Assessment Path Final Report

December 27, 2007

Prepared by:
Kathleen R. Murray
kmurray@library.unt.edu

University of North Texas
Acknowledgements

Special thanks are extended to the many people who participated in and contributed to the assessment activities of Web-at-Risk project.

Project Curators
Arizona State Library; New York University; Stanford University; University of California; University of North Texas

Information Professionals
Archivists and librarians who participated in the focus groups

End Users
Researchers and faculty from academic institutions

Content Providers
Representatives for government agencies and trade unions

Project Staff
California Digital Library; New York University; University of North Texas

Funding Source
Contents

1 Introduction ........................................................................................................ 4

2 Needs Assessment ............................................................................................. 4

  2.1 Participants .................................................................................................. 5

  2.2 The Current Climate .................................................................................... 6

  2.3 Organizational Issues .................................................................................. 6

  2.4 Collection Development Concerns .............................................................. 7

  2.5 Discussion .................................................................................................. 9

3 Collection Development .................................................................................... 9

  3.1 Success Factors for Implementing Web Archives ........................................... 10

  3.2 Collection Plan Content ............................................................................. 13

  3.3 Curators’ Collection Plans ........................................................................ 15

  3.4 Discussion .................................................................................................. 16

4 WAS Evaluation .................................................................................................. 17

  4.1 WAS Releases ............................................................................................. 17

  4.2 Usability Testing ......................................................................................... 19

  4.3 WAS Enhancements .................................................................................. 20

  4.4 Discussion .................................................................................................. 22

5 Summary ........................................................................................................... 23

Appendix A: Web-at-Risk Project Curators ......................................................... 26

Appendix B: Reports and Curator Tools ............................................................... 27

Appendix C: Publications & Presentations ......................................................... 29
1 Introduction

The Web-at-Risk project is a three-year collaborative effort of the California Digital Library, the University of North Texas, and New York University funded in 2004 by the Library of Congress under the National Digital Information Infrastructure and Preservation Program. The project developed a Web Archiving Service (WAS) to enable curators to build, store, and manage collections of web-published materials.

The Web Archiving Service employed a user-centered design approach involving a group of curators. The curators included librarians and archivists from the Arizona State Library, Stanford University, New York University, the University of California campuses, and the University of North Texas. With the exception of one curator who works in a state library, the project’s curators work in large academic libraries. Many of the curators work in government information departments while others are subject specialists in the areas of public policy, trade unions, and political movements. All of the curators have collection management responsibilities and select print materials, electronic resources, and web-published materials for their collections. Over the three-year course of the project, the 36 curators listed in appendix A were involved in one or more assessment activities. While a core group of curators were involved throughout the project, several curators moved on to other responsibilities and were replaced by others. At any one time there were generally 16 – 22 curators actively involved in each assessment activity.

Curators interviewed content providers, identified functional requirements for the WAS, trialed and evaluated each of three software releases of the WAS, participated in usability testing of WAS Release 4, and prioritized desired enhancements for future releases of the WAS. Additionally, curators created web collection plans for the collections they built using the WAS. The content of the collections was captured largely from US federal and state government agency web sites, but also from web-published political policy documents, campaign literature, and information related to political movements and labor unions.

This report describes the major assessment activities, their contributions to the project, and key findings. The assessment work areas included:

- Needs Assessment
- Collection Planning
- WAS Evaluation

Detailed reports for all assessment activities were published on the project wiki¹. Appendix B lists and provides URLs for these reports. Appendix C lists the publications and presentations related to the assessment activities.

2 Needs Assessment

In 2005, 22 project curators, as well as 43 librarians and archivists who primarily work in academic libraries, seven university researchers, and seven content providers participated in needs assessment activities. Three types of assessment activities were conducted: (a) an online survey of curators involved in the Web-at-Risk project, (b) focus groups with librarians working in a variety of settings, and (c) interviews with both potential end users of web archives and providers of web content. The purpose of these activities was to elicit the needs and issues librarians, curators, end users, and content providers have in relation

¹ Web-at-Risk Project wiki: http://wiki.cdlib.org/WebAtRisk/tiki-index.php
to web archives\(^2\). The participants and key findings of these assessment activities are briefly described in this summary.\(^3\)

2.1 Participants

**Survey**
Survey respondents were the 22 curatorial partners involved in the Web-at-Risk project at the time the survey was conducted. In all, 16 surveys were submitted. Ten curators submitted individual surveys while 12 curators submitted a total of six surveys, each of which represented a joint effort between two curators. Four of the surveys were submitted by curators with collection responsibilities in the areas of public policy or political movements. The remaining 12 were submitted by curators with collection responsibilities in local, state, federal, or international government information.

**Focus Groups**
Five focus groups were held in the summer and fall of 2005. Two focus groups were held during national conferences for two organizations, the American Library Association conference in Chicago in June 2005 and the Federal Depository Library Conference (FDLC) in Washington, DC in October 2005. The three remaining focus groups were held at each of the three project partner institutions (New York University, California Digital Library, and the University of North Texas). A total of 43 people participated in the five groups. The majority (\(n = 39\)) worked in colleges or universities and 33% (\(n = 14\)) held library management positions (e.g., Department Heads). About 25% (\(n = 11\)) of the participants indicated they had some prior experience creating web archives.

**Interviews: End Users**
Seven end user interviews were conducted: four with historians, two with political scientists, and one with a professor of hospitality law and management. With the exception of one person whose use of web-published resources was limited to archival finding aids for research, all of the participants used web-published materials in both their research and professional activities, although the extent of their usage varied widely. For some, web-published materials were more likely to be used in their teaching and for others, in their research or professional activities.

**Interviews: Content Providers**
Seven content provider interviews were conducted: three with representatives of union organizations and four with representatives of state government agencies or state government sponsored programs. The unions had existing relationships with a university archive for the preservation of their print materials and, in two cases, these relationships extended over many years. One state government agency had an existing relationship with an archive for the long-term preservation of its major web-publication. Most representatives of state government agencies were sensitive to the issues involved in archiving web-published government information and many were aware that their web sites were already being crawled and captured.

\(^2\) For this needs assessment, a web archive was defined as an archive that contains web-published materials for which an organization has accepted long-term responsibility for both preservation and access.

2.2 The Current Climate

Librarians are facing many challenges as they continue to work in the familiar world of print materials while increasingly accepting responsibilities in the ever-growing world of web-published materials. While interested in embracing the challenges inherent in web-published materials, librarians often lack the technical expertise, the resources, or both. Most acknowledge that collection development models for print materials transfer only at great expense to web-published materials, which are expensive to select, capture, and catalog. In a climate of uncertainty and funding constraints, university libraries find the scope of the preservation effort beyond the capabilities of their IT infrastructures and staffs.

