



Prioritization of WAS Enhancement Ideas

December 21, 2007

Prepared by:

Kathleen R. Murray
kmurray@library.unt.edu

University of North Texas

Contents

| | |
|---|----|
| Overview | 2 |
| Methodology | 2 |
| Results | 3 |
| Discussion | 4 |
| Appendix A. Initial WAS Enhancement Ideas | 6 |
| Appendix B. Curator Refinement and Feedback | 7 |
| Appendix C: WAS Enhancement Ideas: Listed by Ranks with Data Tables | 25 |
| Appendix D. WAS Enhancement Ideas: Listed by Idea Number | 36 |

Overview

Several enhancement ideas for future releases of the Web Archiving Service (WAS) have been identified. In order to refine and prioritize these ideas as well as elicit new enhancement ideas, 16 project curators participated in an online exercise between November 27 and December 16, 2007. The exercise employed a Delphi Technique that consisted of a series of three online questionnaires, each available for a five or six day period.¹ This report describes the exercise and reports the results.

Methodology

Questionnaire 1

The purpose of the first questionnaire was (a) to solicit comments about known enhancement ideas and (b) to generate new ideas. Appendix A lists the 25 WAS enhancement ideas included in the first questionnaire. These were primarily suggested by curators in their evaluations of WAS Releases 2/3 and 4. Curators were asked to consider the following question as they completed the questionnaire:

What feature enhancements to the Web Archiving Service could be made to better address your needs for collection and preservation of web-published materials?

Curators were asked to refine the enhancement ideas by adding any clarifications and to list any benefits or disadvantages associated with the ideas. At the end of the questionnaire, curators could add any new ideas for enhancements, including their benefits and disadvantages. Thirteen curators submitted the first questionnaire.

Questionnaire 2

The initial list of ideas was refined to incorporate curators' clarifications and 10 new ideas. Some initial ideas were merged and one idea was deleted. The second questionnaire consisted of 31 ideas. Curators were given access to a document reporting all curators' comments from the first questionnaire regarding the benefits and potential disadvantages of each idea. (See appendix B.)

In the same manner as the first questionnaire, curators were asked to consider:

What feature enhancements to the Web Archiving Service could be made to better address your needs for collection and preservation of web-published materials?

Likewise, curators were again asked to further refine the enhancement ideas and list any benefits or disadvantages associated with each idea. No new ideas were solicited. Eight curators submitted the second questionnaire.

Questionnaire 3

Two of the 31 ideas from the second questionnaire were further clarified based on curators' comments. Any benefits and disadvantages reported on the second questionnaire were added to those from the first questionnaire and provided to curators for consideration in their final ratings. (See appendix B.)

¹ Of the 16 curators who participated in this exercise, seven submitted all three questionnaires, three submitted two questionnaires, and six submitted one questionnaire.

In the same manner as the earlier questionnaires, curators were asked to consider the following question as they rated each idea:

What feature enhancements to the Web Archiving Service could be made to better address your needs for collection and preservation of web-published materials?

Questionnaire 3 listed the 31 ideas and curators rated the importance of each idea on a 5-point scale: Not important; Very little importance; Moderately important; Very important; and Extremely important. Twelve curators participated in rating the enhancement ideas.

Results

Each of the ratings was assigned a value as follows: 0 = Not important; 1 = Very little importance; 2 = Moderately important; 3 = Very important; and 4 = Extremely important. The average rating for each idea was calculated and these averages were used to rank-order the enhancement ideas.

Table 1 reports in rank-order the 13 ideas whose averages were 3.00 or higher. In the case of tied scores, ideas were given the same rank. Appendix C lists all the ideas in rank-order and includes the tabulated ratings for each idea. For ease of reference, appendix D lists the ideas, including their ranks, in the order they were presented in the second and third questionnaires as well as the order they are listed in appendix B.

Table 1
Ranked Enhancement Ideas with Average Scores of 3.00 or Higher

| Idea # | Average | Rank* | Idea |
|--------|---------|-------|---|
| 22 | 3.67 | 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. |
| 4 | 3.58 | 2 | 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. |
| 10 | 3.45 | 3 | Schedule captures based on one or a combination of the following: <ul style="list-style-type: none"> ▪ 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) |
| 12 | 3.36 | 4 | Develop the ability to capture sites with active content, for example, PHP and .ASP files. |

| Idea # | Average | Rank* | Idea |
|--------|---------|-------|--|
| 1 | 3.33 | 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. |
| 20 | 3.25 | 6 | Generate a report that compares captures so that files that were added or deleted can be readily identified. |
| 26 | 3.25 | 6 | Explain error messages, such as why files were not captured (e.g., server restrictions or capture parameters). |
| 23 | 3.08 | 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. |
| 24 | 3.08 | 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. |
| 6 | 3.00 | 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. |
| 27 | 3.00 | 8 | Give curators the option to 'override' robot exclusions if they have received permission from the web site owner. |
| 28 | 3.00 | 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. |
| 30 | 3.00 | 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, such as articles are downloaded and transmitted to patrons. |

* Tied scores were assigned the same rank.

Discussion

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.

Given the small number of curators ($N=12$) who rated the enhancement ideas, caution is advised when considering the results. The rankings can be better appreciated and applied in light of both the tabulated results in Appendix C and the benefits and disadvantages of the ideas recorded in Appendix B. That said, the results do represent the range of the project's curators fairly well, as indicated by the range of their participation level in the trial of WAS Release 4. Additionally, eight of the nine most active curators during that trial submitted ratings, so presumably their ratings were influenced by a fair amount of familiarity with the WAS.

From the outset of the exercise, curators were informed that it may not be feasible to address their priority enhancement ideas in future WAS releases. One curator requested that in addition to a list of the top priorities for enhancements, it would be helpful to know "which will be most likely implemented, including a timeline of releases where they will be rolled out." 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.

Appendix A. Initial WAS Enhancement Ideas

| | Ideas |
|-----|---|
| 1. | Ability to determine if another curator has already defined and captured a site |
| 2. | Capture options for "Directory +1 link" and "Host +1 link" (Note from author: "Host" was a mistake; it should have been "Page".) |
| 3. | Schedule capture based on a specific date |
| 4. | Schedule capture based on specific time of day |
| 5. | Specify capture frequency: daily, weekly, monthly, annually |
| 6. | Specify file type(s) to be captured from a site |
| 7. | Include a field for recording selector's notes about a site |
| 8. | Include a subject heading thesaurus |
| 9. | Include some Overview Report data in the Manage Site screen |
| 10. | Curator collaboration so that a specified group of curators, for example all curators at one campus, can view the entire archive or a designated portion of it |
| 11. | Search and display multiple capture results for any site in the archive, regardless of which curator captured a site |
| 12. | Simultaneously browse two or more capture results for the same site |
| 13. | View site captures as a tree structure with filenames |
| 14. | Thumbnails of captured sites' home pages in both the 'Manage Sites' and 'View Captures' areas so that sites can be more readily identified |
| 15. | Easier method(s) of dealing with the volume of files in the files list |
| 16. | Ability to print the list of files in its entirety |
| 17. | When viewing PDF files with active links, display the URL for the link along with the linked file |
| 18. | For multiple captures of same site, indicate in capture results if site changed since last capture and retain only non-redundant data |
| 19. | For multiple captures of same site, retain only non-redundant data |
| 20. | Access to all sites in the archive for possible inclusion in collections, regardless of which curator captured a site |
| 21. | Selecting "Help" on any screen opens the general help document at the relevant section for that screen |
| 22. | Explain error messages, for example, why files were not captured |
| 23. | Optionally ignore robot exclusions |
| 24. | 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 |
| 25. | Identify screen resolution and browser settings for optimal display of the user interface |

Appendix B. Curator Refinement and Feedback

Appendix B reports curators' comments regarding the benefits and potential disadvantages of each idea. Curators' clarifications of the initial list of ideas were incorporated into a second list of 31 ideas. Some initial ideas were merged and one idea was deleted. New ideas that emerged in the first round of feedback were also included. Only ideas 2 and 18 were further clarified based on the second round of curator feedback. Comments from both the first questionnaire and the second questionnaire are included.

