



NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM

**The Web-at-Risk:
A Distributed Approach to Preserving our Nation's Political Cultural Heritage**

Content Identification, Selection, and Acquisition Path

Focus Group Report:

Federal Depository Library Conference - Washington DC - October 2005

March 28, 2006

Prepared by:

Kathleen R. Murray
Assessment Analyst, Web-at-Risk Project
University of North Texas
krmurray@unt.edu

Inga Hsieh
Research Assistant, Web-at-Risk Project
University of North Texas
ikh0003@unt.edu

The following people contributed to this document.

University of North Texas
California Digital Library

Valerie Glenn
Tracy Seneca

Contents

1 Introduction..... 3

2 Methodology..... 3

 2.1 Framework..... 3

 2.2 Participants..... 3

 2.3 Data Collection..... 4

 2.4 Data Analysis..... 4

3 Findings..... 4

 3.1 Policy..... 4

 3.2 Selection..... 6

 3.3 Acquisition..... 8

 3.4 Organization..... 9

 3.5 Presentation..... 9

 3.6 Preservation..... 10

4 Discussion..... 10

 4.1 Dealing with Change..... 10

 4.2 What to Preserve..... 11

 4.3 Needs & Issues..... 12

 4.4 Need for Collaboration..... 13

Appendix A. Collection Development for Web Archives..... 15

Appendix B. Participants..... 16

Appendix C. Participant Questionnaire..... 17

1 Introduction

The Web-at-Risk project is one of eight digital preservation projects funded in 2004 by the Library of Congress. The project is a 3-year collaborative effort of the California Digital Library, the University of North Texas, and New York University. The project will develop a Web Archiving Service that enables curators to build, store, and manage collections of web-published materials in distributed repositories located at the three project partner sites. The project will also produce tools and guidelines to assist curators and other information professionals with collection development for web archives.

In support of this effort five focus groups were held in 2005. The purpose of the focus groups was to elicit the needs and issues librarians, curators, and end-users have in relation to web archives. This document summarizes the discussion held on October 17, 2005 at the Federal Depository Library Conference in Washington, DC. The one and one-half hour discussion was facilitated by the Assessment Analyst for the Web-at-Risk project.

The report includes the following three sections: (a) the methodology used to conduct the focus groups and analyze the data, (b) the detailed results of the analysis organized into phases of the collection development process, and (c) a discussion of the key findings.

2 Methodology

2.1 Framework

Collection development for web archives includes three major phases: selection, curation, and preservation. By breaking down collection development into a series of activities within each phase, the functional view shown in Table 1 emerges. Librarians will recognize the activities as those commonly employed in collection planning. (Appendix A provides a brief explanation of the activities in each phase as they apply to collection development for web archives.)

Table 1. Collection Development Framework for Web Archives

PHASES		
SELECTION	CURATION	PRESERVATION
Selection	Description	Preservation
Acquisition	Organization	
	Presentation	
	Maintenance	
	Deselection	

2.2 Participants

A total of nine librarians participated in the group discussion (Appendix B). All participants were attending the Federal Depository Library Conference and worked in government information or government documents positions. Six of the nine participants worked at university libraries, two at university law libraries, and one worked in a state library. Four participants held managerial positions.

Several participants were familiar with web archiving and related technologies. Four participants indicated they had experience building web archives, either within their institutions or in collaborations with other organizations or government agencies. One library is using DSpace and another subscribes to OCLC's Digital Archive service.

Three of the remaining five participants were in various stages of participation, planning, or proposing archival projects or programs either within their institutions or with other institutions or government agencies. One library is involved in a LOCKSS project, a few had experience capturing and archiving websites of interest to faculty members on CDs, and some have created print archives of web-published documents.

2.3 Data Collection

The discussion was recorded and subsequently transcribed. Additionally, two note-takers attended the focus group and created records of the discussion as well as summaries of the key points that emerged. Participants completed a questionnaire (Appendix C) that identified demographic characteristics and captured their thoughts regarding:

- User needs addressed by an archive
- Critical areas their organization needs to address to successfully implement a web archive
- Hurdles their organization faces in creating an archive

2.4 Data Analysis

Collection development provided the overall framework (Appendix A) for analyzing the focus group discussion. Based on a discussion in May of 2005 with curators involved with the Web-at-Risk project, an initial categorization of concerns and issues within each collection development phase was created.