Librarians generally agree that the organization or individual responsible for producing web-published materials ought to take responsibility for preserving them. In practice, however, librarians perceive these content producers as either unaware of the need to preserve their web-published materials or unable or unwilling to accept the challenge. Libraries have traditionally accepted preservation responsibility for print publications, but they lack the resources to extend this practice to web-published materials. On the other hand, most content providers interviewed share a view of a web archive as a safe repository for specific web-published materials of historical value that are beyond the purview of providers’ own retention mandates and possibly beyond their technical or financial ability to preserve.

In regard to federal government information and the continuing shift from print documents to web-published materials, some major research libraries are not certain they can either wait for or rely solely upon federal government preservation efforts. Librarians express concerns regarding the sustainability of government programs in future funding cycles. This uncertainty drives these libraries to assess their need for local preservation programs and to begin planning to meet those needs.

Responsibility for preserving state government publications is often unclear or non-existent and many publications are simply disappearing. State libraries are in a logical position to preserve state government publications but are often understaffed and resource-constrained, resulting in hit-and-miss efforts in regard to preserving the web-published materials of state agencies. The concern for preservation and access to the web-published materials of federal and state agencies extends to those of local government entities, whose need for assistance in preservation of their web-published materials is quite high.

2.3 Organizational Issues

By means of a questionnaire completed immediately after each of five focus group discussions with librarians and archivists, participants identified the major hurdles they envision for their library or organization in creating archives for web-published materials. The four major hurdles were echoed in the focus group discussions and by survey participants. These hurdles were:

1. Technology
   • IT support
   • Preservation expertise
2. Policies
   • Lack of organizational focus for preservation of web-published materials
   • Agreement regarding which materials to archive
   • Agreement regarding what archive technology to implement
3. Management commitment
   • Senior management support for an archive effort
• Political will to drive an archiving effort through the organization

4. Funding
• Limited money and budget constraints
• Staffing issues – a shortage of people and time

The project’s curators estimated the magnitude of the financial challenges involved in creating and maintaining their collections of web-published materials for the Web-at-Risk project. In rank order, the top four financial challenges they anticipated were: cataloging, preservation, IT support, and staff training.

2.4 Collection Development Concerns

While collection development activities for web-published materials conceptually parallel activities for print materials, most librarians find they are more labor-intensive. In particular the activities of selection and acquisition require more up-front work and often involve individual review of materials. These activities are especially challenging in collection development for less-established disciplines for which web-published materials often represent the bulk of available information. A discussion of the key collection development concerns that emerged in the needs assessment follows.

2.4.1 Selection

Identifying what to preserve was a major issue for most participants. The two basic questions librarians ask in regard to identifying web-published materials for preservation are: “Should we save this?” and “Is someone else already saving it?” Overall, the important materials targeted for preservation by librarians fell into four categories:

1. Government Information
   • National
   • State
   • Regional & Local
2. Information in Support of Academic Institutions
   • Teaching & the Curriculum
   • Scholarship
   • University Operations
3. Information Pertaining to Key Events
4. Information Pertaining to or Produced by Organizations

Librarians identified the following materials as currently falling through the cracks of preservation programs: smaller journals, state and local government publications, and institutional web-published materials. The sense was that these types of publishers did not have the historical models or the financial resources to commit to preservation.

2.4.2 Unit of Selection: Content v. Context

For certain research disciplines or types of research, source material context is critically important and therefore the web site would be the unit of selection. In terms of contextual importance, one historian made the analogy between a web site and a newspaper observing that placement of material on a web site has meaning much in the same way placement of an article in a newspaper has meaning. For other research fields, such as statistical research in sociology, the original web-context of the source materials is not always critical and users would be better served by interacting directly with the statistical datasets.
2.4.3 Acquisition
All participants were generally concerned with the frequency with which web-published materials change. Survey respondents identified three important considerations for collection building practices:

1. Assessing the change rate of the source materials
2. Establishing the interval at which collection materials will be captured
3. Articulating criteria for retention of earlier versions

2.4.4 Authenticity
Different users assess authenticity differently. Some need to trust and have assurance that the materials in the archive are reliable and have not been altered. Others, such as persons involved in legal issues related to documents that will be accepted in a court of law, need assurance that the original document has been authenticated by the publishing agency through some type of software means. Survey respondents were concerned that multiple versions of source materials captured at different points in time and multiple formats of the same object might pose a threat to the authenticity of those materials. Amplifying this concern, focus group participants indicated that establishing “fixed” versions and dates for web-published materials (i.e., dates that were never altered once established) is a critical area an archive should address. Many researchers would like an archive to identify the location of original source materials.

2.4.5 Metadata
Survey respondents identified cataloging as the top financial and technical challenge they see in regard to building web archives. Librarians in general anticipate that in creating collections of web-published materials, the biggest challenge will be the application of metadata. Focus group participants reported their libraries currently do not have enough catalogers for their non-web-published materials. Librarians recognize that evaluating web-published materials and applying metadata requires a specialized skill set. New approaches that utilize technology and include users might apply “indicators of usefulness” to materials and provide new mechanisms for users to evaluate archived materials.

2.4.6 Organization
Librarians anticipated users would expect full-text search capability in a web archive. Librarians also thought users would want to search by subject category and thought it would be important to “provide some higher-level topical access, even if it is derived from the title as opposed to the actual content.” Researchers indicated the most important types of searches are “topic or subject” and “full-text using any keyword” and they also indicated they would like to browse a web archive via a subject directory structure.

2.4.7 Presentation: Look-and-Feel
For some content providers, their databases and datasets are the meat of their content, and to varying extents all other content on their web sites is superfluous. These content providers do not think that replication of their web sites’ “look-and-feel” is important when archived materials are presented. In a like observation, librarians agreed that preserving the content of journal articles would suffice. However, many participants thought other types of materials would need to be presented in their original web context. This was of particular importance for historical research in many disciplines. For some librarians and researchers, web sites in an archive were basically viewed as historical records and, as such, the librarians and researchers thought that the archived web sites should be presented in such a way that they mirror the source web sites.
2.4.8 Presentation: Authenticity Indication

Researchers assert that web archives should make it clear that users are interacting with archived material and not “live” material. For certain types of research purposes, a web archive must also be able to provide and present some assurance that what users are seeing is a reliable copy of the originally published material. In legal research, it is critical that archived materials have a designation of authenticity from the original publisher. For maps and GIS data regarding environmental or natural resources and agricultural reports, both an indication of reliability as well as version date is critical. Content providers were also concerned about how an archive might represent itself; they would like users of an archive to be made aware that they are dealing with copies of materials from the original publishers and not with the publishers’ “live” materials.