| | |
|--|--|
| 1. Idea | Curator collaboration so that a specified group of curators, from either a single campus or multiple campuses, share authority for and access to joint collections. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Sounds very beneficial--I'm envisioning that collaborative collection development of this sort could become one of the larger uses of the WAS. ▪ Inter-institutional cooperation needed to avoid duplication of effort ▪ No opinion ▪ Important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ This is especially useful if curators are collaborating to build the same collection or if multiple curators are collecting materials about the same theme (e.g., a specific natural disaster) or subject area (e.g., state government). ▪ Allows for work to be split up between curators. Allows for larger collections--some group members may discover sites that others do not. ▪ Enable coordination between curators. ▪ Enables groups of curators . . . to work together, view each other's captures, develop common standards, and avoid duplication. ▪ Very good idea. Collaborative online collection building. ▪ Enables more efficient collaboration between curators working on similar projects, or on larger collections where the work is divided between curators. ▪ essential for collaboration ▪ Seems like a nice feature. ▪ Could be useful for collaborating on collections. ▪ is valuable for improving future captures |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ I worry about overwriting work--need clear distinction between 'versions' of captures. |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Could be too large to be useful if a campus (or other grouping) were to capture hundreds of sites. ▪ May be difficult to set up user rights for every curator. Curators may inadvertently interfere with each other's work. ▪ Possibility that a curator could inadvertently overwrite another's capture if this feature is not set up w/ safeguards against that. ▪ Need mechanism so sites can't be inadvertently deleted, or settings changed by another curator. This is where the ability to include curator notes would be helpful. |

| | |
|--|---|
| <p>2. New Idea in Q2</p> | <p>An automated workflow that allows curators to identify sites for inclusion in collections and allows non-curatorial staff to create site descriptions, schedule captures, and evaluate capture results. This might be accomplished by associating different levels of authority with WAS user IDs based on users' job responsibilities.</p> |
| <p>Benefits: 2nd Questionnaire</p> | <ul style="list-style-type: none"> ▪ This would certainly enable curators to free up some time by delegating work to others--I'm in favor. A workflow would ensure that the work was reviewed and approved before final acceptance, which would make things run smoother and more reliably. ▪ YES! Any automated process that reduces workload will help us make more efficient use of the WAS by allowing curators to delegate more routine tasks. ▪ Very important, as we expect to capture a large number of sites, and to employ non-curatorial staff as indicated ▪ No opinion ▪ Important. |
| <p>Disadvantages: 2nd Questionnaire</p> | <ul style="list-style-type: none"> ▪ It may be easier to create this workflow off line. |
| <p>3. New Idea in Q2</p> | <p>For the sites they define and capture and the collections they build, allow curators to set access levels for other curators. Access levels might include full editing permission (i.e., full collaboration), search and display permission, or permission to include a site's capture(s) in another curator's collection.</p> |
| <p>Benefits: 2nd Questionnaire</p> | <ul style="list-style-type: none"> ▪ Again, anything that encourages collaboration among curators sounds fantastic to me. ▪ Would prevent other curators from inadvertently destroying an important capture/setting. ▪ Could be very useful for collaboration. ▪ Important. |
| <p>Disadvantages: 2nd Questionnaire</p> | <ul style="list-style-type: none"> ▪ If the first curator leaves, who then can access the site? ▪ Not sure why permission to include a site captured by another curator in a collection one is building is necessary -- after all, the site is already public. |
| <p>4. Idea</p> | <p>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.</p> <p>NOTE: This idea merges two ideas from the initial list. The benefits and disadvantages of both initial ideas are included.</p> |
| <p>Benefits: 2nd Questionnaire</p> | <ul style="list-style-type: none"> ▪ This would be so, so wonderful to not duplicate work! This seems quite important in the digital world--we need to regularly practice not duplicating preservation efforts, so that time can be spent on new captures. ▪ Helpful for those of us interested in collaborating ▪ Essential to avoid wasted effort |

| | |
|--|--|
| | <ul style="list-style-type: none"> ▪ Can help avoid duplication. ▪ Very important. Would help share expertise and not 'reinvent the wheel'. Would want the access to be 'Read Only' |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Other curators may be capturing a California or U.S. government site that would fit in our collection as well ▪ Obviously a duplication of efforts must be avoided as much as possible. ▪ Eliminates duplicated effort. ▪ Better coordination between multiple staff. ▪ avoid duplication ▪ There is no reason to duplicate searches. ▪ This would be a better allocation of resources. ▪ Duplication of captures could be avoided. ▪ Allows curator to decide if they want to follow through with capture. Automated way for curators to learn about other collections with similar scope or subject overlap. ▪ necessary for collaboration ▪ Enhance collaborative collection development. ▪ Reduces redundant captures. ▪ Useful for coordinating with other selectors. ▪ May help curators avoid duplicating effort and multiple captures. Makes the tool more usable as a search tool. ▪ Enable coordination between curators. ▪ This sounds like it could be a useful discovery tool. ▪ Curators may not be able to meet physically, but they could view results jointly and confer by phone. ▪ for researchers this is essential to effectively compare; for curators, this is essential for collaboration ▪ Would allow curators to share expertise and reduced 'reinventing the wheel'. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ There probably needs to be a way to add the capture of another curator to a collection without giving the second curator the ability to change the definition of the capture. ▪ Archive could be too large to be useful. ▪ . . . if searched sites captured by another user could [not] be included in one's collections, the site would still have to be [captured] twice for that purpose. |
| 5. Idea | Add site parameters for capture scope to include "Directory +1 link", where a directory and only the linked content on pages within that directory are captured, and "Page +1 link", where only a specific page and the linked content on that page are captured. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This would be helpful for collections of documents-only (not complete websites). I'm in favor, although I don't think I would use this feature a lot. ▪ Would allow ability to tightly constrain the capture parameters - narrow the focus of a capture to a single location/page instead of having to capture the entire site. ▪ N/A for [our institution] |

| | |
|--|--|
| | <ul style="list-style-type: none"> ▪ Could be useful when information desired on a site is predictably limited to the one directory. ▪ Very important. I think the focus should be on enhancements to the tools ability to successfully capture content first, then work on other 'after the content is captured' enhancements. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ More precision of capture: one can capture only the parts of the site needed but still link to important external pages ▪ May free up capture time. ▪ Shorter searches and less review time ▪ It would provide more control over the scope of the capture. This might also address some copyright issues. ▪ also very important to capture the look and feel of a site -- key archival concern ▪ The more capture options available will increase successful capture results. ▪ 'Directory +1 link' is a benefit when one wants to capture down stream + 1 link, rather than the host. All of my Release 4 captures were targeted directory captures that might have benefited from the link. ▪ More flexibility for configuring the capture is better. Best would be flexibility to configure directory or host AND number of links from 1-5. The Web Archives Workbench has this functionality. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Not everyone seemed to have the same concept of the host + 1 link at the last meeting. ▪ Another layer of potential confusion ▪ You wouldn't always know what was included in a single directory. Also, the links would probably be to other directories in the same site. I don't know if that is a huge problem. ▪ There will always be address to add the 'do not crawl' list but they will fewer. |
| 6. Idea | <p>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.</p> |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This is a must-have. Our collections won't really be cohesive or make much sense unless we have this ability. ▪ Essential to have an abstract -- standard archival description ▪ Valuable for the continuation of capturing sites, even after the specific surator who began capturing has moved on. ▪ Very important. This would be very helpful in terms of building a knowledge base and curator capture expertise. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Allows for consistency when curatorial duties are taken over by a new staff member. Allows for consistency if a selector has forgotten how/why something was done in the past. ▪ Reminders, notes about problems with previous captures, etc. ▪ Very good idea. Enhanced metadata. ▪ These notes would also help other people understand the crawl. |

| | |
|--|---|
| | <ul style="list-style-type: none"> ▪ It gives the person looking at the capture an understanding of the criteria used in defining the capture. ▪ Helps curators manage captures in a single environment (as opposed to post-it note reminders). Especially helpful if more than one curator is capturing site or if site is transferred to another curator in the future ▪ very useful ▪ We use a paper based version of this feature extensively. ▪ Sure |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ These fields are only as good as what selectors put in them. Selectors still may not use these fields or leave detailed notes. ▪ Curator should have option to make comments shared (with other curators) or public (all users) or private (lead curator only) |
| 7. Idea | Include a controlled vocabulary for subject headings and the ability to modify it for specific collections or topical areas. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ a thesaurus of frequently used terms will save time and promote authority control; ideal would be 'hot-key' access, as in cataloging systems, to all the terms used in that library's opac |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Allows for consistency of subject terms, which makes for better searching, especially in conjunction with a keyword search. Will be especially useful for place names. ▪ Eventually enabling controlled vocabulary searching ▪ Hm.... Not sure. What subject headings? ▪ consistent metadata ▪ This would help in assigning descriptors. It could also help the end user navigating the site. ▪ Shared/controlled vocabulary. When multiple curators are collaborating prevents use of synonymous terms (ex: SP News versus News-Spanish) in a single collection. Co-locate sites by subject ▪ very useful for grouping sites -- essential for consistent description -- it is also essential to be able to add local terms |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Not sure that this is necessary as the free text field now allows this. ▪ Seems like a lot of trouble (multiple thesauri??). Would not be high on my list of enhancements to work on first. |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Creating a thesaurus is time-consuming. ▪ May depend on the thesaurus used. ▪ Could take a long time to develop a thesaurus and new terms need to added frequently. ▪ Additional time required to find and assign subject headings ▪ Seems lees important then other enhancements listed. |

