Web Archiving Update
DLF Fall 2006 Forum Panel
Web Archiving Service Release 1 Demonstration

Tracy Seneca
California Digital Library
Grant Scope

• 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

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
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
<table>
<thead>
<tr>
<th>Release 2: Search &amp; Display</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated search and display tools</td>
</tr>
<tr>
<td>Improvements to site navigation</td>
</tr>
<tr>
<td>Ability to browse capture results</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Release 3: Analysis &amp; Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated search and display tools</td>
</tr>
<tr>
<td>Improved reports of capture results</td>
</tr>
</tbody>
</table>
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