

Web Archiving Update

DLF Fall 2006 Forum Panel

Web Archiving Service Release 1 Demonstration



Tracy Seneca
California Digital Library

National Digital Information Infrastructure Preservation Program
Library of Congress

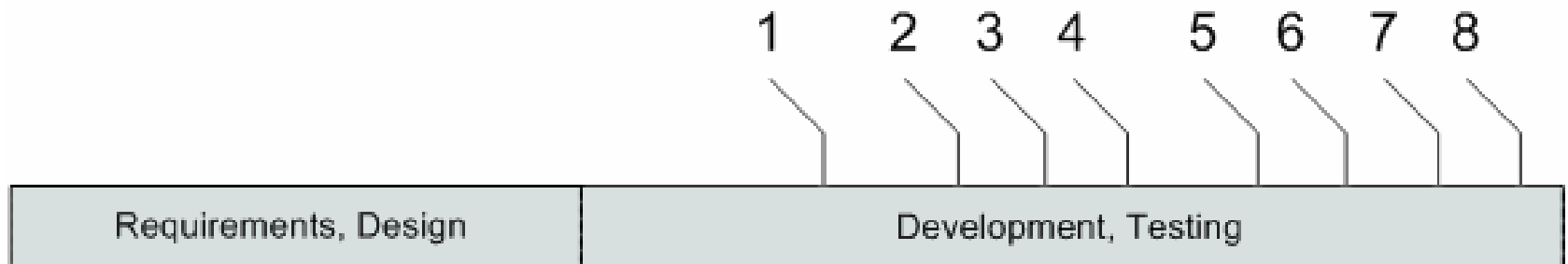
Grant Scope

Jan 2005 – Jan 2008

- Build tools to allow librarians to capture, curate and preserve web-based government and political information.
 - Create topical and event-based archives
 - Capture individual sites and documents
- Assess the impact of these tools on traditional collection development practices.
- Explore web archiving service sustainability.

Web Archiving Service Development Path

Jan 2005 – Dec 2007



Release 1: Basic Capture

Release 2: Search & Display

Release 3: Analysis & Reports

Release 4: Collection Building

Release 5: Admin & Rights

Release 6: Event Capture

Release 7: Preservation

Release 8: Enhancements

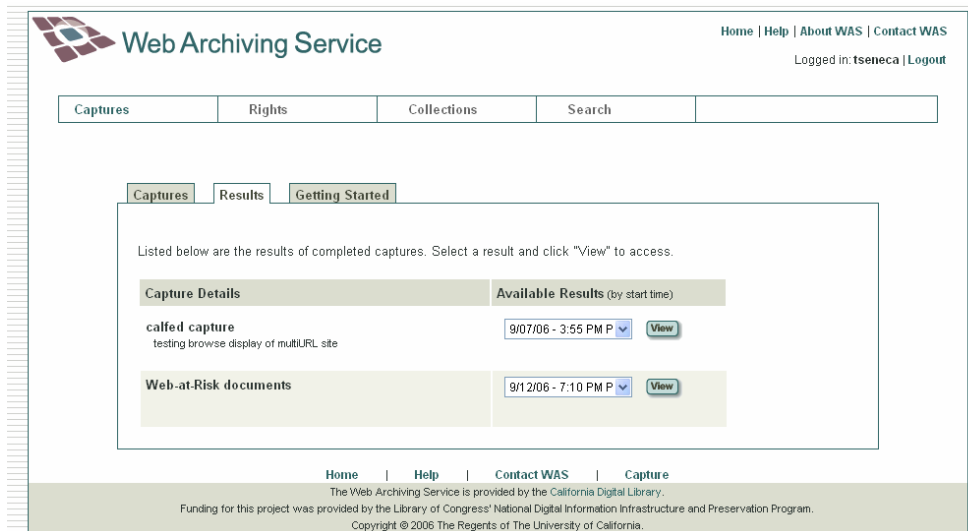


Release 1: Basic Capture

Issue one-time web captures

Review default capture reports

Display results using interim display tool



The screenshot displays the Web Archiving Service (WAS) web interface. At the top, the WAS logo is on the left, and navigation links (Home, Help, About WAS, Contact WAS) are on the right. Below the navigation, a horizontal menu contains 'Captures', 'Rights', 'Collections', and 'Search'. The 'Captures' tab is active, showing sub-tabs for 'Captures', 'Results', and 'Getting Started'. The main content area lists completed captures with instructions to select a result and click 'View'. Two capture entries are shown: 'calfed capture' (testing browse display of multiURL site) and 'Web-at-Risk documents'. Each entry has a date/time dropdown and a 'View' button. The footer contains additional navigation links and funding information.

Web Archiving Service

Home | Help | About WAS | Contact WAS

Logged in: tseneca | Logout

Captures Rights Collections Search

Captures Results Getting Started

Listed below are the results of completed captures. Select a result and click "View" to access.

Capture Details	Available Results (by start time)
calfed capture testing browse display of multiURL site	9/07/06 - 3:55 PM P View
Web-at-Risk documents	9/12/06 - 7:10 PM P View

Home | Help | Contact WAS | Capture

The Web Archiving Service is provided by the California Digital Library.
Funding for this project was provided by the Library of Congress' National Digital Information Infrastructure and Preservation Program.
Copyright © 2006 The Regents of The University of California.

Release 1 Results

- **Overall: successful, intuitive**
- **Hard to judge crawler effectiveness due to**
 - time limitations on capture settings
 - difficult to interpret default reports
- **Need to link tasks more closely**
- **Need faster UI development environment**
 - Moving from JSP/Struts to Ruby on Rails
 - Backend will remain in Java
 - Connected via SOAP layer



Release 2: Search & Display

Integrated search and display tools

Improvements to site navigation

Ability to browse capture results

Release 3: Analysis & Reports

Integrated search and display tools

Improved reports of capture results



Release 4: Collections

Create collections

Add captured materials to collections

Remove materials from collections

Search collections

Release 5: Rights & Admin

Rights management features &
reports

“My Account” features



Rights Features

- Designate rights status by site
 - Free to capture
 - Notification only
 - Permission required
- Record contact information and other rights metadata
- Track permission status



Release 6: Event Capture

Allow public site nominations

Nomination approval / rejection

Co-curation features

Quick-capture form



Release 7: Preservation

Generate alternate formats

HTML > Plain text

HTML > JPG

Checksum reports

Addition of preservation metadata



Further Information

- Expanding the CDL Digital Preservation Repository for New Projects

Shifra Pride Raffel, David Loy

Session 8. 2:30PM–4:00 Thursday

Ballroom Foyer, Main Lobby Level

- <http://wiki.cdlib.org/WebAtRisk>