| | |
|--|---|
| 8. New Idea in Q2 | In the description of a site, include the collection(s) to which it belongs. This should be automatically generated when sites are added to collections. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Allows the curators to see the overall structure quickly and from multiple modules. ▪ This would be fantastic, particularly if automated. ▪ prevents needles re-entry; important in that it allows collection assignment to be done later ▪ Might be a good idea. ▪ Less important. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 9. New Idea in Q2 | When capturing linked content, allow curators to limit captured content by specifying addresses or domains whose linked content should be included (e.g., a host address or the .gov domain) and, conversely, allow curators to specify domains whose linked content should be excluded (e.g., .com.). |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Would help focus the results to only those of interest. One of the most popular sites has been Adobe, for users needing to download the reader. If a specific similar site could also be defined out, then it would tighten the results. ▪ This would avoid unnecessary capturing. Especially useful in avoiding archiving banner ads from commercial sites. ▪ N/A for [our institution], but agree in principle ▪ Any narrowing of parameters for capturing sites is a good thing. ▪ Would be nice but not sure how many curators would actually use (know how to use) this feature. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 10. Idea | <p>Schedule captures based on one or a combination of the following:</p> <ul style="list-style-type: none"> ▪ 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) <p>NOTE: This idea merges three scheduling ideas from the initial list. The benefits and disadvantages of all three initial ideas are included.</p> |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ These would be great abilities--not absolutely necessary, but quite useful. It would really help the ability to stagger times effectively to not overload the servers. ▪ This should be a priority ▪ Obviously essential -- we would use a combination of a set interval, plus fixed dates [elections, etc.] ▪ Very useful as the curator generally is familiar with sites to know how often or when a capture is needed to retain all |

| | |
|---|---|
| | <p>content and changes.</p> <ul style="list-style-type: none"> ▪ Very important. |
| <p>Benefits: 1st Questionnaire</p> | <p>Date</p> <ul style="list-style-type: none"> ▪ The changes to web sites are particularly important information. What has been removed, replaced, or newly added carries with it information about relationship to other events. ▪ Curators can catch known upcoming changes without having to log in and manually run a capture. Curators can capture new (or old) pages without duplicating pages already in the archive. Will allow users to capture coverage of a specific event (say, election day). ▪ Could be useful if you know updates occur on the 1st of every month or anticipate a major announcement on the 10th. ▪ Not very important for me, but perhaps for others, e.g people covering elections ▪ Reduce duplicated material and the number of crawls. ▪ This would allow the capture to be designed before the anticipated event. ▪ Provides automated mechanism for managing workflow, especially for sites that change in a predictable way (news and announcements, etc.) ▪ Extremely important for following key events such as elections ▪ Important, standard feature. ▪ Depends on the site ▪ Could be useful to set capture on low-traffic day for the site being captured. <p>Time of Day</p> <ul style="list-style-type: none"> ▪ Could be a better use of resources and remove contention for the resources, including bandwidth. ▪ Same as with scheduling based on date: Curators can catch known upcoming changes without having to log in and manually run a capture. Curators can capture new (or old) pages without duplicating pages already in the archive. Will allow users to capture coverage of a specific event (say, election day, right after the polls close). ▪ I'd schedule for late night hours on the assumption that the process will be faster due to lower demands on the server. ▪ Not very important for me, but I can see why it might be for others doing time-sensitive events. ▪ This could result in more efficient searches. The searches would be less likely to be rejected due to server load issues. ▪ Allows scheduling for off-peak times (night) of busy sites [i.e., Curator could schedule a one time initial capture or repeat captures on a pre-determined schedule. (Example: capture site at 10:00PM on the 15th of every month using initial capture settings)] ▪ do not see the need for this for our collections ▪ Depends on the site ▪ Could be useful to set capture at low-traffic time for the site being captured. <p>Frequency (daily, weekly, monthly, etc.)</p> <ul style="list-style-type: none"> ▪ Changes over time are significant in the behavior of some |

| | |
|--|--|
| | <p>organizations.</p> <ul style="list-style-type: none"> ▪ Curators will be able to track site updates on a constant and consistent basis. ▪ This is important, especially if the time of day could also be specified. I could schedule captures during off peak hours. ▪ This has obvious advantages as it allows for regularly scheduled captures, which can be based on the perceived frequency of web site updates. ▪ The aim should be to develop focused automatic crawls that show new material. ▪ necessary to capture a representative set of iterations, i.e., an archival record of the changes in the site ▪ Important. ▪ This would be a benefit for web publications like student newspapers, newsletters and content that is regularly published ▪ Very important for capturing sites that refresh content frequently. |
| <p>Disadvantages: 2nd Questionnaire</p> | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| <p>Disadvantages: 1st Questionnaire</p> | <p>Date</p> <ul style="list-style-type: none"> ▪ May make for some uncertainties in captures. Captures may be empty or filled with unexpected data if the wrong date parameters are set. ▪ You could run into problems when two or more curators are looking at the same sites. ▪ Could curator override if the content changes or doesn't change on that date? Would need a list of scheduled captures so curator will remember which sites are on auto-capture and their frequency. <p>Time of Day</p> <ul style="list-style-type: none"> ▪ Same as with scheduling based on date: May make for some uncertainties in captures. Captures may be empty or filled with unexpected data if the wrong date parameters are set. ▪ Not so important. <p>Frequency (daily, weekly, monthly, etc.)</p> <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| <p>11. Idea</p> | <p>Specify file type(s) (e.g., audio, video, and document) or file extensions (e.g., PDF, DOC, MP3, and AVI) to be included or excluded from a site capture.</p> |
| <p>Benefits: 2nd Questionnaire</p> | <ul style="list-style-type: none"> ▪ Would allow curators to sift through a site for more 'traditional' publications (ex: pdfs) and possibly export them to a local catalog (if exporting and persistent URLs are enabled) ▪ Required information for standard archival documentation practice ▪ Very important for certain sites. We have the need to capture recorded sermons from radical Islamic clerics and the remained of the site carries content of no scholarly interest. ▪ Very important. |
| <p>Benefits: 1st Questionnaire</p> | <ul style="list-style-type: none"> ▪ Some collections relate only to specific data on the sites. ▪ Curators can sift out the more official 'documents.' Especially |

| | |
|--|---|
| | <p>useful for hunting down online versions of government documents and other publications that might be buried within a site.</p> <ul style="list-style-type: none"> ▪ enable curators to weed out certain file types ▪ This is important as curators could specify especially desired file types, e.g. PDF, Word Excel. ▪ It might be desirable to exclude streaming audio, visual files, flash files, etc. ▪ Allows targeting specific content on site (reports, etc.) instead of entire site which may not change in any meaningful way ▪ Necessary for archival documentation of the provenance and content of the site ▪ Can envision gov docs librarians asking for all pdfs from a site. ▪ I would prefer a way to harvest file types after the capture and analysis. ▪ Could be useful if collecting documents only (tend to be pdf files). |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Newer information may be lost if the site's original creator changes a document's format when updating. ▪ I guess I tend to want to grab the entire site, rather than specific file types loaded on it. ▪ May be difficult to convey/display to users if they understand WAS captures entire site. Would need to connect these file-type-only captures with the captured site as a whole (if the site has been captured as a whole) or other file-type-only captures from the same site. |
| 12. New Idea in Q2 | Develop the ability to capture sites with active content, for example, .PHP and .ASP files. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Would allow us to capture more kinds of sites. Especially important as more and more sites are coded with cms's. ▪ Vital to capture look and feel of a site, and to capture a 'record' version of the site ▪ Valuable to capture the 'look and feel' of a site to recreate for the user the original experience of viewing the site at the time of capture. ▪ This would be my number 1 enhancement. Would greatly increase the ability to successfully capture sites and inform web archiving tool development in general |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 13. New Idea in Q2 | Ability to sort sites by the collection(s) to which they belong. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This is a must-have feature. ▪ The ability to sort and more importantly to search collections is crucial. ▪ Important for effectively using the manage sites module ▪ No opinion on this. |

| | |
|--|---|
| | <ul style="list-style-type: none"> ▪ Less important. Seems related to #8. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 14. Idea | <p>Provide easier method(s) of dealing with the volume of files in the files list by:</p> <ul style="list-style-type: none"> ▪ Increasing the file sorting options to include file name, server, and directory ▪ Displaying sites in a directory tree, from which clicking on a directory displays its files |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Helpful to understand and document structure ▪ Valuable in order to determine the success of the capture and how to present to end users. ▪ Less important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Also very important. I had a great deal of difficulty sorting and viewing all of these. Way too many files, apparently no logical order to them. ▪ It might be easier to read than a list of files representing the entire capture. ▪ while not really clear about the issues, all in favor of an improvement |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Seems less important than other enhancements listed. |
| 15. Idea | <p>View the structure of captured sites as a tree structure that includes filenames. If this is not feasible, show a site's structure as a tree that includes directories and the number and size of files within each directory by file type (e.g., /directory_name: 5 HTML files, 50kb; 2 PNG files, 60kb).</p> |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Helpful to understand and document structure ▪ No opinion on this. ▪ Less important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ I don't see this as very useful. ▪ Easier to browse the results. ▪ Would be useful with #2 above, 'Capture options for 'Directory +1 link' and 'Host +1 link''. Information from the 2 could help guide smaller future captures if the main interest is in just part of the site. ▪ helps to organize sites logically ▪ Yes! It is very important to see the results of a capture as 'the real web site' as perceived by a user. The main host site should be 'the trunk.' Logically ordered branches. The files as captured now seem to be sorted a big pile of junk! ▪ This could help in designating certain files for 'no capture or crawl' ▪ This might help refine site captures. ▪ Enable curator to quickly determine if major changes have been |

