/virtual_engagement/_jcr_content/slidemobileimage.image.jpg_2160.jpg/1592828835933.jpg
200	image/jpeg	156867	GET https://www.odu.edu/news/2020/9/admissions/_jcr_content/slidemobileimage.image.jpg_2160.jpg/1599840032272.jpg
200	image/jpeg	88643	GET https://www.odu.edu/news/2020/9/health_science/_jcr_content/slidemobileimage.image.jpg_2160.jpg/1600088615337.jpg
200	image/jpeg	96132	GET https://www.odu.edu/news/2020/9/athletics/_jcr_content/slidemobileimage.image.jpg_2160.jpg/1600088322869.jpg
200	image/jpeg	90300	GET https://www.odu.edu/news/2020/9/education/_jcr_content/slidemobileimage.image.jpg_2160.jpg/1600088118637.jpg
200	application/javascript	5059	GET https://www.odu.edu/etc/designs/odu/clientlibs/min.js
200	image/png	352388	GET https://www.odu.edu/news/2020/9/nikole_hannah_jones/_jcr_content/highlightimage.image.jpg_2160.jpg/1599741590346.jpg
200	image/png	200602	GET https://www.odu.edu/news/2020/9/constitution_day_102/_jcr_content/highlightimage.image.jpg_2160.jpg/1599679897036.jpg
200	image/png	277074	GET https://www.odu.edu/news/2020/9/tom_barkin/_jcr_content/highlightimage.image.jpg_2160.jpg/1599848584903.jpg
200	image/png	362374	GET https://www.odu.edu/news/2020/9/luther_covid_19_and_he/_jcr_content/highlightimage.image.jpg_2160.jpg/1599159970756.jpg
200	image/png	355983	GET https://www.odu.edu/news/2020/9/sports/_jcr_content/highlightimage.image.jpg_2160.jpg/1599254576432.jpg
200	image/png	369179	GET https://www.odu.edu/news/2020/9/jared_benton/_jcr_content/highlightimage.image.jpg_2160.jpg/1599675254343.jpg
200	image/png	205071	GET https://www.odu.edu/news/2020/9/engineering_fellows/_jcr_content/highlightimage.image.jpg_2160.jpg/1599233321021.jpg
200	image/png	87177	GET https://www.odu.edu/news/2020/9/subsidence_strategies/_jcr_content/highlightimage.image.jpg_2160.png/1599163542459.png
200	image/png	278349	GET https://www.odu.edu/content/dam/odu/admin/univ/img/campus-shot-6.png/_jcr_content/renditions/cq5dam.thumbnail.319.319.png
200	image/png	306483	GET https://www.odu.edu/content/dam/odu/admin/univ/img/campus-shot-7.png/_jcr_content/renditions/cq5dam.thumbnail.319.319.png
200	image/png	224116	GET https://www.odu.edu/content/dam/odu/admin/univ/img/campus-shot-8.png/_jcr_content/renditions/cq5dam.thumbnail.319.319.png
200	image/png	221579	GET https://www.odu.edu/content/dam/odu/admin/univ/img/campus-shot-3.png/_jcr_content/renditions/cq5dam.thumbnail.319.319.png
200	image/png	234584	GET https://www.odu.edu/content/dam/odu/admin/univ/img/campus-shot-5.png/_jcr_content/renditions/cq5dam.thumbnail.319.319.png
200	image/png	277331	GET https://www.odu.edu/content/dam/odu/admin/univ/img/campus-shot-4.png/_jcr_content/renditions/cq5dam.thumbnail.319.319.png
200	image/png	17	GET https://www.archiveis.com/v1/0.1.0/2020/9/oduhomepage/contents/177167427212993_3354100250686481447_n.jpg?no_header=1&content=amt-2.1.chinastagram.com_nc_cat=105&nc_oh=&REPN=231kAXRM72TzAoh=5777cd41f4de76c523df1f5ee4893koeg5F88RAEF
200	text/plain	17	GET https://www.archiveis.com/v1/0.1.0/2020/9/oduhomepage/contents/177167427212993_3354100250686481447_n.jpg?no_header=1&content=amt-2.1.chinastagram.com_nc_cat=105&nc_oh=&REPN=231kAXRM72TzAoh=5777cd41f4de76c523df1f5ee4893koeg5F88RAEF
200	image/jpeg	26015	GET https://scontent-amt-2.1.chinastagram.com_nc_cat=105&nc_oh=&REPN=231kAXRM72TzAoh=5777cd41f4de76c523df1f5ee4893koeg5F88RAEF

