

# Archiving the Deepwater Horizon Oil Spill



Tracy Seneca  
California Digital Library

<http://was.cdlib.org>

# Archive Scope

**527 sites**

**10402 captures**

**May 5 to present**

tapering to less frequent captures of key sites,  
about 200 captures per month

**76 million + documents**

**2 TB**



The Web has revolutionized our access to information, but Web publications are fragile, and ready access to Web resources cannot be taken for granted. The Web Archiving Service enables librarians and scholars to meet that challenge.

### Information for

- Potential Partners
- Researchers
- WAS Curators
- Webmasters

### Search Archives

Subject experts have been capturing and preserving critical web sites to ensure that you have lasting access to web content. Public archives can be browsed or searched and provide persistent links to archived documents. Access to published archives is open to all. Click on an archive name on the right to search and view archived materials.

[Learn more](#) ►

### Build Archives

With a WAS account you can capture, analyze and archive web sites and documents. Archives can be published or kept for private study. The WAS curator tools are easy to use, fully hosted, and allow collaborative collection building. CDL provides [training and guidance](#) for WAS curators.

[Learn more](#) ►

### Featured Archive

#### Monterey Bay Area Local Government Web Archive



**Created by:** University Library, UC Santa Cruz

**Sites:** 86    **Oldest site:** 02/13/07    **Most recent site:** 11/04/10

**Description:** The Monterey Bay Local Government Project collection contains websites of local and regional government agencies in the greater Monterey Bay Area. Coverage includes the counties of Santa Cruz, Monterey, San Benito, San Mateo, and Santa Clara. Also included are websites for federal regional agencies... [more >](#)

#### At-a-glance

35 public archives  
16 partners  
2713 web sites  
281,038,597 documents  
15.4 TB of data

#### WAS News

- [Preview of Forthcoming Web Archiving Service Features](#)
- [Web Archiving Service: Blog Posts and Articles](#)
- [Recently Released: University of Michigan Web Archive](#)
- [See all](#) ►

### The Archives

Move pointer over an archive name for more information

- [2003 California Recall Election](#)
- [2007 Southern California Wildfires Web Archive](#)
- [2009 H1N1 Influenza A \(Swine Flu\) Outbreak](#)
- [2010 Winter Olympics](#)
- [AFL-CIO - Change to Win: the open web archive](#)
- [African Politics Web Archive](#)
- [Alternative Mass Media / News Web Sites Web Archive](#)
- [Anarchism Web Archive](#)
- [Arts & Cultural Left Web Archive](#)
- [California Political Blogs](#)
- [California State Government: .ca.gov Web Archive](#)
- [California Water Districts Web Archive](#)
- [Communism, Socialism, Trotskyism Web Archive](#)
- [Feminism & Women's Movements Web Archive](#)
- [Guantanamo Bay Detention Camp & War Crimes \(U.S.\) Web Archive](#)
- [International Development Organizations Web Archive](#)
- [Jewish American Left Web Archive](#)
- [Labor Unions and Organizations \(U.S.\) Web Archive](#)

# Archive Selection & Context

## Planned archives

- Advance subject expertise
- Time for evaluation
- Time for QA
- Focus on comprehensive capture
- Traditional collection development
- Control over scale

## Event archives

- Act quickly
- No one is the expert
- Collaboration required
- Every efficiency matters
- Frequent shallow captures / rapidly changing sites
- Massive scale