These categories were used to analyze the content of the first focus group. Additional categories were added as necessary. This process was repeated for each of the four focus groups that followed.

Two analysts categorized the transcripts and notes from each focus group. Discrepancies between the analysts were discussed and resolved.

3 Findings

3.1 Policy

Collection Policies, Practices, & Plans

- State libraries have mandates to preserve/archive government publications. One participant's library is working now to modify their plan to include preservation of web-published materials.
 - Quote: "We are currently participating in a department-wide effort to produce an electronic records in publication plan to deal with both the records side and the public, or web-published materials."
- There was general agreement that long-term preservation of web-published materials require new guidelines. However, most have not yet included these in their plans.
 - Quote: "We have a statement about where that [web materials] goes in the catalog."
 - Quote: "Our main libraries right now are working on a digital preservation policy." This will include:

- Subject selectors will need to consider long term preservation when they select.
- Training will accompany this policy.
- Quote: “We are revising our collection development policy because we now have an electronic resources and services council that is devising all the policies for campus-wide electronic information.”
- Depository libraries are grappling with preservation and archiving of documents published only electronically.
 - Quote: “Our statute still reads that if it exists in print, we get it. But if print doesn’t exist then there’s a whole other set of [issues].”
 - Quote: “We select at 80%. So what does long term preservation mean for us? Are we going to preserve that... whatever percentage of that (it’s a large percentage) is e-only?”
- Contractual obligations affect collection policies and practices.
 - Examples: Depository library agreements; Archiving contracts with government agencies

Organizational Support

- One participant’s institution is identifying the key web archiving ‘talking points’ to present to their university administration. The points need to articulate:
 - The benefit and significance of a web archive to the institution
 - The technical and operational methods for an archiving solution
 - The model for sustainability and collaboration
 - Example talking point: Redundancy of government information archives is important because “administrations change” and “funding streams change”. The GPO “can be a trusted 3rd-party but that’s putting all of your eggs in one basket.”
- Demonstration of value to an organization can ensure funding even with serious budget constraints.
 - Quote: “A huge selling point in getting the agencies to cooperate with us is that we go out and do a little dog-and-pony show. . . . We didn’t create a new place for anyone to look [for archived materials]. . . . We enhanced an existing database. In a really grim series of budget years, that’s protected funding for something that is a real big ticket item and otherwise would have been an easy target, I think.”
- In some cases (e.g., creating dark archives), faculty and academic departments have greater organizational support for infrastructure and funding than university libraries.
 - Quote: “As a faculty member I can archive everything I wanted to. I could get the server space. They’d support me as my faculty right.”
 - Quote: “It could be studied as a project in deep archiving that a doctoral student or graduate student in the Information school could follow.”

Institutional Repository

- One participant’s institution is using D-space to create their institutional repository, which is being populated by the institution’s scholarly research.

Financial Challenges

- State agencies, including state libraries, are often strapped for funds. Within one state government, the department of information technology is responsible for standardizing the state’s web presence however they have no general fund appropriations. This department generates operating revenue by charging fees for their services to other agencies or through grant funding.
- Funding from NSF grants is assisting two archival efforts:
 - Prioritization of Supreme Court briefs to archive
 - Digitization of a subset of Congressional bills

- One state library is not permitted to own servers. The option to contract with a web archiving service provider, who retained ownership of the servers, enabled the agency to create its archive.
 - Quote: "It's not inexpensive, but it's much cheaper than people."
 - Quote: "In our case there's always a vested interest in finding a secure location, a third-party server that is trustworthy but outside the authority of the department."
- One participant said that their university did not envision even having a library ten years from now. This makes it extremely difficult for the library to obtain funding for web archives.
 - Quote: At "this university, the role of the library is very much under pressure. We're trying to prove it in because the campus plan doesn't see a need for a library in ten years."