<http://archive.is/>

# archive.is: Replay ODU Homepage



Archived webpage URI: <https://archive.is/wlbE9>

# Conifer



A screenshot of a web browser displaying the Conifer homepage. The address bar shows "Conifer | Homepage" and the URL "https://conifer.rhizome.org". The page features a large illustration of a wooden filing cabinet with several drawers, some of which are open to reveal documents. To the right of the illustration, the word "Conifer" is prominently displayed in a large, bold, serif font. Below it, the tagline "Collect and revisit web pages." is written in a smaller, sans-serif font. A paragraph of text explains what Conifer is: "Conifer is a web archiving service that creates an interactive copy of any web page that you browse, including content revealed by your interactions such as playing video and audio, scrolling, clicking buttons, and so forth." Below this text are two buttons: a green "Create a Free Account" button and a blue "Existing Users Login" link. At the bottom of the page, there is a note about free accounts and storage, followed by links for becoming a supporter and information about the software used.

- On-demand web archiving
- High-fidelity captures (renders and executes JavaScript)
- Previously known as Webrecorder.io

[Conifer | Homepage](#)

# Conifer: Save ODU Homepage

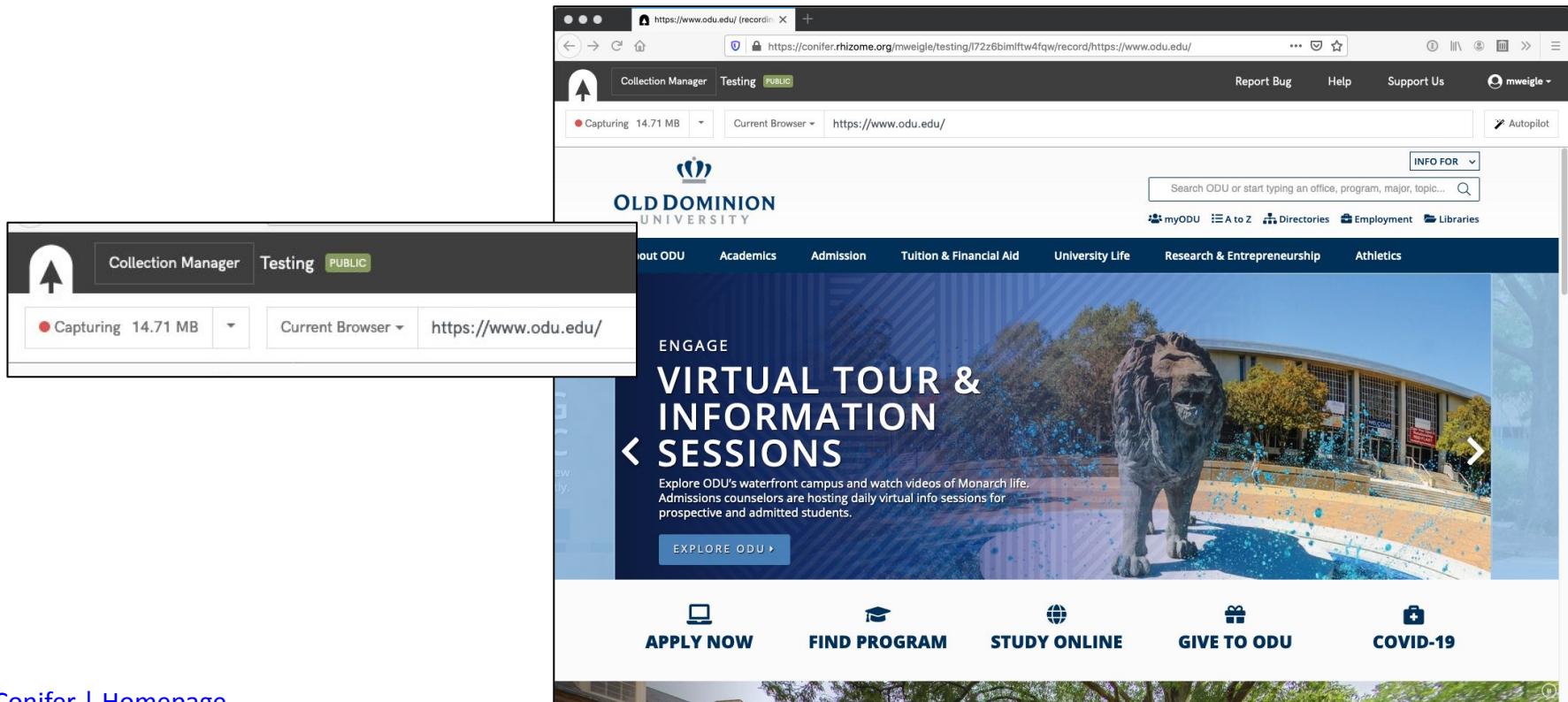
The screenshot shows the Conifer web application interface. At the top, there's a header bar with a back button, a search bar containing the URL <https://conifer.rhizome.org/mweigle>, and various navigation icons. Below the header, the Conifer logo is on the left, and a dark navigation bar contains links for "Report Bug", "Help", "Support Us", and a user account dropdown for "mweigle".