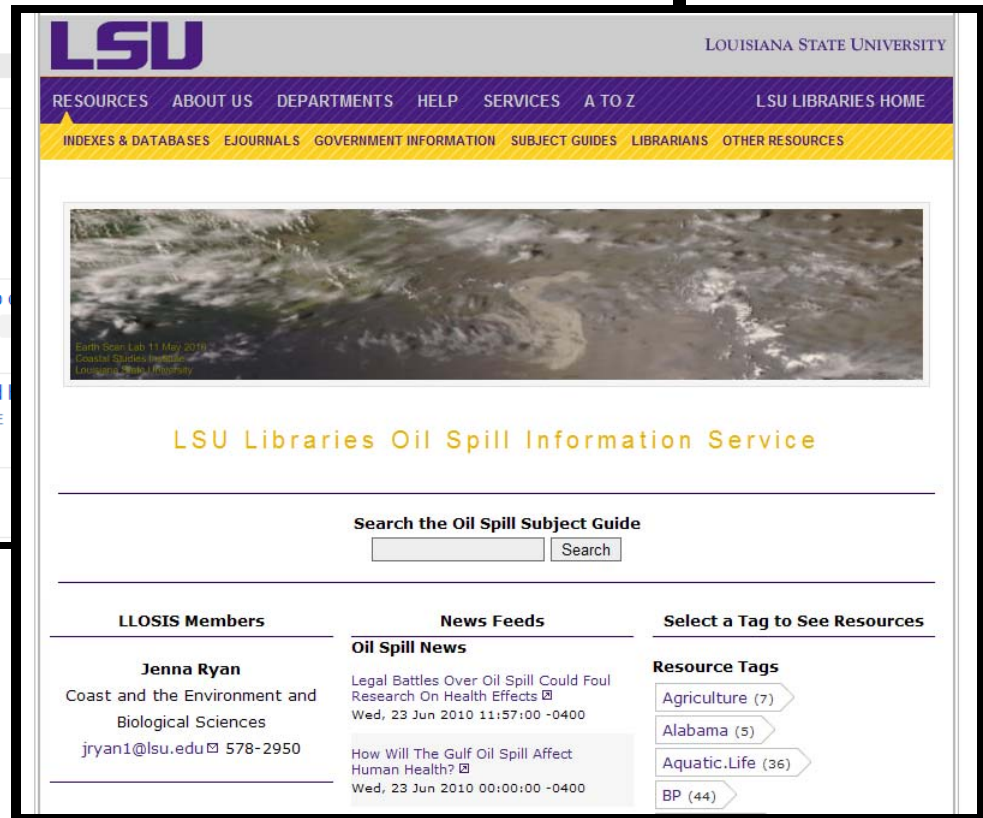
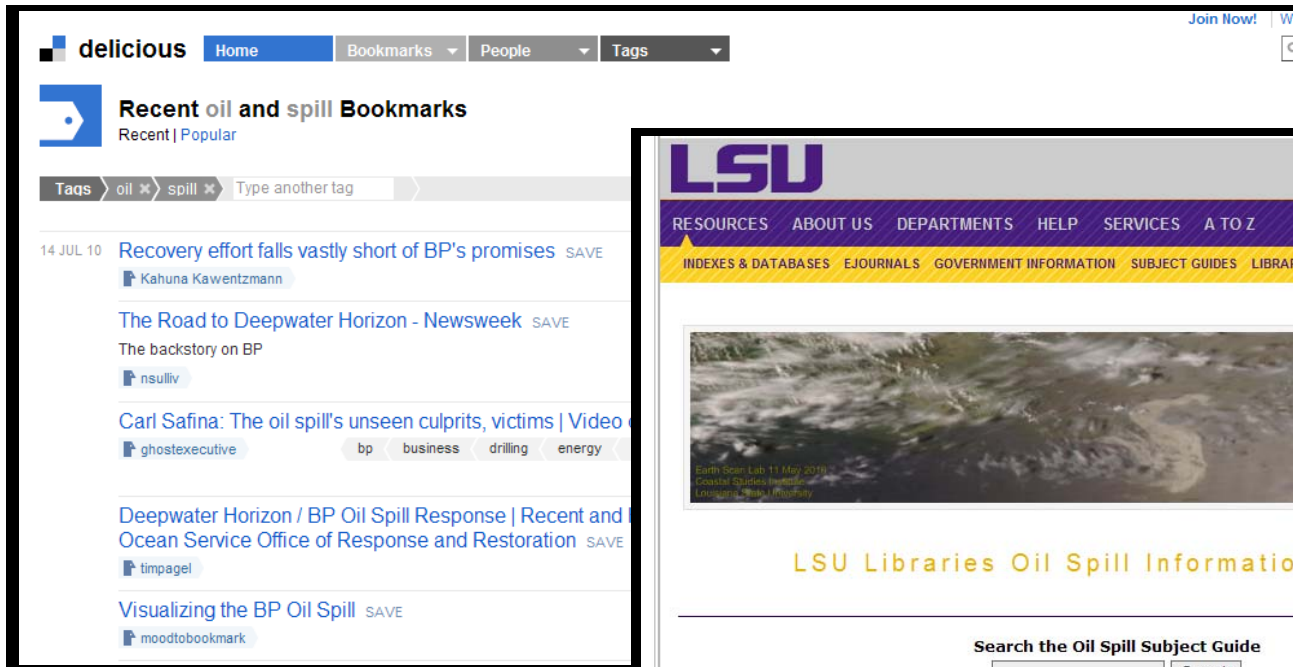
# 3 Challenges