Roles & Responsibilities

- One participant created an archive of web-published information related to a government environmental cleanup site. The topic is of local interest to users and the organization funded a snap drive for the archive. The archive eventually grew to a size for which the IT group would not provide back-ups, although it was their responsibility to do so.
- The state library is the regional depository library for one of the participants. However, this library itself is "at risk" and experiencing a good deal of turnover. As a result, the academic library serves some of the regional depository library functions.
- One participant reported discord between the library's IT group and the campus IT organization regarding support for server space and content management. Essentially the library needs support from the campus IT organization in these areas.
 - Quote: "Our ITS, which is the campus computer folks, and our support people in the library don't see eye-to-eye. So we haven't gotten a whole lot of support from the campus ITS. . . . The library doesn't want to buy a lot of servers and do that sort of support so, so far, that's sort of been falling through [the cracks]."
- Ideally individual government agencies would be responsible for archiving their own materials, but "this isn't a perfect world".

3.2 Selection

Identification of Source Materials

- University-generated materials
 - Faculty course materials
 - Scholarly research
- Government information: websites, documents, etc.
 - Foreign
 - Issues of Canadian journals
 - Federal agencies
 - Supreme Court briefs
 - Congressional bills (1873 - 1937)
 - Census data
 - Administrative offices of US courts
 - Quote: This material is "hard to figure out, hard to come by, hard to get versioning on. It's 'bad' material. It's kind of available but it's hard to use."
 - State agencies
 - Regional and local government
- Topical information & issues in support of users & faculty research:
 - Homelessness
 - Federal tax reform initiatives
 - Immigration

- Diplomacy: international, federal, & state
- Social welfare movement in American history & social work
- Peace commission reports
- Web-published information on specific topics of local interest that may cut across state agencies, local agencies and outside advocacy groups:
 - Regional environmental issues
 - Environmental Impact Statements (EIS) for a region
 - State gambling issues
 - State electronic voting
- Collections that are both consistent with the mission of the institution and within the institution's archiving capability
 - Quote: "At our place we would have to go real small. I was looking at peace commission reports. I think it would be fairly easy to sell our administration because we have the president's peace commission there. Every semester there's a big workshop that's done on peace issues and it kind of fits in with our mission of being Catholic."
 - Quote: "Could the library offer something that did not tie into our [the university's] stated mission? No. And this is a problem we're having with [a specific archive] because it does not fit our stated mission on campus."
- The sheer amount of data in some areas makes it difficult to determine the institution's archiving 'niche'.
 - Example: There is an enormous amount of numeric data in existence from both government and other sources. The issue is what information the institution should archive.
 - Information that is deemed to be "at risk"?
 - Information that is heavily used within the institution?
 - Quote: "Census stuff is a really good example [of heavily used data]. On the other hand ICPSR [Inter-university Consortium for Political and Social Research], which [we're] a member of, [has] captured all the 2000 decennial census stuff. Do we feel like, 'OK, that's taken care of so we'll put our efforts elsewhere'?"
- Topical, issue-related, or event-related information often spans many government agencies, organizations, and groups. This presents selection and archiving challenges.
 - Quote: "We've been trying to get a handle on how to bring that all together so we have some complete picture of what's been at issue and where it's going. I'm not sure how we can handle it."
 - "If you wanted to get everything that the government produces that has to do with your area" it would be a lot harder than capturing everything produced by a single agency because it would cut across all agencies, "but it would be really valuable".
- Multiple models will be necessary for selection because the model will depend on the materials that are being collected.
 - Subject-centric, cross-agency models
 - Agency-centric, cross-subject models
- In some cases the frequency with which the list of targeted web sites changes may affect selection decisions.
 - Example: One library has a contract to archive a federal agency's publications but does not archive embassy web sites because "it's impossible. They change especially as our diplomatic relations change".
- National security issues may affect selection of items identified for a collection.