| | |
|--|--|
| | <p>made to the homepage.</p> <ul style="list-style-type: none"> ▪ greater understanding of site structure is intrinsically useful, it seems to me ▪ I think that this would be a very good idea and would add a way to navigate to the 'root' file, rather than going through screens of twenty filenames, one screen at a time (especially when you have 50,000 files). ▪ Very useful if you want to collect part of a site. Understanding the directory structure helps to configure the capture. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Showing a tree representing every file may be too much information to navigate all at once. ▪ Seems less important than other enhancements listed. |
| 16. Idea | Include the option to view thumbnails of captured sites' home pages in both the 'Manage Sites' and 'View Captures' areas so that sites can be more readily identified. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This would really help those visually-minded of us. ▪ A good idea ▪ Would make visual scanning of sites much quicker. ▪ Less important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Depending on the size of the thumbs! ▪ Would look snazzy. One could see major variations/site redesigns between versions without launching the site. ▪ Easier user interface... ▪ sites can be instantly identified. ▪ Good idea. ▪ very useful |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ More clutter on the capture and manage site pages. ▪ Seems less important than other enhancements listed. |
| 17. Idea | Provide the ability to print the list of files in its entirety. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Standard archival documentation practice requirement ▪ No opinion on this. ▪ Less important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Ability to keep a print record of what was captured. ▪ essential for the archival record -- we must be able to document the contents of the site |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This does not seem to be a good use of functionality of the WAS |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ That would have taken a week or so for me ▪ Seems less important than other enhancements listed. ▪ I don't know why I would want to do this. I think that generating a file comparison report (as mentioned in 13) and |

| | |
|--|--|
| | printing the report might be useful, but I wouldn't want to print a list of all (frequently 4000+) files. Downloading might potentially be useful, not printing a long list of files. |
| 18. New Idea in Q2 | Enable searching of the entire archive using curator name and subject. 'Subject' should be defined. For example, does it mean any subject keyword or the specific subject terms recorded in metadata? |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Good way to track what a person has done, especially if that person is yourself or your predecessor who didn't leave any notes. ▪ This would be useful for curators collaborating on similar projects or who have similar collection interests. ▪ Definitely a help if all you can remember is your name and a subject. ▪ of course -- just as one is able to search the entire contents of an OPAC; also for management by curators ▪ Might be useful. ▪ Not important. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 19. Idea | Allow simultaneous browsing of two or more captures for the same site, or of a live site and a capture, to enable comparisons and evaluate capture results. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This would be helpful, but it's hard to see how this could really work technically. ▪ essential for easy curator/researcher purposes ▪ Would be extremely useful, particularly if one needs to modify the capture parameters. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Would allow for more seamless browsing of the site, especially when page are missing in one capture but not another. ▪ Comparing results will help to see how often a site is updated. ▪ enable curators to compare different captures of the same site. ▪ This would allow you to see if the frequency for the captures made sense. If there was complete duplication, you might want less frequent. If there was no duplication you might want more frequent. ▪ Allows easier comparison of sites and ability to detect changes to site over time. ▪ essential for researchers [and curators] to effectively and efficiently compare iterations, a key activity for both ▪ Allow curator to view changes. ▪ Useful for identifying new content. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Browser might not realize that a capture was incomplete and that they are looking at an earlier/later version of the page. |
| 20. New Idea in Q2 | Generate a report that compares captures so that files that were added or deleted can be readily identified. |

| | |
|--|---|
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This is a good extension of #19 ('Allow simultaneous browsing of two or more captures for the same site...'). It partly automates the process and would assist in the comparison. In addition, it ties with #22 about changed results. ▪ This would be greatly helpful. ▪ Vital -- one should not have to hunt for this information ▪ Could be very useful. ▪ Very important. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 21. Idea | When viewing PDF files with active links, display the URL for the link along with the linked file. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ to provide direct access ▪ No opinion on this. ▪ Less important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ very handy |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Not sure I see the benefit. ▪ If these displayed within the PDF, it could distort the visual presentation of the PDF. ▪ Seems less important than other enhancements listed. |
| 22. Idea | 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. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ see #20 [Vital -- one should not have to hunt for this information] ▪ Extremely useful when the changes in a site impart important information about changes and how they may relate to external events. ▪ Very important. Goes with #20. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Quicker scanning. ▪ Would be useful for establishing a capture schedule for a site. ▪ Not a bad idea. Could save some time. ▪ This would help to better define site captures. For example, making the captures more frequent or less frequent. ▪ This would be useful in determining the frequency of future captures. ▪ Provides more information for the curator. ▪ essential -- so that researchers and curators can know if they want to/need to compare iterations ▪ Ranks high on my list of enhancements. ▪ Very useful. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |

| | |
|--|--|
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 23. Idea | For multiple captures of same site, provide an option to only retain non-redundant data. However, keep records of the capture dates and times for (a) fully redundant captures that are not retained and (b) specific redundant files that are not retained. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ saves storage space; OK as long as redundant data is secure ▪ Outstanding idea! ▪ Like options #20 and #22 better than this option. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Saves space. Perhaps saves crawling time or opens it up to items that have changed but might be missed because the crawl times out. ▪ useful in saving disk space ▪ Very good idea. Why repeat? ▪ This would help manage the size of the archive. ▪ There would not be many duplicates of PDF files that don't change between captures. ▪ Results in less file to capture and evaluate. Would be a good option for sites that don't change regularly - other than new content added in form of links to reports. ▪ One way to indicate new content. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Overhead in reconstructing compression of stored data and reconstructing various states of the site at various times may be more than the end result is worth. ▪ Not a complete site capture each time. ▪ The major downside would be in a later display of the site. All pages should display seamlessly no matter when it was captured. ▪ Without good cross-referencing between captures, this could make the captures difficult to navigate. ▪ There may be some use in knowing when an institution decided to take files off its web site. ▪ Would this interfere with the link structure within the capture? ▪ Seems less important than other enhancements listed. ▪ Any greater danger of permanent loss if there is only one 'copy' of data/files? |
| 24. Idea | 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. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ The ability to search a site or collection is crucial for future usability and sustainability of the collections we are capturing. It would be most helpful to be able to search within a site, across a specified set of captures for a single collection, and across a set of loosely related captures. Ability to search the entire WAS by keyword would also be helpful. |

| | |
|--|--|
| | <ul style="list-style-type: none"> ▪ Of course -- by analogy, one wants access to Worldcat, not just to one's individual OPAC ▪ Could be very useful in a collaborative effort. ▪ Nice feature. Rights issues? |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Could be useful for collaborative projects. ▪ Would save duplication of effort and server space for captures. ▪ Begins to create the archive 'library.' In the future this would also allow curators to see what sites/subjects were missing - collection analysis in a way. ▪ Not a bad idea at all. This could be a good discovery tool, especially if you have teams of curators working on a subject oriented project, e.g. 'U.S. natural disasters' ▪ It would avoid the need for duplicate captures. It could be used to give the capture creators a better understanding of how the data were being used. ▪ necessary for collaboration ▪ Would increase access to captured material. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Probably none. Depending on other limitations, might not allow a curator to use the site settings on a capture needed for their collection. ▪ Some curators might not want to share complete access (such as editing settings, etc.) to all other curators. May need to allow curators the option to share a capture or make it private or limit permissions ▪ May have copyright implications? ▪ Could be overwhelming. Would have to be a way to select sites of interest. |
| 25. Idea | Selecting "Help" on any screen opens the general help document at the relevant section for that screen and also allows easy navigation to the contents list of the entire help document. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Important, as the help menu will be regularly used, I believe ▪ No opinion on this. ▪ Less important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Should get the curator to the answer to their question sooner. ▪ Sure! The benefits for this seem obvious. ▪ It would streamline the process of navigating the help screens. ▪ very handy, and becoming common practice elsewhere |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Needs to still be easy to navigate to the main help screen and other help pages if the user is multitasking, like I often am. ▪ Seems less important than other enhancements listed. |
| 26. Idea | Explain error messages, such as why files were not captured (e.g., server restrictions or capture parameters). |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ High priority. There should also be an explanation of symbols that cause the system to break. Ex: ampersand (&). |