2.5 Discussion

Librarians who participated in the focus group discussions were asked to identify the top three user needs web archives could address at their institutions or organizations. The most important need they identified was persistent access to the information users need for teaching and research. The participants also identified two additional needs an archive could address: provision of value-added information services, such as aggregation of content from disparate sources, and persistent access to the institution’s history and intellectual products in an institutional repository. Articulating the benefits of a web archive or institutional repository and identifying the risks of not preserving web-published materials of importance to researchers and other users should help libraries as they build business cases they can present to administrators and funding agencies.

Often when looking to cut expenses and realign budgets, both universities and state governments target libraries for downsizing and elimination. At the same time, libraries and archives are responding to an urgent and growing need to collect and preserve web-published materials. This effort cannot be addressed without partnering both internally with other departments in their organizations and externally with other organizations and government agencies. In addition to partnerships, software tools are needed to address several aspects of web archiving. In particular librarians and archivists need tools to help with metadata application, evaluation for selection and capture, and version comparison. Additional tools are needed for preservation of web-published materials, including tools for file format validation and integrity assurance.

Lastly, registry services for web archives would provide an answer to librarians’ need to know if some other organization is already preserving specific materials and collections of web-published materials. Clearly it would be of value to create a shared registry service for web archives. The benefits of such a registry service for libraries include expanding access to materials, eliminating redundancy of effort, and controlling preservation costs.

A registry service in combination with collaborations, partnerships, and a web archiving service would provide a suite of solutions to the major hurdles librarians and archivists envisioned as they considered the challenges of building and preserving collections of web-published materials. The WAS is a service solution that begins to address significant needs of information professionals that emerged in this needs assessment.

3 Collection Development

Findings from the needs assessment, as well as a review of web archiving policies and practices, informed the creation of a set of collection planning guidelines for web-published materials. Guidelines were developed for the general task of creating collection policies and
plans for web-published materials in libraries as well as for the specific use of the project’s curators who created collection plans for the collections they intended to build using the Web Archiving Service. A companion template for creating their collection plans was also developed for the project’s curators. Appendix B lists and provides URLs for the guidelines and template.\(^4\)

Librarians and archivists who participated in the focus groups in 2005 (\(N=43\)) identified the seven factors in Figure 1 as critical to the successful implementation of web archives in their organizations. Whether a library plans to create its own web archive or utilize an external archive service, each of these factors should be explored prior to creating collection plans for web-published materials. Doing so should help libraries identify critical areas where policies or practices need to be established. Section 3.1 briefly discusses each factor. Section 3.2 presents a content model for a web collection plan. Section 3.3 lists the collection plans created by the project curators.

![Critical Success Factors for Web Archive Implementation](figure1.png)

**Figure 1. Critical Success Factors for Web Archive Implementation**

### 3.1 Success Factors for Implementing Web Archives

#### 3.1.1 Consortial Approach

Libraries have a rich tradition of collaborations and consortial efforts that provide models for sharing the preservation responsibilities for web-published materials. While “preserving the Web” is an endeavor beyond the mission and resources of individual institutions and their libraries, identifying web-published materials that support an institution’s mission and establishing procedures to preserve them is of importance to most institutions. Often these goals can be advanced with consortial approaches.

---

3.1.2 User Evaluation

3.1.2.1 Identifying User Groups
A fundamental task in the formulation of a policy for the preservation of web materials is to define the user groups for whom information is being collected and preserved. The Reference Model for an Open Archival Information System\(^5\) (OAIS) refers to these groups as a Designated Community. Because a Designated Community may consist of disparate user groups, it is important to identify each of the user groups and evaluate their specific needs in regard to archived web-published materials.

3.1.2.2 Involving User Groups
Members of the Designated Community should initially be consulted to identify their information needs. Subsequently the community should be engaged in evaluating both collections of web-published materials and a web archive’s effectiveness in regard to meeting the community’s needs.

Within an academic institution, user groups will likely include researchers, faculty, students, members of the public, administrative staff, and alumni. Each of these user groups would be part of an institution’s Designated Community but their unique information needs might predicate different requirements in regard to:

- depth of the collection
- frequency with which materials should be captured
- level of descriptive metadata
- discovery mechanisms required in the user interface
- manner in which materials should be presented
- expectations as to preservation of the collected materials
  - certification of authenticity
  - format migration
  - retention of versions

3.1.3 Copyright

3.1.3.1 Legal Considerations
Since a web archive must honor all legal restrictions regarding copyright and intellectual property, it should be a matter of policy that web sites be evaluated in this regard. Ideally this evaluation should occur in the planning stages for a collection and prior to acquisition of the web sites. The Digital Millennium Copyright Act (DMCA) allows digital reformatting and migration over time of print source materials but does not specifically extend this allowance to born-digital source materials. The DMCA requires permissions from creators before making any copy of digital materials and encourages actions, such as encryption, password protection, and other security mechanisms, to prevent copy violations of born-digital materials from occurring in the first place. Additionally, the DMCA makes it illegal to create tools to thwart such preventative actions. These actions present challenges to the typical collection or capture method for web-published materials, specifically to web crawlers. Furthermore, the characteristics of the DMCA also present challenges to digital preservation methods such as migration and creation of redundant copies of born-digital materials.

---

It is critical to define as specifically as possible what rights the archiving agency has over the materials in its web archive. At a minimum, the rights holder’s responsibilities in the preservation and dissemination of the data must be defined. Preferably, the archiving agency would be allowed to act on behalf of rights holders to execute changes to the content for clearly defined preservation activities. These types of changes might include:

- Reformatting of materials for continued access when necessary hardware and software become obsolete
- Changes to preservation metadata to record preservation activity

3.1.3.2 Other Considerations

Collection Perspectives

Two general approaches to capturing web-published materials can be observed in web archiving efforts: “opt-in” and “opt-out.” An opt-in policy is one in which explicit permission from content owners is sought before web-published materials are captured. An opt-out policy requires that content owners explicitly request that their materials not be captured either by robots.txt exclusions or official take-down requests.

Embedded Information Objects

As Lyman\(^6\) points out, it is important to realize that a web page may consist of multiple items (e.g., sounds, images, etc.) and that each of these items may also be protected by intellectual property rights.

Privacy

Lyman also identifies potential privacy issues to consider. Some web sites collect data about their customers in order to provide a customized environment. While exposure of this data to a web crawler is not likely, any collection of this data may be regarded as an invasion of privacy.

3.1.4 Policies

Most physical or print materials are in forms that endure for some predictable period of time allowing for preservation actions and policies to be outlined within that interval. The urgency to address the preservation of web-published materials is that their longevity is unpredictable and materials are often lost in relatively short time frames. Therefore, preservation of web-published materials must be addressed in organizational policy in order to ensure the materials will not be lost due to a delay of action.