Risk Management

- One participant created an archive of web-published information related to a government environmental cleanup site. Some materials originally downloaded from the Department of Energy's Information Bridge were removed from the archive after 9/11 and replaced with notices indicating the materials had been removed from the Bridge from which they

had been downloaded. It was not known who accessed and removed the files from the library's archive, which had not been backed up.

- One participant's institution is exploring redundancy of archived government information as a web archiving 'talking point' to present to their university administration.
 - Example: The GPO "can be a trusted 3rd-party but that's putting all of your eggs in one basket." Redundancy of government information archives is important because "administrations change" and "funding streams change". At the end of the day, users will still need access to government information.

Lost Materials

- One state agency published its annual county-level statistical report on the web in 1998 for the first time. The next year, the agency replaced the 1998 report with the 1999 report.
 - Quote: "That has pretty much become our standard bad example."

3.3 Acquisition

Authenticity of Materials

- The contractual agreement between one library and the federal government agency whose information the library archives stipulates that all available information must be "official". This requires authentication of materials by a federal government agency.
- Users often don't question the authenticity of government information provided by libraries.
 - Quote: "People don't have a great concern about either the source or the format of the information. They want an answer to a question. Whether [the information source] is local, state, or federal is immaterial to them and the only thing that is less material to them is whether or not it's executive, judicial, or legislative."
 - Quote: "I think there's an implicit level of trust in us [libraries], whether or not it's deserved."
- Legal researchers and lawyers care a great deal about the authenticity of the citations in their publications. This is a bit problematic because the law citation manual requires that references cite print resources, many of which are unavailable to researchers in print format. So, the references are generally located and studied electronically and then cited from their print publications.
- Judges and courts care a great deal about the authenticity of materials and documents.
 - Members of the general public acting as their own attorneys want photocopies or printouts of government information certified as "authentic" by librarians.
 - Quote: Members of the general public "want me to authenticate something as a true photocopy or printout of the thing that is in the library. They want me to sign it."
- Some users are quite indiscriminate about the source of their information, being satisfied to find any written source that supports their views.
 - Quote: "They'll take any information. It's almost like an article of faith that if they can find something out there that proves what they already believe, they're going to take it and say 'I don't care where it comes from'."

Source Material Versions & Formats

- Some materials are inherently difficult for users to discover and to understand. Collecting and storing this information in a web archive where it could be made more accessible and be preserved would be a service to users.

3.4 Organization

User Expectations

- Users would like to access cross-agency, comprehensive government information specific to their local area but this is difficult to provide.
 - “We do get a lot of questions about local information. Users look at you like ‘why don’t you have that?’ Well, because it’s just not indexed that way.”

3.5 Presentation

Intellectual Property Issues

- One participant expressed a concern that government information might become increasingly available through licensing agreements that would put limits on its access.
 - Penny Hill Publishing, a commercial vendor, recently put access restrictions in its licensing agreement for the Congressional Research Service publications. The restriction limits access to the purchaser’s organizational intranet.
 - NTIS harvests web-published technical reports from various government organizations and sells them in PDF formats. The purchase carries access restrictions, which in general allow the purchaser to make the documents accessible within their organization. This is an attractive format change from the previously available microfiche format but one participant questioned if the new purchasing agreements might become licensing agreements, with unacceptable access restrictions.
 - Quote: “This generally becomes an issue and I think it will become an issue even more as GPO moves more aggressively to thinking of themselves as a vendor and a supplier of government information. I can see the potential for more license-like restrictions on government information.”
 - Quote: “It’s a legislative mandate. We’ve seen it with NTIS. We’ve seen it with StatUSA. We’ve seen it with cataloging distribution centers. I mean, if we don’t keep paying for the red books, they’re not going to keep making the red books.”
- One Canadian depository library is allowed to make Canadian journals available as long as users are authenticated and only members of this library’s communities are allowed access.