| | |
|--|--|
| | <ul style="list-style-type: none"> ▪ Essential for archival record/documentation ▪ Would be very useful in determining if a site capture was successful, or not useful. ▪ Very important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Curators can attempt a new capture with parameters that might fix the problem. ▪ Better information would let me know more quickly if the problem was something I could fix, or beyond my control. I recaptured sites and tied up system resources for problems that weren't related to anything a recapture would have solved. ▪ Yes! I had some jobs that were interrupted after a very long capture session and did not know why. ▪ This would help the person defining the capture to review the file exclusions. ▪ Allows curators to formulate alternate approaches to captures (for example, if site cannot be captured because "&" appears in the URL, curator can look for an alternate URL that serves the same/similar purpose). Helps curator decide whether or not it is worthwhile to re-try the capture at a later time ▪ again, essential for the archival record, and for planning technical improvements ▪ If this is possible it would greatly aid the non-specialist person trying to archive a website. This would rank very high on my list of enhancements. Not sure it is replicated in other tools so would benefit the entire open source web archiving community. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ The error may not be something that could be fixed, e.g. the WAS apparently can't capture data on a Google map. |
| 27. Idea | Give curators the option to 'override' robot exclusions if they have received permission from the web site owner. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ If curators have the site owners permission to capture then the capture should be allowed. This is another argument in favor of a checkbox or ability to add notes about the collection ▪ Yes – allows more complete capture ▪ Can be extremely important as sometimes one has connections that will allow this override. ▪ Very important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ This would be extremely useful for capturing from local government agencies online documents centers, which seem to use proprietary software designed for businesses. ▪ Would ensure everything is captured. ▪ Not desirable -- again, knowledge of excluded files is essential for the archival record ▪ Allows all sites to be captured. ▪ Again, the Web Archives Workbench includes this functionality. We've used it to ignore robot exclusions on government docs. |

| | |
|--|---|
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ The more I think about this, the more uneasy I am with it. That remains true after reading the comments from the 1st round. |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ This feature should probably only be used with permission of the site's owner. ▪ I'm not entirely comfortable about this. Shouldn't we respect a site owner's wishes. ▪ Sounds like this could be dangerous... ▪ This could put a load on sites that is not desired. ▪ not knowing what was excluded ▪ Ignores wishes of site designer. |
| 28. Idea | 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. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ This would be very helpful for serving the documents out to users. ▪ Good idea ▪ Obviously an excellent suggestion. ▪ Nice feature. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ This would be especially useful for things considered publications so I could refer patrons to them and/or link to them from the library catalog. ▪ Very good idea! This would enable easy identification of capture jobs. ▪ Terrific idea. Something like this will be necessary to allow us to integrate the captured files into our libraries' collections and to give our users easy access to the captured files. ▪ Would allow use of site without having to enter WAS; Allows researchers to accurately cite sources retrieved by the WAS. ▪ very useful ▪ Important feature. Will increase access. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ The archive is closed and dark, requiring a login. No need to point outside users to it until it is available (and searchable) to them. |
| 29. Idea | In WAS documentation identify screen resolution and browser settings for optimal display of the WAS user interface. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ Also essential for archival record ▪ Would be very useful. ▪ Not important. |
| Benefits: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ useful -- again, essential for reproducing the look and feel of the site |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |

| | |
|--|--|
| Disadvantages: 1 st Questionnaire | <ul style="list-style-type: none"> ▪ Seems less important than other enhancements listed. |
| 30. New Idea in Q2 | 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. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ May or may not be necessary once the archive becomes 'light.' ▪ This function along with those in #28 above would provide targeted access and use of captured material without having to have access to WAS. ▪ vital for [our collections] ▪ No opinion on this. ▪ Nice feature. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |
| 31. New Idea in Q2 | Functionality to manage several thousand sites for a single institution. |
| Benefits: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ It's what we're going to need since our collections are only going to grow. ▪ Ability to search the site is a key functional requirement. Without ability to search collections, our users will not take advantage of the collections. Important for the more active institutions. |
| Disadvantages: 2 nd Questionnaire | <ul style="list-style-type: none"> ▪ <i>None reported</i> |

Additional Comment:

- Questionnaire 2: It would be helpful to have a list of the top five to ten priorities for enhancements, and which will be most likely implemented. Including a timeline of releases where they will be rolled out.
- Questionnaire 3: #12 - ability to capture active content is the most important enhancement – [management] and collaboration are secondary to actually capturing sites in their totality, and to continually upgrading the WAS so that it can capture content/file types as they emerge

Appendix C: WAS Enhancement Ideas: Listed by Ranks with Data Tables

Note: Thirty-one ideas were rated by curators; however, when the ideas were ranked some ideas had identical scores. In these cases ideas were given the same ranks, which resulted in 23 ranks.