The main content area is titled "New Capture". It features a text input field with the URL <https://www.odu.edu>. To the right of this is a dropdown menu labeled "Testing" under the heading "Add to collection". Below these fields are two dropdown menus: "Select browser" set to "Use Current Browser" and "Session settings". A prominent green button labeled "Start Capture" is located at the bottom right of this section.

At the bottom of the main content area, there are two buttons: "+ New Collection" and "Upload". Below this, a "Bookmarks" section is shown, featuring a collection named "mweigle" which was created over a year ago, has a size of 17.99 MB, and is marked as "PUBLIC". This section also includes "Manage Collection" and "+ New Session" buttons.

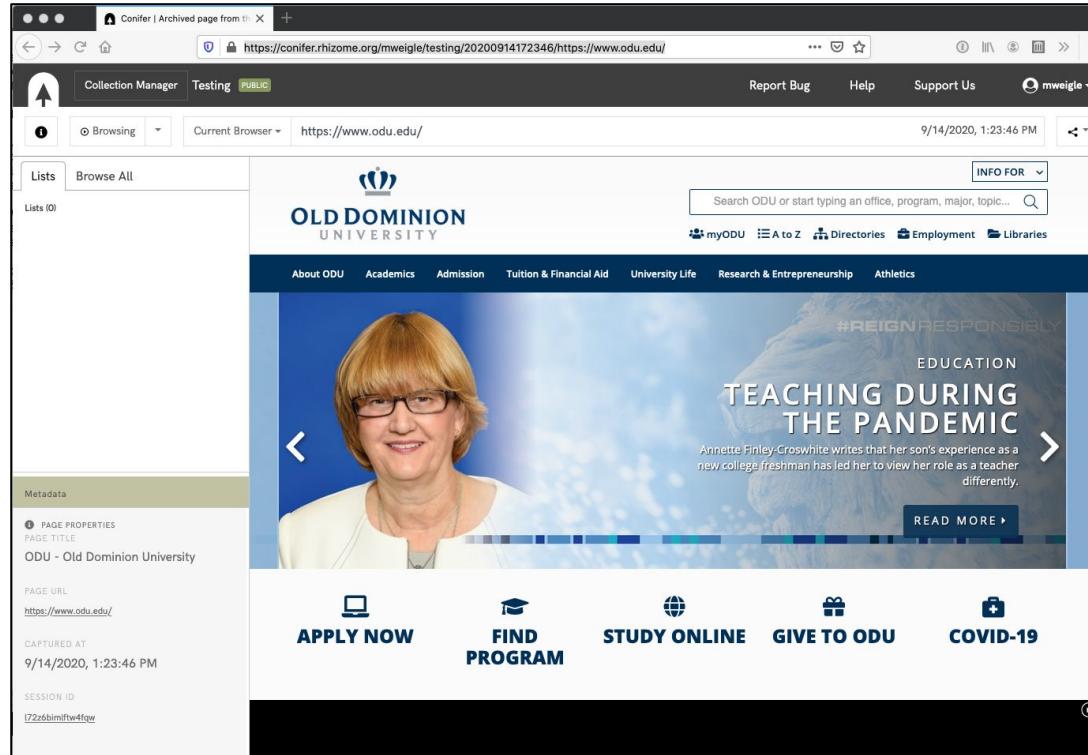
[Conifer | Homepage](#)

# Conifer: Saving ODU Homepage



[Conifer | Homepage](#)

# Conifer: Replay ODU Homepage



source URI: <https://conifer.rhizome.org/mweigle/testing/20200914172346/https://www.odu.edu/>

# Objectives

- Explain the importance of web archiving.
- Describe how to find and access archived webpages using the Internet Archive's Wayback Machine.
- Explain how the Memento Framework allows for content negotiation in the time dimension.
- Differentiate between a URI-R, URI-M, and URI-T in Memento terminology.
- Use the Memgator Memento Aggregator service to download a TimeMap and explain the contents of what is returned.