Dark Archives

- While not a preference, one participant indicated they would continue to archive state government information even if they were unable to provide access to it.
- While there was a general sentiment that collecting web materials for a dark archive would have value there were variances in whether a library would be able to do it.
 - Public institutions might be able to maintain a dark archive.
 - Quote: “[The State Library] would argue that we were collecting [information] for the benefit of the United States Government. . . . They find us when they need us . . . it [the archive] turns into a ‘just-in-case’ kind of scenario.”
 - Quote: “There’s a value in [archiving without providing access] but I think it might be a hard value to sell upstream, especially at a private university.”
 - Quote: “In the library we are very limited in what the campus will allow us to do with web archives.”
 - Quote: “. . . the mission of our library is so oriented to access. We have a buzz phrase: ‘the any time, any place library’. That’s helped to guide our collection decisions. A dark archive doesn’t really fit with that, for me.”
- Archives have always accepted collections that have ‘embargo periods’ associated with them. This practice could extend to web archives.

- Quote: “But [the collections] are not things that people would give to us and say keep this for me but you can never show it to anybody. That would be an impossible sell, I think.”
- If the university library, per se, could not create a dark archive within its mission, it might be possible for the university to do so or for a department within the university to do so.
 - A commitment outside of the purview of the university’s main mission might embrace building a dark archive.
 - Example: Stated commitment on the part of the university to the community.
 - A research project under the auspices of a university department could create a dark archive.
 - Example: Doctoral research regarding dark archives within the Information School.

3.6 Preservation

Stewardship

- Ideally individual government agencies would be responsible for archiving their own materials, but “this isn’t a perfect world”

4 Discussion

4.1 Dealing with Change

Building & Preserving Collections

Access to Government Information

A concern was expressed by two participants in this focus group regarding “purchasing” government information in digital formats versus print formats. In general practice, the purchase of government information in print format allowed a library to exercise control over material distribution and access. However, some digital materials available for “purchase” also have “licenses” specifying the extent to which the material can be made available on a licensee’s network. As more government information moves to digital-only publications, the fear is that license-imposed access restrictions to government information will increase, making it more difficult for people to gain access to government information in the future.

Roles & Responsibilities

Security

One participant indicated that sensitivity to “very large national security issues” was needed when creating an archive comprised of government information published by various federal government agencies. Another participant had archived information related to an environmental cleanup operation but some of the information was removed soon after 9/11. It was not known who accessed and removed the files from the archive. This highlights the importance of replication and security for web archives.

Funding

Looking to cut expenses and realign budgets, both universities and state governments are targeting libraries for downsizing and elimination. One university plans to eliminate its library in the next 10 years, apparently anticipating that the research and scholarship needs of its faculty and students can be met without the current library infrastructure and organization. One state library's existence is currently "at risk" and the library is experiencing a good deal of staff turnover. In a climate of uncertainty and funding constraints, the library is unable to fulfill all of the functions it has in the past. How likely is it that many university and state libraries will be able to assume web archiving responsibilities in the future? What are the implications for long-term preservation of state and federal government information?

Technical Infrastructure & Support

Information in support of the university's curriculum and the state government's operations is increasingly web-published and web-accessible. The technical infrastructure and support for preservation of this information in web archives challenges both university and state libraries to look for solutions beyond their own organizations and staffs. Collaborative efforts between libraries and their information technology organizations are a necessity for successful web archiving efforts.

For universities, web archiving predates the involvement of their campus Information Technology (IT) organizations. The scope of the required archival effort is beyond the capabilities of university libraries' IT infrastructures and staffs.

State libraries operate within their state's information resources guidelines and practices, which influence their selection of technology solutions. For example, while the library may not be permitted to purchase or operate its own servers in support of a web archive, the library may be permitted to contract with a third-party provider for this service or to team with their information resources department to create an archive.