3.1.5 Organizational Support

Creation and preservation of collections in web archives requires enormous effort and resources, spanning several departments within an organization. Organizational support in terms of a long-term commitment to funding is required for any web archiving effort to succeed. It is important to articulate how collection and preservation of web-published materials supports the organization’s mission, benefits the organization, and provides a valuable service to the community it serves.

Information management professionals, whether librarians, curators, or archivists, have expertise in collecting and preserving materials, but often do not have the technical expertise necessary to create and preserve an extensive collection of web-published materials. Information technology professionals do have expertise working with networks.

and digital storage, but rarely understand the long-term implications inherent in curation and preservation of stored content. It is clear that these organizational units will need to work together to achieve success in collecting and preserving web-published materials.

3.1.6 Resources

3.1.6.1 The Resource Challenge
In many libraries, resources such as money, people, and infrastructure, are stressed by ever-growing expectations from both management and end users. Generally, identifying web-published materials and making them accessible has been incorporated into library selectors’ responsibilities. However, in many cases, these are labor-intensive responsibilities that have not been addressed with increases in funding or staff. Likewise, the library’s IT infrastructure and internal support staff is typically unable to provide archival support for these increasingly important areas within library collections.

3.1.6.2 Importance of Web-Published Materials
Prior to gaining the resources necessary to build and preserve collections of web-published materials, a library will generally need to document how collecting these materials promotes and supports the institution’s mission. This process can generate internal selling points for creating these collections and identify the risks to the organization of not preserving web-published materials. It can also articulate the importance of web archives to a library’s end users.

3.1.7 Technology
Technology challenges at all stages of collection development for web-published materials are directly related to the challenges of web archiving in general. Characteristics of source materials may impact the degree to which web-published materials are successfully captured. Source materials present multiple challenges to web capture tools (i.e., web crawlers) including implementation-specific challenges such as use of Macromedia Flash, PHP, Java, and JavaScript. Some capture tools handle these challenges better than others. In addition, source materials may present challenges that no existing tools can overcome, such as password-protected source materials and web pages generated in response to users’ database queries.

Once materials are collected, it can be technically challenging to apply adequate metadata to them. Materials that are re-collected on a regular basis require that versions be separately identifiable. This presents challenges in regard to both evaluation of the differences among versions and identification of different versions in presentation.

Storage, presentation, and preservation of materials require an extensive technical infrastructure and expertise in storage media, data replication, networking, and risk management. Additionally, technical understanding of the requirements involved in content preservation is critical. These requirements include maintaining the renderability and understandability of content bitstreams as well as preserving the integrity of content over time.

3.2 Collection Plan Content
Web collection plans provide guidance for managing collections of web-published materials created for specific groups of users within an institution. This is not unlike the general role collection plans or collection policies serve for traditional collections within a library. Figure 2 identifies the major phases and activities involved in web collection development.
Figure 2. Collection Development Phases

A collection plan should provide sufficient guidance for conducting activities in each phase. The plan can be organized in sections that roughly correspond to each of the major activities involved in collection development. Table 1 lists eight sections to include in a web collection plan and identifies the topics to be addressed in each one.

Table 1. Collection Plan Content

<table>
<thead>
<tr>
<th>Section 1. Mission &amp; Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Mission Statement</td>
</tr>
<tr>
<td>B. User Group(s)</td>
</tr>
<tr>
<td>C. Collection Subject, Theme, or Event</td>
</tr>
<tr>
<td>D. Curator(s)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 2. Selection Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Seed List</td>
</tr>
<tr>
<td>B. Capture Scope</td>
</tr>
<tr>
<td>C. Rights Metadata</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 3. Web Site Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Frequency of Capture</td>
</tr>
<tr>
<td>B. Capture Scope</td>
</tr>
<tr>
<td>C. Material Types &amp; Formats</td>
</tr>
<tr>
<td>D. Interactive &amp; Dynamic Content</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 4. Descriptive Metadata Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Level of description</td>
</tr>
<tr>
<td>B. Metadata elements</td>
</tr>
<tr>
<td>C. Controlled vocabularies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 5. Presentation &amp; Access Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Discovery</td>
</tr>
<tr>
<td>B. Access</td>
</tr>
<tr>
<td>C. Look-and-Feel</td>
</tr>
<tr>
<td>D. Dynamic Content</td>
</tr>
<tr>
<td>E. Multiple Types/Formats</td>
</tr>
<tr>
<td>F. Authenticity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 6. Maintenance &amp; Weeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Maintenance Activities</td>
</tr>
<tr>
<td>B. Deselection Guidelines</td>
</tr>
<tr>
<td>C. Collection Evaluation</td>
</tr>
</tbody>
</table>
3.3 Curators' Collection Plans

In creating their collection plans, the project curators used a template that included the sections in Table 1. Additionally, the curators were given a set of guidelines that explained each section and provided guidance for completion. Curators used their collection plans in the trials of WAS Releases 2/3 and 4 to capture seed URLs and begin building their collections. Following are the collection plans created by the curators.7

The Tamiment Library at New York University
Michael Nash and Peter Filardo; New York University

Islamic and Middle Eastern Political Web
John A. Elts; Stanford University

International Government Organizations and Developing Countries
James A. Church; University of California, Berkeley

AFL-CIO/Change to Win Web Archive
Terence K. Huwe; University of California, Berkeley

California Political Blogs and Interest Group Websites
Nick Robinson; University of California, Berkeley

Local Government and Local Area Flood Control Collection
Marcia Meister and Juri Stratford; University of California, Davis

UCI Orange County Government Information Web Collection
Yvonne Wilson; University of California, Irvine

UCLA Online Campaign Literature Archive
Gabriella Gray and Scott Martin; University of California, Los Angeles

UCLA NGO & Local Government Information
Kris Kasianovitz; University of California, Los Angeles

UCR Inland Empire (CA) Web Archive
Lynne Reasoner and Kenneth Furuta; University of California, Riverside

UCSB Santa Barbara, Ventura, and San Luis Obispo Counties Local Planning Documents and Water Archive
Sherry DeDecker and Janet Martorana; University of California, Santa Barbara

3.4 Discussion

From an organizational and policy perspective, collection plans within a library articulate the role a collection has within the organization and state the organization’s commitment to building such collections. Often collection plans identify policies, guidelines, and standards that affect collections. These might include technical standards regarding the format of web-published materials suitable for collections, web archive policies regarding metadata requirements, or guidelines regarding required copyright clearances.