- Site selection
- Site / capture management
- Quality assurance

# Getting Volunteers

- Tried bringing volunteers into service
  - “Add to WAS” browser button
- Tried external nomination tool
- TAP INTO WHAT USERS ARE ALREADY DOING

# LSU tags relevant sites in Delicious CDL imports Delicious JSON feed into WAS



~ 50% delicious  
~ 45% 1 curator  
~5% everything else

<http://was.cdlib.org>

# Site Management - From:

## Manage Sites Capture Status

1-10 of 159 ◀ Prev 1 2 3 4 ... 16 Next ▶ Limit:  Sort:

<input type="checkbox"/>	SITE NAME	STATUS	CAPTURES	ACTIONS
<input type="checkbox"/>	211 San Diego Wildfire Emergency	Finished Last captured 6 months ago	5	  <input type="button" value="Capture"/> 
<input type="checkbox"/>	Blog: And Still I Persist Conservative analysis of IT trends, climate cha...	Finished Last captured 6 months ago	4	  <input type="button" value="Capture"/> 
<input type="checkbox"/>	Blog: barboni.org	Finished Last captured 6 months ago	5	  <input type="button" value="Capture"/> 
<input type="checkbox"/>	Blog: California Fire News	Finished Last captured 6 months ago	5	  <input type="button" value="Capture"/> 

1-10 of 159 ◀ Prev 1 2 3 4 ... 16 Next ▶

**Fixed table**  
**Not enough control**  
**Few batch actions**

# To

## Manage Sites

1-25 of 518 ◀ Prev 1 2 3 4 ... 21 Next ▶

display: 25 | 50 | 100 | all  
brief records | **site name** | seed URL

Select all

sort by: **site name** | URL

1115.org West Coast Cap Peelers Blog

150 Dead Sea Turtles

2theadvocate.com | Powered by The Advocate and WBRZ News 2 Louisiana — Baton Rouge, LA

6 Ways to Help the Gulf Coast Today - The Causemopolitan

A Mennonite Disaster Service

ABC News: The Conversation: Oil Spill Blame Game

Absence of fatalities in blowouts encouraging in MMS study of OCS incidents 1992-2006

Acadiana RC&D Council

Active Leases and Infrastructure -- MMS

Agilent to show customers how to test for Gulf oil spill contamination

AGU: Gulf Gusher: Oil-spill science has advanced, but data is sparse

Alabama Coastal Foundation

### Limit site list by:

- All active sites
- In progress
- Never captured
- Inactive sites

Captured since:

month ▼ day ▼ year ▼

Tag:

Everything ▼

Scheduled:

Any ▼

Keyword:

- site name
- seed URL

limit

clear

## Manage Sites

display: 25 | 50 | 100 | all  
brief records | site name | seed URL

1-25 of 79 ◀ Prev 1 2 3 4 Next ▶

Select all

sort by: site name | URL

[2theadvocate.com | Powered by The Advocate and WBRZ News 2 Louisiana — Baton Rouge, LA](#) (86)

Seed URL: <http://2theadvocate.com>

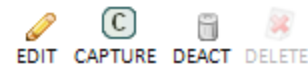
Tags: [Louisiana](#), [News](#)

Status: [Scheduled \(Monthly\)](#)