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|---|----------------|---|---|---------------|---|----|------------------------|---|----|----------------------|---|----|----------------|---|-----|---------------------|---|-----|
| 22 | 3.67 | 1 | <p>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.</p> <table border="1" data-bbox="1115 558 1602 769"> <thead> <tr> <th>Idea 22 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>1</td> <td>8%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Extremely Important</td> <td>9</td> <td>75%</td> </tr> </tbody> </table> | Idea 22 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 1 | 8% | Very Important | 2 | 17% | Extremely Important | 9 | 75% |
| Idea 22 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 1 | 8% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 9 | 75% | | | | | | | | | | | | | | | | | | | |
| 4 | 3.58 | 2 | <p>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.</p> <table border="1" data-bbox="1115 948 1602 1159"> <thead> <tr> <th>Idea 4 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>1</td> <td>8%</td> </tr> <tr> <td>Very Important</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Extremely Important</td> <td>8</td> <td>67%</td> </tr> </tbody> </table> | Idea 4 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 1 | 8% | Very Important | 3 | 25% | Extremely Important | 8 | 67% |
| Idea 4 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 1 | 8% | | | | | | | | | | | | | | | | | | | |
| Very Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 8 | 67% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|--|----------------|---|---|---------------|---|----|------------------------|---|----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 10 | 3.45 | 3 | <p>Schedule captures based on one or a combination of the following:</p> <ul style="list-style-type: none"> ▪ 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) <table border="1" data-bbox="1115 506 1602 717"> <thead> <tr> <th>Idea 10 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>2</td> <td>18%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>18%</td> </tr> <tr> <td>Extremely Important</td> <td>7</td> <td>64%</td> </tr> </tbody> </table> | Idea 10 (N=11) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 2 | 18% | Very Important | 2 | 18% | Extremely Important | 7 | 64% |
| Idea 10 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 7 | 64% | | | | | | | | | | | | | | | | | | | |
| 12 | 3.36 | 4 | <p>Develop the ability to capture sites with active content, for example, .PHP and .ASP files.</p> <table border="1" data-bbox="1115 831 1602 1042"> <thead> <tr> <th>Idea 12 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>1</td> <td>9%</td> </tr> <tr> <td>Very Important</td> <td>5</td> <td>45%</td> </tr> <tr> <td>Extremely Important</td> <td>5</td> <td>45%</td> </tr> </tbody> </table> | Idea 12 (N=11) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 1 | 9% | Very Important | 5 | 45% | Extremely Important | 5 | 45% |
| Idea 12 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 1 | 9% | | | | | | | | | | | | | | | | | | | |
| Very Important | 5 | 45% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 5 | 45% | | | | | | | | | | | | | | | | | | | |
| 1 | 3.33 | 5 | <p>Curator collaboration so that a specified group of curators, both from a single campus or multiple campuses, share authority and access to joint collections.</p> <table border="1" data-bbox="1115 1156 1602 1367"> <thead> <tr> <th>Idea 1 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Extremely Important</td> <td>7</td> <td>58%</td> </tr> </tbody> </table> | Idea 1 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 3 | 25% | Very Important | 2 | 17% | Extremely Important | 7 | 58% |
| Idea 1 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 7 | 58% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|---|----------------|---|---|---------------|---|----|------------------------|---|----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 20 | 3.25 | 6 | <p>Generate a report that compares captures so that files that were added or deleted can be readily identified.</p> <table border="1" data-bbox="1115 345 1602 557"> <thead> <tr> <th>Idea 20 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>1</td> <td>8%</td> </tr> <tr> <td>Very Important</td> <td>7</td> <td>58%</td> </tr> <tr> <td>Extremely Important</td> <td>4</td> <td>33%</td> </tr> </tbody> </table> | Idea 20 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 1 | 8% | Very Important | 7 | 58% | Extremely Important | 4 | 33% |
| Idea 20 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 1 | 8% | | | | | | | | | | | | | | | | | | | |
| Very Important | 7 | 58% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| 26 | 3.25 | 6 | <p>Explain error messages, such as why files were not captured (e.g., server restrictions or capture parameters).</p> <table border="1" data-bbox="1115 670 1602 881"> <thead> <tr> <th>Idea 26 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Very Important</td> <td>5</td> <td>42%</td> </tr> <tr> <td>Extremely Important</td> <td>5</td> <td>42%</td> </tr> </tbody> </table> | Idea 26 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 2 | 17% | Very Important | 5 | 42% | Extremely Important | 5 | 42% |
| Idea 26 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Very Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| 23 | 3.08 | 7 | <p>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.</p> <table border="1" data-bbox="1115 1060 1602 1271"> <thead> <tr> <th>Idea 23 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>4</td> <td>33%</td> </tr> <tr> <td>Very Important</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Extremely Important</td> <td>5</td> <td>42%</td> </tr> </tbody> </table> | Idea 23 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 4 | 33% | Very Important | 3 | 25% | Extremely Important | 5 | 42% |
| Idea 23 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| Very Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|---|----------------|---|---|---------------|---|----|------------------------|---|----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 24 | 3.08 | 7 | <p>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.</p> <table border="1" data-bbox="1115 410 1602 621"> <thead> <tr> <th>Idea 24 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>4</td> <td>33%</td> </tr> <tr> <td>Very Important</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Extremely Important</td> <td>5</td> <td>42%</td> </tr> </tbody> </table> | Idea 24 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 4 | 33% | Very Important | 3 | 25% | Extremely Important | 5 | 42% |
| Idea 24 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| Very Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| 6 | 3.00 | 8 | <p>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.</p> <table border="1" data-bbox="1115 833 1602 1044"> <thead> <tr> <th>Idea 6 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Very Important</td> <td>6</td> <td>50%</td> </tr> <tr> <td>Extremely Important</td> <td>3</td> <td>25%</td> </tr> </tbody> </table> | Idea 6 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 3 | 25% | Very Important | 6 | 50% | Extremely Important | 3 | 25% |
| Idea 6 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| Very Important | 6 | 50% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| 27 | 3.00 | 8 | <p>Give curators the option to 'override' robot exclusions if they have received permission from the web site owner.</p> <table border="1" data-bbox="1115 1157 1602 1369"> <thead> <tr> <th>Idea 27 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>3</td> <td>27%</td> </tr> <tr> <td>Very Important</td> <td>5</td> <td>45%</td> </tr> <tr> <td>Extremely Important</td> <td>3</td> <td>27%</td> </tr> </tbody> </table> | Idea 27 (N=11) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 3 | 27% | Very Important | 5 | 45% | Extremely Important | 3 | 27% |
| Idea 27 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |
| Very Important | 5 | 45% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|---|----------------|---|---|---------------|---|----|------------------------|---|-----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 28 | 3.00 | 8 | <p>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.</p> <table border="1" data-bbox="1115 378 1602 589"> <thead> <tr> <th>Idea 28 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Moderately Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Extremely Important</td> <td>6</td> <td>50%</td> </tr> </tbody> </table> | Idea 28 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 2 | 17% | Moderately Important | 2 | 17% | Very Important | 2 | 17% | Extremely Important | 6 | 50% |
| Idea 28 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 6 | 50% | | | | | | | | | | | | | | | | | | | |
| 30 | 3.00 | 8 | <p>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.</p> <table border="1" data-bbox="1115 735 1602 946"> <thead> <tr> <th>Idea 30 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>5</td> <td>42%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Extremely Important</td> <td>5</td> <td>42%</td> </tr> </tbody> </table> | Idea 30 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 5 | 42% | Very Important | 2 | 17% | Extremely Important | 5 | 42% |
| Idea 30 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| 11 | 2.91 | 9 | <p>Specify file type(s) (e.g., audio, video, and document) or file extensions (e.g., PDF, DOC, MP3, and AVI) to be included or excluded from a site capture.</p> <table border="1" data-bbox="1115 1060 1602 1271"> <thead> <tr> <th>Idea 11 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>4</td> <td>36%</td> </tr> <tr> <td>Very Important</td> <td>4</td> <td>36%</td> </tr> <tr> <td>Extremely Important</td> <td>3</td> <td>27%</td> </tr> </tbody> </table> | Idea 11 (N=11) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 4 | 36% | Very Important | 4 | 36% | Extremely Important | 3 | 27% |
| Idea 11 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 4 | 36% | | | | | | | | | | | | | | | | | | | |
| Very Important | 4 | 36% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|--|----------------|---|---|---------------|---|----|------------------------|---|-----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 3 | 2.83 | 10 | <p>For the sites they define and capture and the collections they build, allow curators to set access levels for other curators. Access levels might include full editing permission (i.e., full collaboration), search and display permission, or permission to include a site's capture(s) in another curator's collection.</p> <table border="1" data-bbox="1115 410 1602 621"> <thead> <tr> <th>Idea 3 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>5</td> <td>42%</td> </tr> <tr> <td>Very Important</td> <td>4</td> <td>33%</td> </tr> <tr> <td>Extremely Important</td> <td>3</td> <td>25%</td> </tr> </tbody> </table> | Idea 3 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 5 | 42% | Very Important | 4 | 33% | Extremely Important | 3 | 25% |
| Idea 3 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| Very Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| 13 | 2.83 | 10 | <p>Sort sites by collection to which they are assigned.</p> <table border="1" data-bbox="1115 703 1602 914"> <thead> <tr> <th>Idea 13 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Moderately Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Very Important</td> <td>4</td> <td>33%</td> </tr> <tr> <td>Extremely Important</td> <td>4</td> <td>33%</td> </tr> </tbody> </table> | Idea 13 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 2 | 17% | Moderately Important | 2 | 17% | Very Important | 4 | 33% | Extremely Important | 4 | 33% |
| Idea 13 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Very Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| 25 | 2.75 | 11 | <p>Selecting "Help" on any screen opens the general help document at the relevant section for that screen and also allows easy navigation to the contents list of the entire help document.</p> <table border="1" data-bbox="1115 1060 1602 1271"> <thead> <tr> <th>Idea 25 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>6</td> <td>50%</td> </tr> <tr> <td>Very Important</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Extremely Important</td> <td>3</td> <td>25%</td> </tr> </tbody> </table> | Idea 25 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 6 | 50% | Very Important | 3 | 25% | Extremely Important | 3 | 25% |
| Idea 25 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 6 | 50% | | | | | | | | | | | | | | | | | | | |
| Very Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 3 | 25% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|--|----------------|---|---|---------------|---|----|------------------------|---|-----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 5 | 2.73 | 12 | <p>Site parameters for capture scope to include "Directory +1 link", where a directory and only the linked content on pages within that directory are captured, and "Page +1 link", where only a specific page and the linked content on that page are captured.</p> <table border="1" data-bbox="1115 410 1602 621"> <thead> <tr> <th>Idea 5 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>1</td> <td>9%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>2</td> <td>18%</td> </tr> <tr> <td>Very Important</td> <td>6</td> <td>55%</td> </tr> <tr> <td>Extremely Important</td> <td>2</td> <td>18%</td> </tr> </tbody> </table> | Idea 5 (N=11) | # | % | Not Important | 1 | 9% | Very Little Importance | 0 | 0% | Moderately Important | 2 | 18% | Very Important | 6 | 55% | Extremely Important | 2 | 18% |
| Idea 5 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 1 | 9% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Very Important | 6 | 55% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| 31 | 2.73 | 12 | <p>Functionality to manage several thousand sites for a single institution.</p> <table border="1" data-bbox="1115 703 1602 914"> <thead> <tr> <th>Idea 31 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>2</td> <td>18%</td> </tr> <tr> <td>Moderately Important</td> <td>3</td> <td>27%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>18%</td> </tr> <tr> <td>Extremely Important</td> <td>4</td> <td>36%</td> </tr> </tbody> </table> | Idea 31 (N=11) | # | % | Not Important | 0 | 0% | Very Little Importance | 2 | 18% | Moderately Important | 3 | 27% | Very Important | 2 | 18% | Extremely Important | 4 | 36% |
| Idea 31 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 4 | 36% | | | | | | | | | | | | | | | | | | | |
| 2 | 2.70 | 13 | <p>An automated workflow that allows curators to identify sites for inclusion in collections and allows non-curatorial staff to create site descriptions, schedule captures, and evaluate capture results. This might be accomplished by associating different levels of authority with WAS user IDs based on users' job responsibilities.</p> <table border="1" data-bbox="1115 1125 1602 1336"> <thead> <tr> <th>Idea 2 (N=10)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>1</td> <td>10%</td> </tr> <tr> <td>Moderately Important</td> <td>4</td> <td>40%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>20%</td> </tr> <tr> <td>Extremely Important</td> <td>3</td> <td>30%</td> </tr> </tbody> </table> | Idea 2 (N=10) | # | % | Not Important | 0 | 0% | Very Little Importance | 1 | 10% | Moderately Important | 4 | 40% | Very Important | 2 | 20% | Extremely Important | 3 | 30% |
| Idea 2 (N=10) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 1 | 10% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 4 | 40% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 20% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 3 | 30% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|--|----------------|---|---|---------------|---|----|------------------------|---|----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 14 | 2.67 | 14 | <p>Provide easier method(s) of dealing with the volume of files in the files list by:</p> <ul style="list-style-type: none"> ▪ Increasing the file sorting options to include file name, server, and directory ▪ Displaying sites in a directory tree, from which clicking on a directory displays its files <table border="1" data-bbox="1115 443 1602 654"> <thead> <tr> <th>Idea 14 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>1</td> <td>8%</td> </tr> <tr> <td>Moderately Important</td> <td>4</td> <td>33%</td> </tr> <tr> <td>Very Important</td> <td>5</td> <td>42%</td> </tr> <tr> <td>Extremely Important</td> <td>2</td> <td>17%</td> </tr> </tbody> </table> | Idea 14 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 1 | 8% | Moderately Important | 4 | 33% | Very Important | 5 | 42% | Extremely Important | 2 | 17% |
| Idea 14 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 1 | 8% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| Very Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| 19 | 2.67 | 14 | <p>Allow simultaneous browsing of two or more captures for the same site, or of a live site and a capture, to enable comparisons and evaluate capture results.</p> <table border="1" data-bbox="1115 768 1602 979"> <thead> <tr> <th>Idea 19 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>5</td> <td>42%</td> </tr> <tr> <td>Very Important</td> <td>6</td> <td>50%</td> </tr> <tr> <td>Extremely Important</td> <td>1</td> <td>8%</td> </tr> </tbody> </table> | Idea 19 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 0 | 0% | Moderately Important | 5 | 42% | Very Important | 6 | 50% | Extremely Important | 1 | 8% |
| Idea 19 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 5 | 42% | | | | | | | | | | | | | | | | | | | |
| Very Important | 6 | 50% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 1 | 8% | | | | | | | | | | | | | | | | | | | |
| 8 | 2.64 | 15 | <p>In the description of a site, include the collection[s] to which it belongs. This should be automatically generated when sites are added to collections.</p> <table border="1" data-bbox="1115 1092 1602 1304"> <thead> <tr> <th>Idea 8 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>1</td> <td>9%</td> </tr> <tr> <td>Moderately Important</td> <td>5</td> <td>45%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>18%</td> </tr> <tr> <td>Extremely Important</td> <td>3</td> <td>27%</td> </tr> </tbody> </table> | Idea 8 (N=11) | # | % | Not Important | 0 | 0% | Very Little Importance | 1 | 9% | Moderately Important | 5 | 45% | Very Important | 2 | 18% | Extremely Important | 3 | 27% |
| Idea 8 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 1 | 9% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 5 | 45% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|---|----------------|---|---|---------------|---|-----|------------------------|---|-----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|-------------|---|----|
| 18 | 2.60 | 16 | <p data-bbox="806 250 1902 376">Enable searching of the entire archive using curator name and subject. 'Subject' should be defined. For example, does it mean any subject keyword or the specific subject terms recorded in site metadata, whether or not a controlled vocabulary is used?</p> <table border="1" data-bbox="1115 409 1602 620"> <thead> <tr> <th>Idea 18 (N=10)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>1</td> <td>10%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>3</td> <td>30%</td> </tr> <tr> <td>Very Important</td> <td>4</td> <td>40%</td> </tr> <tr> <td>Extremely Important</td> <td>2</td> <td>20%</td> </tr> </tbody> </table> | Idea 18 (N=10) | # | % | Not Important | 1 | 10% | Very Little Importance | 0 | 0% | Moderately Important | 3 | 30% | Very Important | 4 | 40% | Extremely Important | 2 | 20% | | | |
| Idea 18 (N=10) | # | % | | | | | | | | | | | | | | | | | | | | | | |
| Not Important | 1 | 10% | | | | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 30% | | | | | | | | | | | | | | | | | | | | | | |
| Very Important | 4 | 40% | | | | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 2 | 20% | | | | | | | | | | | | | | | | | | | | | | |
| 9 | 2.50 | 17 | <p data-bbox="806 639 1860 737">Specify the addresses or domains to be included in a site capture (e.g., a host address or the .gov domain) and, conversely, specify domains to be excluded from a site capture (e.g., com or .org.).</p> <table border="1" data-bbox="1115 769 1602 1013"> <thead> <tr> <th>Idea 9 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Moderately Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Very Important</td> <td>5</td> <td>42%</td> </tr> <tr> <td>Extremely Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>No Response</td> <td>0</td> <td>0%</td> </tr> </tbody> </table> | Idea 9 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 3 | 25% | Moderately Important | 2 | 17% | Very Important | 5 | 42% | Extremely Important | 2 | 17% | No Response | 0 | 0% |
| Idea 9 (N=12) | # | % | | | | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 3 | 25% | | | | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 2 | 17% | | | | | | | | | | | | | | | | | | | | | | |
| Very Important | 5 | 42% | | | | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 2 | 17% | | | | | | | | | | | | | | | | | | | | | | |
| No Response | 0 | 0% | | | | | | | | | | | | | | | | | | | | | | |
| 7 | 2.42 | 18 | <p data-bbox="806 1029 1877 1094">Include a controlled vocabulary for subject headings and the ability to modify it for specific collections or topical areas.</p> <table border="1" data-bbox="1115 1127 1602 1338"> <thead> <tr> <th>Idea 7 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Moderately Important</td> <td>3</td> <td>25%</td> </tr> <tr> <td>Very Important</td> <td>4</td> <td>33%</td> </tr> <tr> <td>Extremely Important</td> <td>2</td> <td>17%</td> </tr> </tbody> </table> | Idea 7 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 3 | 25% | Moderately Important | 3 | 25% | Very Important | 4 | 33% | Extremely Important | 2 | 17% | | | |
| Idea 7 (N=12) | # | % | | | | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 3 | 25% | | | | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 25% | | | | | | | | | | | | | | | | | | | | | | |
| Very Important | 4 | 33% | | | | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 2 | 17% | | | | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|---|----------------|---|---|---------------|---|----|------------------------|---|-----|----------------------|---|-----|----------------|---|-----|---------------------|---|-----|
| 29 | 2.33 | 19 | <p>In WAS documentation identify screen resolution and browser settings for optimal display of the WAS user interface.</p> <table border="1" data-bbox="1115 345 1602 557"> <thead> <tr> <th>Idea 29 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>1</td> <td>8%</td> </tr> <tr> <td>Very Little Importance</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Moderately Important</td> <td>7</td> <td>58%</td> </tr> <tr> <td>Very Important</td> <td>2</td> <td>17%</td> </tr> <tr> <td>Extremely Important</td> <td>2</td> <td>17%</td> </tr> </tbody> </table> | Idea 29 (N=12) | # | % | Not Important | 1 | 8% | Very Little Importance | 0 | 0% | Moderately Important | 7 | 58% | Very Important | 2 | 17% | Extremely Important | 2 | 17% |
| Idea 29 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 1 | 8% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 7 | 58% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 2 | 17% | | | | | | | | | | | | | | | | | | | |
| 21 | 2.25 | 20 | <p>When viewing PDF files with active links, display the URL for the link along with the linked file.</p> <table border="1" data-bbox="1115 670 1602 881"> <thead> <tr> <th>Idea 21 (N=12)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>1</td> <td>8%</td> </tr> <tr> <td>Moderately Important</td> <td>7</td> <td>58%</td> </tr> <tr> <td>Very Important</td> <td>4</td> <td>33%</td> </tr> <tr> <td>Extremely Important</td> <td>0</td> <td>0%</td> </tr> </tbody> </table> | Idea 21 (N=12) | # | % | Not Important | 0 | 0% | Very Little Importance | 1 | 8% | Moderately Important | 7 | 58% | Very Important | 4 | 33% | Extremely Important | 0 | 0% |
| Idea 21 (N=12) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 1 | 8% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 7 | 58% | | | | | | | | | | | | | | | | | | | |
| Very Important | 4 | 33% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| 15 | 2.09 | 21 | <p>View the structure of captured sites as a tree structure that includes filenames. If this is not feasible, show a site's structure as a tree that includes directories and the number and size of files within each directory by file type (e.g., /directory_name: 5 html files, 50kb; 2 png files, 60kb).</p> <table border="1" data-bbox="1115 1060 1602 1271"> <thead> <tr> <th>Idea 15 (N=11)</th> <th>#</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>Not Important</td> <td>0</td> <td>0%</td> </tr> <tr> <td>Very Little Importance</td> <td>4</td> <td>36%</td> </tr> <tr> <td>Moderately Important</td> <td>3</td> <td>27%</td> </tr> <tr> <td>Very Important</td> <td>3</td> <td>27%</td> </tr> <tr> <td>Extremely Important</td> <td>1</td> <td>9%</td> </tr> </tbody> </table> | Idea 15 (N=11) | # | % | Not Important | 0 | 0% | Very Little Importance | 4 | 36% | Moderately Important | 3 | 27% | Very Important | 3 | 27% | Extremely Important | 1 | 9% |
| Idea 15 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 4 | 36% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |
| Very Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 1 | 9% | | | | | | | | | | | | | | | | | | | |