Most collection plans also address the preservation of materials in the collection. Web-published materials have unique preservation requirements that may require new policies or modifications to existing policies. To successfully develop collections of web-published materials, it is important for an institution to develop policies and guidelines that support preservation and other requirements unique to collection development for these materials. Such policies and guidelines require management endorsement as well as committed support from all units and people involved in selection and ongoing maintenance of the institution’s collections.

In addition to documenting and implementing relevant organizational policies, collection plans also document the activities necessary to create and manage a collection of materials for a specific user group or entity within an organization or a library. These activities include describing the user group(s), describing the collections and specifying their contents, and creating metadata/catalog records.

The collection planning guidelines and associated template created for the project curators assisted them in developing plans for the collections they created using the project’s Web Archiving Service. Some of the plans describe web collections consisting of a group of related web-sites, and others describe groups of related discrete web-publications, such as a set of public policy documents related to a common subject. The curators’ collection plans identified URLs to be captured using the WAS. Curators subsequently added captured sites, both entire web sites and discrete items, to the collections they built using the WAS.

The collection plans serve as management tools that curators can use for ongoing maintenance and development of their web collections. They also provide a handbook for continued collection development should a new curator assume responsibility for a collection.

The collection planning guidelines may also be useful to librarians, archivists, and curators who are not involved in the Web-at-Risk project. The Chesapeake Project, “a two-year pilot digital preservation program established to preserve and ensure permanent access to vital legal information currently available in digital formats on the World Wide Web” created their
collection plan using the guidelines and template developed for the Web-at-Risk project curators.\textsuperscript{8}

4 WAS Evaluation

4.1 WAS Releases

Curators trialed three releases of the Web Archiving Service (Table 2). Subsequent to their trials, curators submitted web-based survey evaluations of their experience. The surveys solicited curators’ feedback in regard to: Site Definition, Capture, and Management; Reports and Analysis of Capture Results; Search and Display of Capture Results; Collection Management; and General Features (Login, Help, etc.). Appendix B lists the URLs for the results of each evaluation of the WAS.

Table 2. Trial and Evaluation Dates for WAS Releases

<table>
<thead>
<tr>
<th>WAS</th>
<th>Trial Dates</th>
<th># Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Release 1</td>
<td>September 11, 2006 - September 26, 2006</td>
<td>18</td>
</tr>
<tr>
<td>Release 2/3</td>
<td>January 30, 2007 – February 13, 2007</td>
<td>17</td>
</tr>
<tr>
<td>Release 4</td>
<td>September 18, 2007 – October 12, 2007</td>
<td>17</td>
</tr>
</tbody>
</table>

4.1.1 Release 1

Overall, curators were pleased with the service and found it easy to use. Curators found the help screens and user guide useful and the interface clean. They were generally optimistic about future releases. From a usability perspective, the three functions that created the most confusion were:

1. Starting a capture after defining it
2. Determining when a capture was completed
3. Determining how to view captured results

To address these areas, curators provided the following recommendations:

- Make the RUN CAPTURE icon more obvious; add a RUN button on the capture definition screen to initiate a newly defined capture.
- Provide service-initiated “positive feedback”: after captures complete, send an email message to signal capture completion and to notify curators with a report that includes what was captured

Many problems with capture results were associated with incomplete captures, likely due to the 10 minute capture constraint for this trial. Some other problems seemed related to the Firefox browser or to the Wayback Machine. When evaluating their capture results, a few curators reported problems that might be of concern:

1. File listed in crawl log but not found in the archive
2. Damaged file error message for file that appeared to be intact on the site
3. Displayed images turning gray

\textsuperscript{8} The Chesapeake Project: Legal Information Archive adapted the guidelines and template for use in their project: \url{http://web3.unt.edu/webatrisk/cp_cur/CP/ChesapeakeCollectionPlan_July2Draft.pdf}
4.1.2 Release 2/3

Overall, curators thought the workflow reflected in the redesigned interface was a great improvement from the first release and they noted that the service was intuitive and easy to use. This release effectively addressed the three key functions that posed the most confusion in WAS release 1. These were: (1) Starting a capture after defining it; (2) Determining when a capture was completed; and (3) Determining how to view captured results.

While opinions about the new and enhanced features in this release were not unanimous, each major enhancement was welcomed by some. The features many curators received well were: searching and browsing; reports; filtering and sorting of files; and longer capture times. Several curators noted that the overall performance of this release was very slow and some were frustrated that many or all of their captures were not successful. Additionally, many were unable to display files and some found the browsing interface a bit of a barrier to accessing and viewing a captured site in its entirety. The major functions that were either confusing or problematic in this release were:

1. Estimating max time capture settings for sites
2. Inability to display files and receiving either an error message or a “Not in Archive” message for files listed in the captured files list
3. Unexpected search results and lack of understanding about what content was being searched
4. Discovery and display of discrete files of publications or documents, especially PDF files
5. Discovery and display of the “home page” for seed URLs (versus individual files)

4.1.3 Release 4

Overall, curators were very satisfied with this WAS release, as one curator commented: “I can finally envision how I would use it [the WAS] to build collections.” Another curator wrote: “I’m very optimistic I will be able to make this work for capturing and organizing my collections.”

As with WAS release 2/3, opinions about the benefit and usefulness of new and enhanced features in this release were not unanimous. However, several curators welcomed the improvements in overall workflow, available help, and the speed of captures. The new features curators mentioned as some of the best in this release were the bookmarklet, which allowed curators to define sites as they browsed the Web, the RSS feed, which provided the capture completion status of a curator’s most recent captures, and the ability to compare two different captures of the same site, which one curator thought might be “a backbone feature for WAS.”

This release appeared to effectively resolve three problems in WAS release 2/3: (1) estimating max time capture settings for sites, (2) unexpected search results and lack of understanding about what content was being searched, and (3) discovery and display of discrete files of publications or documents, especially PDF files. A fourth problem curators reported in WAS release 2/3 was partially addressed; while no curators reported receiving either an error message or a “Not in Archive” message for files listed in the captured files list, one curator reported they continued to be unable to display all files from captured sites. A fifth problem, display of the “home page” for seed URLs, remains a problem with 71% of curators (n=12) reporting they had difficulty displaying the content of seed URLs.

The major areas that were either confusing or problematic in this release were:
1. Receiving an error message when attempting to search collections (Message: 'There was an error getting the index for searching or listing files. This index may be very large and is taking a long time with its initial loading.')

2. Display of the “home page” for seed URLs

3. Installation of the bookmarklet and RSS feed at campuses whose IT policies prohibit end users from installing programs/tools such as these

4.2 Usability Testing

Usability testing of WAS Release 4 was conducted in May 2007 with eight of the project’s curators participating in test sessions. Usability relative to four functional areas of the Web Archiving Service was tested:

- Create/Capture a Site
- Create a Collection
- Manage Collections and Capture Results
- View and Add Metadata

Appendix B lists the URL for the results of the testing. Modifications were made to the WAS based on the results prior to the actual trial of WAS Release 4 in September 2007. Listed below are the general findings and recommendations.

