Turning Challenges into Opportunities:

How Law Libraries Can Capture and Preserve Government Web Resources

AALL Annual Meeting
July 17, 2007
Cathy Nelson Hartman, University of North Texas
chartman@library.unt.edu
Overview of Presentation

• Background – New issues encountered when collecting web published resources
• Research
• Collection Development Planning Tool
Background
The Web Environment

- Brings significant changes and challenges for Collection Development
- Struggle to meet service demands of customers
- Realign resources to meet new demands and challenges
Traditional Collection Development Processes

- Select
- Acquire
- Describe
- Organize
- Present for access
- Curate – Maintain, Deselect, Preserve
Web Environment

• Same Collection Development/Planning processes may be followed
• New challenges are encountered in every process
Selection of Materials

- Establish Authority
  - Peer Review Process
  - Reputation of Publisher
  - Reputation of Author
Print on Paper

Selection (cont.)

– Established processes for selection
  • Approval plans
  • Depository Library item selections
  • Vendors / Distributors
Web Environment

Selection of Materials

– What will users/researchers value?
– Should we collect text, video/audio, multimedia versions or all versions available?
Columbia Accident Investigation Board – 2nd Public Hearing

Dr. William Ailor, director of the Center for Orbital Reentry Debris Studies from Aerospace Corporation

"One of the things that we did was we were able to get a portion of this fragment that brushed Lottie Williams on the shoulder and we actually wanted to find out if, in fact, it did come from the launch vehicle or from that vehicle. We analyzed that and found that – if you take the next page – that it did. The trajectory time was consistent. She was out walking at around 3:30 in the morning and actually saw the reentry and then this thing came down and brushed her on the shoulder and she recovered that."
Web Environment

Selection of Materials (Continued)

Compare with the video of the hearing

..\Desktop\20030317_nasa_80k.wmv
Web Environment

Selection (continued)
– How will authority of Web published materials be established?
– How will we locate/identify Web published materials?
Acquisition of Materials

– Publication/item begins and ends with cover of book or other container
– Established practices for acquiring material facilitated by vendors
– Authenticity/stability
Web Environment

Acquisition of Materials

– Where does a publication/item begin and end – Unit of Selection/Acquisition?
  • Links?
  • Databases?
  • Multi-media presentations?

– Can we ensure accuracy of captured material?
Describe and Organize

MARC Records – Cooperative Cataloging
(subject/name authority)

Version control
– Changes/updates made
– Item republished as new edition
– Tracked through cataloging/metadata
– Shelved side-by-side on book shelves
Web Environment

Describe and Organize

Metadata – automatic generation?

Less precise Subject control, name authority, etc.

Version control

– Frequent changes made
– Discovery of changed publications/records
– Define a new edition – how many changes required?
– Track changes in metadata
– Serendipitous discovery by browsing less likely
Web Environment

Presentation of Materials to Users

Access rights management

• Licensed products with restrictions
  – Specific user group
  – Many simultaneous users

• Determine access rights for captured material

Usability issues
Print on Paper

Maintain and Deselect

– Shelve, repair, provide reference services
– Weed collections
Web Environment

Maintain and Deselect
- Storage
- Refresh, migrate
- Provide reference services
- Weed collections
Print on Paper

Preservation

– Well developed strategies in place
– Circulating Collections
  • Repairs as needed
  • Acid paper issues
  • Shelf life 75 + years
– Curation of rare materials
Web Environment

Preservation
– Should begin with creation of publication
– Constantly changing technology environments, both hardware and software
– How many copies are needed?
– Standards development – metadata, text mark-up
– Shelf life - ??
Web Environment Issue

Privacy
– Access greatly increased with full-text searching
– Safety issues
– Hackers
– User tracking

Authenticity
– Must begin with creator
– AALL Leadership
Research
Looking for Solutions

National Digital Information Infrastructure and Preservation Program (NDIIPP)

• Sponsored research project – The Web-at-Risk

• Partners include:
  California Digital Library, University of North Texas, and New York University
Data Collection

- **Focus Groups**
  - ALA Chicago, Depository Library Council
  - Librarians - UC System, UNT, NYU

- **Interviews**
  - Data owners
  - Researchers

- **Curator Surveys**

What we learned......
Possible Solutions

- Creation of a Web Archiving Service
- Models for collection planning for Web published materials
Web Archiving Service

- Project creating tools for a Web Archiving Service (WAS) at CDL
- Each tool release assessment (UNT) improves next release
- Tools to be adapted to opensource digital asset management systems – DSpace, Fedora, etc.
- Fee-based services now available from Archive-It.org and OCLC
Collection Planning Tools

- Collection Plan Template and
- Web Collection Plan Overview: Considerations for Project Curators
- Linked from:
  http://web3.unt.edu/webatrisk/cpg.php
Template Organization

Section 1. Mission & Scope
Section 2. Selection
Section 3. Acquisition
Section 4. Descriptive Metadata
Section 5. Presentation and Access
Section 6. Maintenance and Weeding
Section 7. Preservation
Appendix A. Submission Agreements
Appendix B. Web Archiving Service Agreement
Appendix C. Collaboration Agreements
CyberCemetery Collection Plan

Scope - The CyberCemetery is a collection of government websites that have ceased operation (usually websites of defunct government agencies and commissions that have issued a final report). This collection features a variety of topics indicative of the broad nature of government information. In particular, this collection features websites that cover topics supporting the university’s curriculum and particular program strengths.
Section 4. Descriptive Metadata

A. Level of description
   Website level description provides ample subject access to the information.

B. Metadata elements
   Our local metadata schema is listed in Appendix D: UNT Metadata Schema (Descriptive Metadata).

C. Controlled vocabularies
   The CyberCemetery collection will use the Legislative Indexing Vocabulary (LIV) for subject terms. The Legislative Indexing Vocabulary is a thesaurus of over 5,000 terms created by the Congressional Research Service to describe legislative and public policy literature.
Comments or Suggestions

Your comments on the adaptability of the collection planning tools for law libraries appreciated.

chartman@library.unt.edu
More Information

Web Site:  http://web2.unt.edu/webatrisk/

Select “Reports” to see reports of all data collection activities. Select “Collection Plans” to see guidelines and plans from participating institutions.

Project Wiki:  
http://wiki.cdlib.org/WebAtRisk/

Contact:  Chartman@library.unt.edu

Thanks to the University of North Texas NDIIPP Team: Kathleen Murray, Assessment Analyst, Mark Phillips, Head, Digital Projects Unit, Bill Moen, SLIS faculty/ project advisor