4.2 What to Preserve

- Government information: websites, documents, etc.
 - Foreign
 - Issues of Canadian journals
 - Federal agencies
 - Supreme Court briefs
 - Administrative offices of US courts
 - Congressional bills (1873 - 1937)
 - Census data
 - State agencies
 - Regional and local government
- Topical information & issues in support of users & faculty research.
 - Homelessness
 - Federal tax reform initiative
 - Immigration
 - Diplomacy: international, federal, & state
 - Social welfare movement in American history & social work
 - Peace commission reports
- Web-published information on specific topics of local interest that may cut across state agencies, local agencies and outside advocacy groups
 - Regional environmental issues
 - State gambling issues
 - State electronic voting

- Environmental Impact Statements (EIS) for a region
- University-generated materials
 - Faculty course materials
 - Scholarly research

Preserving in the Absence of Repositories

"I have been known to archive web publications by printing them out and having them bound in buckram and then cataloged."

In the absence of digital repositories or web archives at their institutions, some government information librarians create print versions of web-published documents. Others create 'personal' archives of specific web sites on CD-ROMs in support of faculty research.

4.3 Needs & Issues

At the end of the focus group discussion, participants completed the brief questionnaire in Appendix C. The questionnaire elicited information regarding the critical user needs that an archive of web materials would meet in each participant's environment. Additionally, the questionnaire allowed participants to record the critical areas their organization needed to address and the biggest hurdles they faced in building an archive of web-based materials. In general responses echoed and provided a summary of the discussion itself. These results are listed below.

User Needs

In order of importance, the two user needs a web archive would address were:

1. Persistent access to materials for research and reference
 - a. State & federal information from pre-web era
 - b. Information no longer maintained by government agencies
 - c. Web-born government documents published independently from the issuing agencies
 - d. Web-born government publications from first web publication date forward
 - e. Web sites cited in faculty publications
2. Provision of value-added services
 - a. Ease of discovery and access
 - b. Focused collections from diverse sources

Critical Areas to Address

Participants were asked to identify two critical areas their organizations needed to address in order to successfully implement a web archive. The areas are listed below in order of criticality. (*Note: The two areas in item four were of roughly equal importance.*)

1. Resources (funding, staff)
2. Technology (infrastructure, technical expertise, harvesting/archiving tools)
3. Policies related to web materials (what to archive, standards, preservation)
4.
 - (a) Organizational support (bureaucracy)
 - (b) User needs (needs assessment)

Biggest Hurdles

Participants identified staffing issues (priorities, time constraints, and training) as the biggest hurdle their organization faced in creating a web archive. A few participants mentioned management commitment and IT support as organizational hurdles and one participant mentioned funding.

4.4 Need for Collaboration

Finding a Model

The amount of information in many areas precludes exhaustive archiving by any single institution or organization and challenges them to determine their archiving 'niche'. Institutions sometimes struggle to establish criteria to use in making this decision. Should they archive information that is deemed to be "at risk" or information that is heavily used within the institution? What if information of importance to an institution's users is already being archived by another institution or consortia? Will it always be available? Does duplication of effort make sense?

One participant discussed their library's issue with finding a model for collaboration with other libraries. The LOCKSS model is attractive in terms of its 'niche-oriented' or targeted collection focus within a library and its distributed redundancy aspect among several libraries. How might this type of model work with archiving government information?

"Would it be an agency-based model where 'you guys are doing the State Department'? Would we take some agency or some department? Would there be other ways to do it?"

Within one state, a common shortage of funding for web archiving among state agencies has fostered collaboration among the information technology department and the state library to maximize the use of both resources and infrastructure. The information technology department is in a position to receive grant funding to help fulfill its mission in regard to the state's web presence and the state library is able to utilize the IT department's digital information stores to meet their own archiving responsibilities.

Partnerships

The focus of an archive will help define the partners needed for successful collaboration. If the focus is the information published by a federal agency then a university library might partner with the organizations within the federal agency. If the focus is across government agencies at a state, regional, or local level, then a library might partner with the various agencies involved. For example, one participant's library intends to build an archive related to diplomacy, which will involve publications from several federal agencies. In many instances, partnerships might include information providers outside the academic and government arenas.

Examples

1. One participant reported their academic library has had a partnership with a federal agency to archive its website and its sub-agency websites. This is not an exhaustive archive; embassy sites are not included. However, this archive does include web-published agency reports.
2. One academic library is building a web archive comprised of (a) city and county government websites and (b) the websites of non-profit organizations related to homelessness and immigration.