4.2.1 General Findings

- The WAS has come a long way in the last year and is going in the right direction.
  - The WAS interface design is clean and is easy for curators to explore.
  - The WAS functionality relates well to the activities involved in web collection development.
- Terminology in the interface is confusing to curators who are often not clear on the distinctions between the following terms used in various places on the WAS screens and consequently are not sure what action to take.
  - Site, Collection, Job, Capture, File, and Object
  - Manage and View
  - Create, Build, and Add
- The interface favors people who tend to read the explanatory text on screens and contained within sidebars prior to taking actions. Some of the curators did this almost consistently and some almost never did this. The latter group met more often with frustration, got lost navigationally, and made incorrect interpretations of actions and concepts.
- Likewise, some curators readily discover the icons on screens and explore both the mouse-over text descriptions and/or the Icon Keys in the sidebar. However, other curators find the action icons rather small and they are not drawn to either discover their meaning or to select them.
- Many curators, from time-to-time, get lost in the application and are unsure either of how they navigated to a location or of how to navigate to the feature or functionality they need. This was most common when curators attempted to add content to a new collection.

4.2.2 General Recommendations

- Terminology
  - Identify a core set of terms for key actions and concepts within the WAS and use them consistently.
  - Add an indexed and searchable glossary that includes definitions of common terms, concepts, and icons.
Navigation
- Create an obvious and optimal method curators will use for each major activity in the web collection development workflow.
- For each major activity ensure that curators can make a single corresponding selection from the main drop-down navigation bar.
- Simplify the number of avenues available for task completion. Merge related features/content, for example, perhaps the Capture History Tab and View Results screen can be merged and made accessible via Sites → Capture Results.
- Consider adding breadcrumbs to each screen to indicate where a user has navigated to within the application.
- Ensure that the browser’s back button works appropriately.

Action Icons
- Decrease the number of action and navigation options on screens. For each screen, determine the most important tasks/actions a user would undertake from that screen and optimize placement of the associated icons to attract users to them.
- Add text labels to icons.
- Ensure that each icon’s hover text and Icon Key text in the sidebar are the same, for example, “view capture history” vs. “view job history”.
- Ensure that each icon’s action or meaning is consistent within the application, for example:
  - Magnifying glass: "view capture history" vs. "view job history"
  - Trash can: “deactivate” a site vs. “delete” a collection, site, or file

Documentation
- Develop an indexed or searchable library of simple “how-to” guides that can be accessed from any screen by selecting ‘Help’.
- Consider developing step-by-step wizards as tutorials for common tasks, such as those in the usability testing.

4.3 WAS Enhancements
Several enhancement ideas for future releases of the Web Archiving Service were identified from the feedback provided by curators in their evaluations of the three WAS Releases. In order to refine and prioritize these ideas, project curators participated in an online exercise using a Delphi Technique. In all, 31 enhancement ideas resulted. Curators were asked to consider the following question as they refined and rated the ideas: What feature enhancements to the Web Archiving Service could be made to better address your needs for collection and preservation of web-published materials? Table 3 lists the top 13 ideas in priority order.

Table 3. Enhancement Ideas in Rank Order

| Rank | Idea |  
|------|------|---
| 1    | For multiple captures of the same site, indicate in capture results if the site changed since its last capture. If the site changed, allow easy identification of specific files that changed. |

9 The Delphi Technique consisted of a series of three online questionnaires, each available for a five or six day period. Of the 16 curators who participated in this exercise, seven submitted all three questionnaires, three submitted two questionnaires, and six submitted one questionnaire.
<table>
<thead>
<tr>
<th>Rank*</th>
<th>Idea</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Curator access to the entire archive so that an individual curator could readily determine if another curator has already defined a site, what parameters the curator specified, precisely when the site was captured, and if captures were successful.</td>
</tr>
</tbody>
</table>
| 3     | Schedule captures based on one or a combination of the following:  
- on a specific date  
- between two specific dates  
- at a specific time of day  
- at set intervals to include daily, weekly, monthly, semi-annually, annually  
- at shorter intervals (e.g., a number of hours) for exceptional events (e.g., natural disasters) |
| 4     | Develop the ability to capture sites with active content, for example, .PHP and .ASP files. |
| 5     | Curator collaboration so that a specified group of curators, both from a single campus or multiple campuses, share authority and access to joint collections. |
| 6     | Generate a report that compares captures so that files that were added or deleted can be readily identified. |
| 6     | Explain error messages, such as why files were not captured (e.g., server restrictions or capture parameters). |
| 7     | For multiple captures of same site, provide an option to only retain non-redundant data. Keep records of the capture dates and times for (a) fully redundant captures that are not retained and (b) specific redundant data/files that are not retained. |
| 7     | Access (e.g., via category or subject searches) to all sites in the archive for possible inclusion in collections, regardless of which curator captured a site. Allow curators to request permission to include a site in a collection from the original curator who archived the site. |
| 8     | Include a field for recording selector's notes about a site. Notes might inform future selectors of the importance of a site, highlight particularly relevant sections of a site, explain why capture parameters were chosen, or state the relationship of a site to a collection. Guidelines for what to include in this field are advisable. |
| 8     | Give curators the option to 'override' robot exclusions if they have received permission from the web site owner. |
| 8     | Create a “perma-link” or “stable URL”, similar to a “tinyurl bibpurl”, for collections, individual files, and captures, so catalogs, websites, and email messages can include the links. |
| 8     | Ability to export specific file types (like PDFs) to another database or for access from a subject guide in order to publicize and transmit specific files to users, much as articles are downloaded and transmitted to patrons. |

* Tied scores were assigned the same rank.

The top enhancement ideas cover a fairly broad range of ideas. However, five of the top ideas relate to two areas: collaboration among curators and comparison of capture results. It has been known since early in the project that curators desire to collaborate with one
another in building and managing collections. The desire to discover and use materials captured by any curator is another, more general, type of collaboration of importance to curators. Curators also want to easily identify changes in capture results; three separate enhancement ideas relate to comparison of capture results to identify changes.

Scheduling options for captures continue to be a top priority. Curators identified a variety of applications for scheduling options, ranging from a number of hours in the case of exceptional events (e.g., natural disasters), to annually, within a date range (e.g., for annual reports from government agencies).