| Idea # | Average | Rank | Idea | | | | | | | | | | | | | | | | | | |
|------------------------|---------|------|--|----------------|---|---|---------------|---|----|------------------------|---|-----|----------------------|---|-----|----------------|---|-----|---------------------|---|----|
| 17 | 2.00 | 22 | <p data-bbox="806 250 1554 282">Provide the ability to print the list of files in its entirety.</p> <table border="1" data-bbox="1115 315 1602 526"> <thead> <tr> <th data-bbox="1121 315 1415 347">Idea 17 (N=11)</th> <th data-bbox="1415 315 1486 347">#</th> <th data-bbox="1486 315 1596 347">%</th> </tr> </thead> <tbody> <tr> <td data-bbox="1121 347 1415 380">Not Important</td> <td data-bbox="1415 347 1486 380">1</td> <td data-bbox="1486 347 1596 380">9%</td> </tr> <tr> <td data-bbox="1121 380 1415 412">Very Little Importance</td> <td data-bbox="1415 380 1486 412">2</td> <td data-bbox="1486 380 1596 412">18%</td> </tr> <tr> <td data-bbox="1121 412 1415 444">Moderately Important</td> <td data-bbox="1415 412 1486 444">5</td> <td data-bbox="1486 412 1596 444">45%</td> </tr> <tr> <td data-bbox="1121 444 1415 477">Very Important</td> <td data-bbox="1415 444 1486 477">2</td> <td data-bbox="1486 444 1596 477">18%</td> </tr> <tr> <td data-bbox="1121 477 1415 509">Extremely Important</td> <td data-bbox="1415 477 1486 509">1</td> <td data-bbox="1486 477 1596 509">9%</td> </tr> </tbody> </table> | Idea 17 (N=11) | # | % | Not Important | 1 | 9% | Very Little Importance | 2 | 18% | Moderately Important | 5 | 45% | Very Important | 2 | 18% | Extremely Important | 1 | 9% |
| Idea 17 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 1 | 9% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 5 | 45% | | | | | | | | | | | | | | | | | | | |
| Very Important | 2 | 18% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 1 | 9% | | | | | | | | | | | | | | | | | | | |
| 16 | 1.91 | 23 | <p data-bbox="806 542 1885 639">Include the option to view thumbnails of captured sites' home pages in both the 'Manage Sites' and 'View Captures' areas so that sites can be more readily identified.</p> <table border="1" data-bbox="1115 672 1602 883"> <thead> <tr> <th data-bbox="1121 672 1415 704">Idea 16 (N=11)</th> <th data-bbox="1415 672 1486 704">#</th> <th data-bbox="1486 672 1596 704">%</th> </tr> </thead> <tbody> <tr> <td data-bbox="1121 704 1415 737">Not Important</td> <td data-bbox="1415 704 1486 737">1</td> <td data-bbox="1486 704 1596 737">9%</td> </tr> <tr> <td data-bbox="1121 737 1415 769">Very Little Importance</td> <td data-bbox="1415 737 1486 769">3</td> <td data-bbox="1486 737 1596 769">27%</td> </tr> <tr> <td data-bbox="1121 769 1415 802">Moderately Important</td> <td data-bbox="1415 769 1486 802">3</td> <td data-bbox="1486 769 1596 802">27%</td> </tr> <tr> <td data-bbox="1121 802 1415 834">Very Important</td> <td data-bbox="1415 802 1486 834">4</td> <td data-bbox="1486 802 1596 834">36%</td> </tr> <tr> <td data-bbox="1121 834 1415 867">Extremely Important</td> <td data-bbox="1415 834 1486 867">0</td> <td data-bbox="1486 834 1596 867">0%</td> </tr> </tbody> </table> | Idea 16 (N=11) | # | % | Not Important | 1 | 9% | Very Little Importance | 3 | 27% | Moderately Important | 3 | 27% | Very Important | 4 | 36% | Extremely Important | 0 | 0% |
| Idea 16 (N=11) | # | % | | | | | | | | | | | | | | | | | | | |
| Not Important | 1 | 9% | | | | | | | | | | | | | | | | | | | |
| Very Little Importance | 3 | 27% | | | | | | | | | | | | | | | | | | | |
| Moderately Important | 3 | 27% | | | | | | | | | | | | | | | | | | | |
| Very Important | 4 | 36% | | | | | | | | | | | | | | | | | | | |
| Extremely Important | 0 | 0% | | | | | | | | | | | | | | | | | | | |