Current settings: Host site only, 1h

Last captured 15 days ago

Next capture in 15 days



[Archdiocese of New Orleans](#) (33)

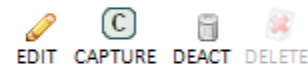
Seed URL: <http://www.archdiocese-no.org>

Tags: [Louisiana](#), [Non-profit Organizations](#)

Status: [Preserved](#)

Current settings: Host site only, 1h

Last captured about 1 month ago



[Army Corps of Engineers May 27, 2010 Permit for Louisiana's Barrier Island Plan](#) (1)

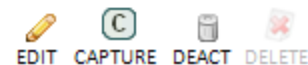
Seed URL: [http://www.lib.lsu.edu/ref/oilspill/Emergency\\_Permit\\_Documents.pdf](http://www.lib.lsu.edu/ref/oilspill/Emergency_Permit_Documents.pdf)

Tags: [Louisiana](#)

Status: [Preserved](#)

Current settings: Page only, 1h

Last captured 7 months ago



### Limit site list by:

- All active sites
- In progress
- Never captured
- Inactive sites

Captured since:

month ▼ day ▼ year ▼

Tag:

- Louisiana ▼
- Everything
- Untagged
- All tagged
- Blogs
- Business
- Education / Universities
- Environment
- Florida
- Government
- Louisiana**
- Mississippi
- News
- Non-profit Organizations
- Social Media
- Technology
- Texas

### Detailed Guides

- [Managing Sites: Guide](#)
- [Tagging Sites: Guide](#)



## Deepwater Horizon Oil Spill Web Archive

California Digital Library

[Home](#) [About](#) [Site List](#) [Search](#) [Help](#) [Contact Us](#)

### Refine site list

lookup by site name

#### Site list by topic:

- [Blogs](#)
- [Business](#)
- [Education / Universities](#)
- [Environment](#)**
- [Florida](#)
- [Government](#)
- [Louisiana](#)
- [Mississippi](#)
- [News](#)
- [Non-profit Organizations](#)
- [Social Media](#)
- [Technology](#)
- [Texas](#)
- 
- [Show All Topics](#)

### Showing sites with the topic *Environment*

- [Alabama Coastal Foundation](#) ⓘ [Show Info](#)
- [Alabama Coastal Foundation Facebook Page](#) ⓘ [Show Info](#)
- [American Birding Association](#) ⓘ [Show Info](#)
- [American Fisheries Society](#) ⓘ [Show Info](#)
- [Benthic Habitat Atlas of Coastal Texas - NOAA Digital Coast](#) ⓘ [Show Info](#)
- [C-CAP Land Cover Atlas - NOAA Digital Coast](#) ⓘ [Show Info](#)
- [CCA Louisiana - Coastal Conservation Association](#) ⓘ [Show Info](#)
- [Center for Biological Diversity - Gulf Disaster](#) ⓘ [Show Info](#)
- [City of Biloxi - Coast Clean-up Volunteers](#) ⓘ [Show Info](#)
- [Coalition to Restore Coastal Louisiana](#) ⓘ [Show Info](#)
- [Coastal Research Center: UC Santa Barbara](#) ⓘ [Show Info](#)
- [Commercial Fisheries Water Quality Sampling Methods & Procedures](#) ⓘ [Show Info](#)
- [Cool Green Science: the Nature Conservancy Blog](#) ⓘ [Show Info](#)
- [Craigslist sites for gulf coast cities](#) ⓘ [Show Info](#)
- [Earth Echo International: Blog](#) ⓘ [Show Info](#)
- [eBird](#) ⓘ [Show Info](#)
- [eBird Gulf Coast Oil Spill Bird Tracker](#) ⓘ [Show Info](#)
- [Economic Impacts to Fisheries and Coastal Habitat - GULFSEAGRANT.ORG](#) ⓘ [Show Info](#)

<a href="#">Sites</a>	<a href="#">Captures</a>	<a href="#">Administration</a>	<a href="#">Public Access</a>
-----------------------	--------------------------	--------------------------------	-------------------------------