While some curators seek to capture discrete publications, others desire to capture materials and web sites in their entirety. One challenging idea, and for five curators an extremely important idea, is the capability to capture active web pages, such as PHP and ASP pages. One archivist in the curator group commented that the “ability to capture active content is the most important enhancement”. For three others, the ability to override robot exclusions, with permission of the content provider, was extremely important in order to capture desired content.

Not surprisingly some curators are looking ahead to patron access to the archived materials. They want to both export materials from the archive and link to materials in the archive from their own environments. Curators also want more information regarding WAS errors so that they can improve their capture results based on better information. Finally, curators were quite receptive to the addition of a curatorial notes field in site descriptions to provide a record of curatorial decisions for future reference.

4.4 Discussion

Curators responded positively to improvements from one WAS Release to the next. They noted and appreciated the improvements in workflow, the responsiveness of the development team to their input, and the addition of new features. They consistently commented upon the clean design of the interface and the ease of navigation. The help features were well received and consistently improved. By Release 4, 76% of the curators rated the side bar information as “very helpful”.

Curators were eager both to capture sites and then to display the sites they captured. Satisfaction was measured by asking them: “Overall, how satisfied were you with the display of your captured content?” The results in Table 4 illustrate the improvements curators generally noted between Releases 2/3 and 4.

Table 4. Curator Satisfaction with Displayed Content

<table>
<thead>
<tr>
<th></th>
<th>R 2/3</th>
<th>R4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Satisfied</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>% 29%</td>
<td></td>
<td>47%</td>
</tr>
<tr>
<td>Somewhat Satisfied</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>% 41%</td>
<td></td>
<td>35%</td>
</tr>
<tr>
<td>Not Satisfied</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>% 24%</td>
<td></td>
<td>18%</td>
</tr>
<tr>
<td>Not Tried</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>% 6%</td>
<td></td>
<td>0%</td>
</tr>
</tbody>
</table>

Each survey asked curators to state the two best features in the release and the two ways the WAS might be improved. This feedback provided ongoing assurances of what WAS features and functions were working well for curators and where improvements were desired. Additionally, the ideas for improvements were used to develop the list of
enhancement ideas for future WAS Releases that were prioritized by curators using a Delphi Technique.

Given the relatively small number of curators ($n=12$) who submitted final ratings for the enhancement ideas, caution is advised when considering the results reported in Section 4.3. The rankings can be better appreciated and applied in light of both the tabulated results (i.e., the number of curators rating ideas as either “extremely important”, “very important”, “moderately important”, of “very little importance”, or “not important”) as well as curators’ statements regarding the benefits and disadvantages of the ideas, which are included in the full report. That said, the results do represent the range of the project’s curators fairly well, as indicated by their participation level in the trial of WAS Release 4. Additionally, eight of the nine most active curators during that trial participated in the ratings, so presumably their ratings were influenced by a fair amount of familiarity with the WAS.

From the outset of the enhancement prioritization exercise, curators were informed that it may not be feasible to develop their priority enhancement ideas in future WAS releases. However, some of the enhancement ideas are already scheduled for implementation in later releases and the remaining ideas offer the WAS development team insight into curators’ requirements. In a larger sense, the ideas also identify a number of desirable features for future applications seeking to address the needs of curators building collections of web-published materials.

5 Summary

A few key issues and concerns discovered during the needs assessment were shared by the project’s curators. These were addressed either in other assessment activities or in the Web Archiving Service. These concerns are listed below and followed by a discussion.

- Selection
  - What web-published materials should be saved?
  - What materials are others already saving?
- Acquisition
  - How can changes in web-published materials be determined?
- Organization
  - How can quality subject access for searching and browsing of archives be provided without resource-intensive metadata creation?
- Presentation
  - Can web archives/crawlers meet archival standards for capturing the entire content of web sites and faithfully rendering the look-and-feel of the original source?

Selection

During the focus groups, two basic questions librarians asked in regard to identifying web-published materials for preservation were: “Should we save this?” and “Is someone else already saving it?” Curators in the Web-at-Risk project asked similar questions. Their web collection plans helped focus curators both locally on what web-published materials they “should” save and facilitated collaboration with other curators for larger collection efforts, for example, the web-based publications of California government agencies.

In regard to discovering what materials other curators were capturing, a future release of the WAS will allow curators to collaborate as they build collections. Searching and discovering the contents of the entire WAS archive is desirable to many curators. This is similar to a finding of the needs assessment that indicated a need for a registry of archived
materials in any archive, so that librarians or curators could consult that source prior to initiating a capture. In the future it seems likely that shared registry services among web archives at different institutions would benefit librarians as they build collections of web-published materials.

**Acquisition**

In regard to the capture of web-published materials, focus group participants were generally concerned with the frequency with which these materials change. Curators responding to the needs assessment survey identified important considerations for collection building practices including:

1. Assessing the change rate of the source materials
2. Establishing the interval at which collection materials will be captured

Two years after reporting these requirements for collection planning in the needs assessment, curators echoed these concerns in their prioritization of WAS enhancements. In regard to assessing the change rate in source materials, curators expressed the need for the WAS to compare a site’s capture results with the results of a previous capture of the same site and clearly indicate if the site had changed. Further, curators want to know what files changed and some want the WAS to only retain non-redundant data. Curators may be responding, at least in part, to the vast numbers of files often resulting from captures. Additionally, some curators are only collecting discrete publications and they have no need to capture a publication again if it has not changed since the last capture.

In regard to establishing the frequency with which a site is captured, curators identified a range of options applicable to the various types of materials and web sites in their collections. Eliminating the need for intensive “eyes-on” decisions about retention of captured materials and websites is a priority for curators. Both this need and the need to know if some other curator has already captured materials of interest supports a strong desire to avoid duplication of effort.

**Organization of Materials**

Both researchers and librarians participating in the needs assessment indicated a need for users to search and browse archives by subject or topic. Directly related to this need, one enhancement idea curators considered was to “include a controlled vocabulary for subject headings and the ability to modify it for specific collections or topical areas.” Some curators clearly saw a need for this; however, there was not general agreement that this was a top enhancement priority. In the priority ranking, the idea ranked 18th of 23, with six curators rating this idea as either ‘very important” or “extremely important” and six curators rating it as either “moderately important” or of “very little importance”. The divergence of views is represented in the following curator comments regarding the benefits and disadvantages of this idea.

- **Benefits**
  - consistency of subject terms
  - better searching, especially in conjunction with a keyword search
  - especially useful for place names
  - enabling controlled vocabulary searching
  - would help in assigning descriptors
  - would help the end user navigating the site
  - When multiple curators are collaborating, prevents use of synonymous terms in a single collection.
  - co-locate sites by subject; very useful for grouping sites
Disadvantages
- time-consuming
- new terms need to be added frequently

More curators commented on the benefits than the disadvantages of the WAS providing a controlled vocabulary, which may reflect a common understanding of the value of such vocabularies for material organization and access. Yet, it seems that the resource-intensive nature of the creation and maintenance of controlled vocabularies is a major deterrent to curators. This, in conjunction with curators’ desire to avoid duplication of effort in capturing materials, speaks to the high value curators place on their limited resources.

