Law Library of Congress

Legal Blawgs: Collecting for the Future.
Donna Scheeder
July 25, 2009
Law Library of Congress

• Why archive legal blogs?
  - Temporary
  - Relevant
  - Cited elsewhere
  - Enhance collection
  - Serve our mission
Law Library of Congress

- **A Collection of Blawgs**
  - Started capturing about 90 legal blogs in 2007
  - Added 38 in 2008
  - Goals:
    - Increase our international and comparative legal blogs
    - Anticipate capturing up to 200 blogs
Law Library of Congress

- Who worked on this collaboration?
  - Curators from the Law Library
  - Staff from web archiving in OSI
  - Staff from web services to design and integrate into LC/LL web pages
Law Library of Congress

- What did we need to do first?
  - Selection criteria
  - Organization
  - Workflow
Law Library of Congress

• How were blogs selected?
  - Variety
  - Authoritative
    • Frequently cited
    • Widely read
    • Honors or awards
  - Nominee Statistics
Law Library of Congress

- **Organization**
  - Categorize into bucket terms
  - Monitor
    - Still publishing
    - Scholarship
Law Library of Congress

Webcapture Team Mission

• Charged with
  - building a Library-wide process and technical infrastructure for capturing Web content.
  - identifying policy issues, establishing best practices and building tools to collect and preserve Web content.
The Library has been collecting Web content since 2000.
Eleven collections cataloged and accessible.
Six collections in post-processing phase.
Current ongoing collections include:
- Congressional/Legislative
- Indonesia General Election 2009
- Legal Blawgs
- Presidential Transition in Times of Crises

Partnered with the Internet Archive since 2000.
- Capture and access tool infrastructure
Manages and tracks the status of recommended websites from nomination to post-processing.
Law Library of Congress

Web Archive Permissions Process

Each collection has a permissions plan.

Legal Blawgs content owners must give us permission to crawl and all but one thus far have given permission to display off-site.

The tracking tool’s primary purpose is to manage the permission process.
**Law Library of Congress**

**Web Archive Goals**

**QUALITY WEB ARCHIVES**

Have to specify to the crawler what content to capture/links to follow.

Capturing the depth and breadth of a Web site can be an ongoing challenge in defining its boundaries.
A Metadata Objective Descriptive Schema (MODS) catalog record is created for each captured Web site.

Metadata comes from several sources:

- **From the tracking tool**
  - On or off-site permission status
  - Language
  - Primary URL

- **From the archive**
  - First and last capture date
  - HTML title, keywords, and abstract

- **From the catalogers**
  - Other subject cataloging.

- **From the Web capture team**
  - Any quality review comments to assist the catalogers.

Compiled into an XML file for cataloger review
Legal Blawgs Web Archive

**Title:** blackprof.com

**Date Captured:** September 6, 2007 - OPEN

**Subject(s):**
- Law--Blogs
- African American law teachers--Blogs
- United States--Race relations--Blogs

**Language(s):** English

**Genre:** Web site

**Access Condition:** None

**URL at time of capture:** http://www.blackprof.com/

**Citation ID:** http://www.blackprof.com/

**Record ID:** mrva0015.0057

**Collection Title:** Legal Blawgs Web Archive
Law Library of Congress

Web Integration Challenges

- How to integrate into loc.gov standard design?
  - Use standard design and conventions to help user navigate
  - Make visible in search results with other resources
  - Webarchives materials should appear in context – not be in a silo

- How to integrate into existing Law Library content?
  - In context with other specific resources
  - Use Law design and navigation
Law Library of Congress

Working with the Wayback Application Data

- **Interface Challenge:**
  How to scale for captures of varying frequency and amounts, from 1 time only to daily over the course of many years.
Law Library of Congress

Working with the Wayback Application Data

• Our Approach: use a modified calendar format that breaks the archive for a site down by year, and for each year, into months.
  - Uses well known organizational scheme
  - Uses date formats compatible with Wayback
  - Works with LOC standard design
The limit of 12 months in a year and max of 31 links in a month provides a browseable view of when the captures were made and link to the individual captures in the archive.

In cases of multiple captures in a day, a date links to a page with all the captures for that day.

This approach puts some extra clicks between the identification of the blog and the view of the archived page - But, the path is clear, organized, and leverages a familiar organizing principle - the calendar.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>2007</td>
</tr>
</tbody>
</table>

Legal Blawgs Web Archive

http://aidpblog.org/

*denotes when site was updated

**2008**

<table>
<thead>
<tr>
<th>Month</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>01/03/08 *</td>
</tr>
<tr>
<td>February</td>
<td>02/07/08 *</td>
</tr>
<tr>
<td>March</td>
<td>03/06/08 *</td>
</tr>
<tr>
<td>April</td>
<td>04/03/08 *</td>
</tr>
<tr>
<td>May</td>
<td>05/01/08 *</td>
</tr>
<tr>
<td>June</td>
<td>06/05/08 *</td>
</tr>
<tr>
<td>July</td>
<td>07/03/08 *</td>
</tr>
<tr>
<td>August</td>
<td>08/07/08 *</td>
</tr>
<tr>
<td>September</td>
<td>09/04/08 *</td>
</tr>
<tr>
<td>October</td>
<td>10/02/08 *</td>
</tr>
<tr>
<td>November</td>
<td>11/06/08 *</td>
</tr>
<tr>
<td>December</td>
<td>12/05/08 *</td>
</tr>
</tbody>
</table>

**Total:** 24 archived captures
Law Library of Congress

Header Bar

- Establishes LOC/Law Library as the source of the content
- Shows where capture lives within the collection
- Allows user to navigate through the collection from the archived site, without resorting to bouncing back and forth between the Law Library and archived site
- User can hide header
Law Library of Congress

*Future – Enhanced Searching, Bib Records, etc*

- **Planned Enhancements:**
  - Conversion of the Web Archives site to standard design
  - Allow searching and browsing of catalog/bibliographic records
  - Better integration with other collections
Law Library of Congress

• Future challenges: What can you do?
  - Collect and preserve if you have the resources.
  - Help us to improve the site: Nominate blawgs, submit comments
  - Recommendations for guidelines on blawg administration
  - Standardize some practices
Law Library of Congress

• Biggest challenge ahead will be to keep pace with changes and trends in scholarly publishing. Books published a chapter at a time and articles published one by one are in our near future. Will Librarians need to become aggregators in order to collect and preserve?
Law Library of Congress
Contact Information

Gina Jones, gjon@loc.gov, (202) 707-6604

William Kellum, wike@loc.gov, (202) 707-9752

Andrew Weber, anwe@loc.gov, (202) 707-4758