## QA: Failed captures

1-25 of 51 ◀ Prev 1 2 3 Next ▶

display: 25 | 50 | 100 | all

<input type="checkbox"/> select all		status	actions
<input type="checkbox"/>	Congressman Allen Boyd : Florida 2nd District (1/11/2011 0:00) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE
<input type="checkbox"/>	Congressman Allen Boyd : Florida 2nd District (1/18/2011 0:00) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE
<input type="checkbox"/>	Congressman Allen Boyd : Florida 2nd District (1/25/2011 0:00) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE
<input type="checkbox"/>	Congressman Anh "Joseph" Cao (1/22/2011 20:58) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE
<input type="checkbox"/>	Congressman Anh "Joseph" Cao (1/15/2011 20:58) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE
<input type="checkbox"/>	Congressman Anh "Joseph" Cao (1/8/2011 20:58) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE
<input type="checkbox"/>	Congressman Charlie Melancon : Louisiana 3rd District (1/25/2011 6:49) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE
<input type="checkbox"/>	Congressman Charlie Melancon : Louisiana 3rd District (1/18/2011 6:49) settings: host, 36hr <b>scheduled: weekly</b>	failed	<a href="#">edit site</a>   <a href="#">view history</a>  DELETE

### All QA Reports

Failed captures ▼

### Tips

The sites in this list could not be successfully captured or were cancelled.

A status of "failed" or "error" usually occurs when the seed URL provided does not lead to a web page.

This can also happen if a site is temporarily down at the time you attempted to capture it.

You should check and correct it if needed.

You should delete the failed, errored or cancelled captures.

If the site no longer exists, you should turn off any scheduling options. You can also deactivate the site to remove it from your "Manage Sites" list. This

# Collection Observations

- Of ~350 sites from the Hurricane Katrina archive, only about 120 were initially relevant to the oil spill
  - Different responding organizations
- The relevant sites
  - Political offices / government agencies in the region
  - News sources in the region
  - Environmental organizations

# Manage Sites

1-58 of 58

display: 25 | 50 | 100 | all

[brief records](#) | [site name](#) | [seed URL](#)

Select all sort by: [site name](#) | [URL](#)

[150 Dead Sea Turtles](#)

[2theadvocate.com | Powered by The Advocate and WBRZ News 2 Louisiana — Baton Rouge, LA](#)

[ABC News: The Conversation: Oil Spill Blame Game](#)

[Anderson Cooper 360 Blog: Oil Spill](#)

[Anderson Cooper 360: Oil spill worse than expected](#)

[Atlanta News, Sports, Atlanta Weather, Business News | ajc.com](#)

[BBC - Oil Spill Coverage](#)

[BBC News - What is a 'top ki](#)

[Bill O'Reilly on Obama's resp](#)

[BP "Small People" Comment](#)

[BP and the Gulf of Mexico Oi](#)

[BP and the Gulf of Mexico Oi](#)

[BP and the Gulf of Mexico Oi](#)

[BP CEO Tony Hayward's test](#)

[BP to try again to cut off leak](#)

[BP: Hayward relieved of day-t](#)

[CBS News: Disaster in the Gulf](#)

[CNN.com: Oil Spill Coverage](#)

[Disaster News Network - Reporting disasters and response efforts](#)

[Disaster unfolds slowly in the Gulf of Mexico - The Big Picture ...](#)

[Exclusive: BP Chairman Talks To Sky](#)

## Site Summary: CNN.com: Oil Spill Coverage

Capture Settings

Scheduling

Descriptive Data

Capture History

**Seed URLs:** <http://www.ireport.com/ir-to...-stories.jsps?topicId=437986>  
<http://www.cnn.com/SPECIALS/2010/gulf.coast.oil.spill/>

**Scope:** Page + linked pages

**Max Time:** 1h

Edit Site

Capture Now

Add New Site

# Reminders

Use the tools you build

At larger scale than your users

Take advantage of existing workflows

*Collection building drives innovation*

# Next Steps

Release public archive

Review with Louisiana State University librarians

Web Archiving Service

- <http://was.cdlib.org>
- [www.facebook.com/webarchiving](http://www.facebook.com/webarchiving)