Presentation: Content v. Context
Content providers, researchers, and librarians participating in the needs assessment all identified types of web content that could be preserved in archives without benefiting from being rendered in their original context in the future. Statistical datasets are a good example of this type of web content. However, for some researchers and librarians, capturing web sites in their entirety so that they can be rendered in the future to mirror the original source material is critical. Historical research in many disciplines, which often relies on the analysis of change over time, is a good example of the need to preserve and present web-published materials in their original context. The issue of capturing websites in their entirety also emerged in the enhancement ideas in terms of wanting the WAS crawler to be able to capture active content, a current limitation in regard to certain file formats. For preservation of both historical and archival records, capturing web sites in their entirety is a top priority.

Conclusion
The assessment activities in the Web-at-Risk built upon one another and complemented the developmental stages of the Web Archiving Service. The initial needs assessment both informed the collection planning guidelines and provided functional requirements for the WAS. Curators’ collection plans identified the web sites and web-published materials they captured using the WAS. The survey evaluations and usability testing identified both what features and functions of the WAS were well-received by curators as well as areas where improvements were needed. Finally, the evaluations provided enhancement ideas that curators prioritized for possible development in future WAS releases. In the future, it is of interest to expand the WAS user base to discover if selectors in other academic concentrations have special needs and if the WAS design works well for the larger community of information professionals, beyond the government information professionals who participated as curators.

In conclusion, curators were asked in their evaluation of WAS Release 4: “Overall, how satisfied are you with this WAS release?” As a testimonial to the success of the Web Archiving Service, 65% indicated they were “very satisfied” and 24% were “somewhat satisfied.”
## Appendix A: Web-at-Risk Project Curators

<table>
<thead>
<tr>
<th>Institution</th>
<th>Curator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona State Library</td>
<td>Richard Pearce-Moses</td>
</tr>
<tr>
<td>New York University</td>
<td>Peter Filardo</td>
</tr>
<tr>
<td>New York University</td>
<td>Michael Nash</td>
</tr>
<tr>
<td>Stanford University</td>
<td>Elizabeth Cowell</td>
</tr>
<tr>
<td>Stanford University</td>
<td>Charles D. Eckman</td>
</tr>
<tr>
<td>Stanford University</td>
<td>John Eilts</td>
</tr>
<tr>
<td>Stanford University</td>
<td>Ann W. Latta</td>
</tr>
<tr>
<td>UC Archivists Council</td>
<td>Steve Coy</td>
</tr>
<tr>
<td>UC - San Diego</td>
<td>Lisa Mix</td>
</tr>
<tr>
<td>UC - San Francisco</td>
<td>Connie Williams</td>
</tr>
<tr>
<td>UC - Office of the President</td>
<td>Tim Smith</td>
</tr>
<tr>
<td>UC - Berkeley</td>
<td>Jim Church</td>
</tr>
<tr>
<td>UC - Berkeley</td>
<td>Ron Heckart</td>
</tr>
<tr>
<td>UC - Berkeley</td>
<td>Terence K. Huwe</td>
</tr>
<tr>
<td>UC - Berkeley</td>
<td>Amy Kautzman</td>
</tr>
<tr>
<td>UC - Berkeley</td>
<td>Frank Lester</td>
</tr>
<tr>
<td>UC - Berkeley</td>
<td>Nick Robinson</td>
</tr>
<tr>
<td>UC - Berkeley</td>
<td>Jesse Silva</td>
</tr>
<tr>
<td>UC - Davis</td>
<td>Linda Kennedy</td>
</tr>
<tr>
<td>UC - Davis</td>
<td>Juri Stratford</td>
</tr>
<tr>
<td>UC - Davis</td>
<td>Marcia Meister</td>
</tr>
<tr>
<td>UC - Irvine</td>
<td>Yvonne Wilson</td>
</tr>
<tr>
<td>UC - Los Angeles</td>
<td>Gabriella Gray</td>
</tr>
<tr>
<td>UC - Los Angeles</td>
<td>Kris Kasianovitz</td>
</tr>
<tr>
<td>UC - Riverside</td>
<td>Ken Furuta</td>
</tr>
<tr>
<td>UC - Riverside</td>
<td>Lynne Reasoner</td>
</tr>
<tr>
<td>UC - San Diego</td>
<td>Megan Dreger</td>
</tr>
<tr>
<td>UC - San Diego</td>
<td>James R Jacobs</td>
</tr>
<tr>
<td>UC - San Diego</td>
<td>Annelise Sklar</td>
</tr>
<tr>
<td>UC - San Francisco</td>
<td>Kathleen Cameron</td>
</tr>
<tr>
<td>UC - San Francisco</td>
<td>Kirsten Nielsen</td>
</tr>
<tr>
<td>UC - Santa Barbara</td>
<td>Sherry DeDekker</td>
</tr>
<tr>
<td>UC - Santa Barbara</td>
<td>Janet Martorana</td>
</tr>
<tr>
<td>UC - Santa Cruz</td>
<td>Lucia Orlando</td>
</tr>
<tr>
<td>University of North Texas</td>
<td>Valerie Glenn</td>
</tr>
<tr>
<td>University of North Texas</td>
<td>Starr Hoffman</td>
</tr>
<tr>
<td>University of North Texas</td>
<td>Arlene Weible</td>
</tr>
</tbody>
</table>
Appendix B: Reports and Curator Tools

NEEDS ASSESSMENT

Reports

Needs Assessment Summary Report

Survey Report

Focus Group Reports

ALA – Chicago

University of North Texas – Denton

California Digital Library – Oakland

New York University - New York City

Federal Depository Library Conference - Wash DC

Interview Reports

Content Provider Interviews: Summary Report

End User Interviews: Summary Report

Curator Tools

Assessment Toolkit

Survey

Focus Group Discussion Guide

Focus Group Participant Questionnaire
End User Interviewer Questionnaire  

Content Producer Interview Questionnaire  

**COLLECTION PLANNING**

**Reports**

Collection Planning Guidelines  

**Curator Tools**

Web Collection Planning Overview  

Collection Plan Template  

**WAS EVALUATION**

**Reports**

WAS Release 1 Survey Results  

WAS Release 2/3 Survey Results  

WAS Release 4 Survey Results  

WAS Release 4 Usability Test Report  

Prioritization of WAS Enhancement Ideas  
Appendix C: Publications & Presentations

Publications


Presentations