3. One participant suggested that a comprehensive archive of Environmental Impact Statements for a local region might involve the university library's historical archive of EISs and include current information from NOAA and the local Corps of Engineers.
4. A state library is working with 22 executive agencies within the state to archive state government information.
5. One library is discussing the feasibility of breaking up large information content areas, like numeric data, among a collaborative group of libraries in order to determine where the libraries' resources can be best used.

Appendix A. Collection Development for Web Archives

POLICY SETTING	Policy factors influencing web archiving include political mandates, organizational mission, financial parameters, and technical capabilities.	
	SELECTION	
	Selection	Choice of web-published materials for archiving is impacted by the focus of the collection, unit of selection, web boundaries, copyright obligations, and authenticity of materials.
	Acquisition	Web-published materials are acquired or 'harvested' using crawling tools, which either globally or selectively capture web-published materials.
	CURATION	
	Description	Baseline metadata is machine-generated and gathered by a crawler at the time of data capture. Enriched metadata is generally specific to an organization and contains a mixture of human-generated metadata added subsequent to data capture as well as machine-generated metadata.
	Organization	Digital archives of web-published materials typically either retain the organizational structure of the materials as they existed on the web at the time of capture or modify the organizational structure to suit the archive's mission or constraints.
	Presentation	Presentation of web archive materials is related to how the content was captured and to post-harvest descriptive and organizational analysis. For example, archived materials might mirror the web at the time of their capture or might be categorized in accord with selection criteria, such as image files presented by subject.
	Maintenance	Several maintenance functions are critical to ensuring the successful use of materials in web archives: software and hardware training for archive support staff; hardware and software maintenance, performance optimization, backups, and upgrades; and duplicate detection.
	Deselection	Removal of materials from a web archive can be for several reasons: duplication, errors, legal or social considerations (e.g., offensive materials). Risks of removal and retention are weighed against policy and storage costs.
	PRESERVATION	
Preservation	Preservation challenges are numerous. They include persistent naming, format migration and/or emulation, inventory management, volatility, replication, re-validation, curator-operator error, and storage.	

Appendix B. Participants

Kathy Amen
St. Mary's University, Blume Library
Government Information Librarian
San Antonio, Texas

Tim Byrne
University of Colorado
Government Publications Library
Boulder, Colorado

Cass Hartnett
University of Washington Libraries
U.S. Documents Librarian
Seattle, Washington

Julie Linden
Yale University-Seeley G. Mudd Library
Government Information Librarian
New Haven, Connecticut

Scott Matheson
Yale University Law Library
Reference and Government Documents Librarian
New Haven, Connecticut

Colleen Parmer
Bowling Green State University Libraries
Chair, Collections and Technical Services
Head, Government Documents
Bowling Green, Ohio

Aimee Quinn
University of Illinois-Chicago
Asst. Professor, Richard J. Daley Library
Assistant Documents Librarian

Ann Sanders
Michigan State Library
Head, Government Documents
Lansing, Michigan

Bill Sleeman
University of Maryland Law Library
Asst. Director for Technical Services
Coordinator, Collection Development
Baltimore, Maryland

Appendix C. Participant Questionnaire

1. I work in:

- | | | | |
|-------|------------------------------|-------|------------------------------|
| _____ | K-12 School | _____ | Local Government Institution |
| _____ | College or University | _____ | Non-Profit Organization |
| _____ | Federally Funded Institution | _____ | Corporate Institution |
| _____ | State Government Institution | _____ | Specify Other: |
- _____

2. My current position is: _____

3. I have experience creating a web archive: _____ Yes _____ No

4. The two most important user needs that a web archive will address in my library or organization are:

- a. _____

- b. _____

5. Two critical areas my library or organization needs to address in order to successfully implement a web archive are:

- a. _____

- b. _____

6. As I think about the reality of creating a web archive, the biggest hurdle I see for my library or organization is:

7. Your comments are welcomed. Please use back of page if you need more space.

Thanks very much for your help!