Appendix D. WAS Enhancement Ideas: Listed by Idea Number

| Idea # | Average | Average Rank | Idea |
|--------|---------|--------------|--|
| 1 | 3.33 | 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. |
| 2 | 2.70 | 13 | An automated workflow that allows curators to identify sites for inclusion in collections and allows non-curatorial staff to create site descriptions, schedule captures, and evaluate capture results. This might be accomplished by associating different levels of authority with WAS user IDs based on users' job responsibilities. |
| 3 | 2.83 | 10 | For the sites they define and capture and the collections they build, allow curators to set access levels for other curators. Access levels might include full editing permission (i.e., full collaboration), search and display permission, or permission to include a site's capture(s) in another curator's collection. |
| 4 | 3.58 | 2 | 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. |
| 5 | 2.73 | 12 | Site parameters for capture scope to include "Directory +1 link", where a directory and only the linked content on pages within that directory are captured, and "Page +1 link", where only a specific page and the linked content on that page are captured. |
| 6 | 3.00 | 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. |
| 7 | 2.42 | 18 | Include a controlled vocabulary for subject headings and the ability to modify it for specific collections or topical areas. |
| 8 | 2.64 | 15 | In the description of a site, include the collection[s] to which it belongs. This should be automatically generated when sites are added to collections. |

| Idea # | Average | Average Rank | Idea |
|--------|---------|--------------|---|
| 9 | 2.50 | 17 | Specify the addresses or domains to be included in a site capture (e.g., a host address or the .gov domain) and, conversely, specify domains to be excluded from a site capture (e.g., com or .org.). |
| 10 | 3.45 | 3 | Schedule captures based on one or a combination of the following: <ul style="list-style-type: none"> ▪ 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) |
| 11 | 2.91 | 9 | Specify file type(s) (e.g., audio, video, and document) or file extensions (e.g., PDF, DOC, MP3, and AVI) to be included or excluded from a site capture. |
| 12 | 3.36 | 4 | Develop the ability to capture sites with active content, for example, PHP and .ASP files. |
| 13 | 2.83 | 10 | Sort sites by collection to which they are assigned. |
| 14 | 2.67 | 14 | Provide easier method(s) of dealing with the volume of files in the files list by: <ul style="list-style-type: none"> ▪ Increasing the file sorting options to include file name, server, and directory ▪ Displaying sites in a directory tree, from which clicking on a directory displays its files |
| 15 | 2.09 | 21 | View the structure of captured sites as a tree structure that includes filenames. If this is not feasible, show a site's structure as a tree that includes directories and the number and size of files within each directory by file type (e.g., /directory_name: 5 html files, 50kb; 2 png files, 60kb). |
| 16 | 1.91 | 23 | Include the option to view thumbnails of captured sites' home pages in both the 'Manage Sites' and 'View Captures' areas so that sites can be more readily identified. |
| 17 | 2.00 | 22 | Provide the ability to print the list of files in its entirety. |

| Idea # | Average | Average Rank | Idea |
|--------|---------|--------------|---|
| 18 | 2.60 | 16 | Enable searching of the entire archive using curator name and subject. 'Subject' should be defined. For example, does it mean any subject keyword or the specific subject terms recorded in site metadata, whether or not a controlled vocabulary is used? |
| 19 | 2.67 | 14 | Allow simultaneous browsing of two or more captures for the same site, or of a live site and a capture, to enable comparisons and evaluate capture results. |
| 20 | 3.25 | 6 | Generate a report that compares captures so that files that were added or deleted can be readily identified. |
| 21 | 2.25 | 20 | When viewing PDF files with active links, display the URL for the link along with the linked file. |
| 22 | 3.67 | 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. |
| 23 | 3.08 | 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. |
| 24 | 3.08 | 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. |
| 25 | 2.75 | 11 | Selecting "Help" on any screen opens the general help document at the relevant section for that screen and also allows easy navigation to the contents list of the entire help document. |
| 26 | 3.25 | 6 | Explain error messages, such as why files were not captured (e.g., server restrictions or capture parameters). |
| 27 | 3.00 | 8 | Give curators the option to 'override' robot exclusions if they have received permission from the web site owner. |

Prioritization of WAS Enhancement Ideas

| Idea # | Average | Average Rank | Idea |
|---------------|----------------|---------------------|--|
| 28 | 3.00 | 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. |
| 29 | 2.33 | 19 | In WAS documentation identify screen resolution and browser settings for optimal display of the WAS user interface. |
| 30 | 3.00 | 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. |
| 31 | 2.73 | 12 | Functionality to manage several thousand sites for a single institution. |