WAS Release 4 - Usability Test Results

June 26, 2007

Prepared by:
Kathleen R. Murray
krmurray@unt.edu
University of North Texas

Contents

Introduction ................................................................................................ 2
Scenario 1. Create / Capture a Site.............................................................. 6
Scenario 2. Create a Collection.................................................................12
Scenario 3. Manage Collections & Capture Results...............................21
Scenario 4. View and Add Metadata.......................................................26
Additional Comments............................................................................29
Test Script .............................................................................................30
Introduction

Methodology
Usability testing of the Web Archiving Service (WAS) Release 4 was conducted on May 29, 2007. Eight of the project’s curators participated in test sessions, which generally were completed in 45-60 minutes. Two testing teams, each consisting of a facilitator and two observers were involved in the testing.

The Firefox web browser was used in all sessions. Camtasia Studio software captured each curator’s interactions with the WAS application as well as the discussion that occurred during the session. (However, in 3 full sessions and one partial session, the discussion was not captured.)

Each curator was assigned a unique test account that allowed them to access a controlled environment in which the following sites had been captured:

- California Department of Water Resources (1 capture jobs)
- California State Water Resources Control Board (1 capture jobs)
- Bureau of Reclamation Mid-Pacific Region (1 capture jobs)
- Crawler Test (3 capture jobs)

All curators were given the test script in Appendix 1. The script consisted of four scenarios related to four functional areas of the WAS application.

1. Create/Capture a Site
2. Create a Collection
3. Manage Collections and Capture Results
4. View and Add Metadata

Facilitators instructed curators to read aloud each scenario and attempt to complete the specified tasks. Curators were encouraged to explore the application interface and talk aloud as they interacted with the WAS. Three of the eight curators worked through the entire script and four completed the first three scenarios. One curator completed only the first two scenarios due to the need to constrain their test session to 30 minutes. Subsequent to completing the testing, notes and recordings were analyzed by the project’s assessment analyst.

Objectives & Outcomes

Scenario 1 Objectives

1. Do curators know how to choose appropriate settings for crawling their site:
   a. Scope?
   b. Max Time?
2. How likely is a curator to use each of the descriptive metadata fields when defining/adding a site?
3. Do curators understand what selecting the “get subdomains” checkbox will do?

Scenario 1 Outcomes

1. In general curators are uncertain how to determine the appropriate settings for crawling sites. They rely heavily on past experience for both subsequent crawls of
the same site and for initial crawls. Sometimes past experience is a good teacher, however, in many cases curators would benefit from tools or guidance on how to evaluate sites to determine initial settings and how to interpret reports to modify settings for subsequent captures.

2. Curators appear quite likely to use all of the descriptive metadata fields in defining their sites. Curators have different understandings of the “curator” field for site definitions and for the “description” field for both sites and collections.

3. Some curators understand the concept of subdomains and several found the sidebar help text quite clear. Many did not appear to have an accurate understanding of the concept.

**Scenario 2 Objectives**

1. How easily can the user determine how to accomplish the following?
   a. create a new collection
   b. add an entire job to a collection
   c. add individual files to a collection
2. Do curators know where to go to add whole capture jobs to a collection?
3. Do curators know how to add multiple jobs of the same site to a collection or do they add jobs one at a time?
4. Which tabs (on Results screen) do curators think will be the most useful when building their collections?

**Scenario 2 Outcomes**

1. Curators readily created a new collection. All curators had initial difficulty discovering how to add an entire job to a collection. Almost all tried to add a job through Sites \(\rightarrow\) Manage Sites and were unable to determine how to accomplish this. Others tried to add jobs from the View Results screen, but were also unable to determine how to accomplish the task. Several eventually found Collections \(\rightarrow\) Build Collections and were able to easily add entire jobs and specific files from this screen.
2. See question 1.
3. For the collection curators built, there was only one capture job for each site available.
4. Until they were familiar with a site, curators indicated they would use the Files tab in their capture results to identify and add content to a collection.

**Scenario 3 Objectives**

1. How easily can the curator understand the contents of a collection?
2. Can the user easily distinguish if s/he is interacting with an object in a Capture Result or a Collection?
3. Can the user:
   a. Easily select a job to display its capture results?
   b. Understand the search results?

**Scenario 3 Outcomes**

1. There was a good deal of variance in curators’ understanding of the contents of a collection. All could recognize the site names in the Contents, however, some easily expanded the content tree and some struggled with it. Many had some difficulty with the concept of “job”, with one curator expecting there would always only be one job for any site in a collection. Likewise, locating specific files curators added to their
collection was a serendipitous event for many. That said, it does appear that understanding the contents of a collection using a content tree would be easily learned by curators.

2. Eventually, most curators did seem to understand that captured objects (i.e., sites and files) comprised collections. Many further understood that specific captures of sites could be members of collections. However, these understandings were not initially present for some curators. Additionally, conceptual clarity often did not translate to navigational and interactive clarity as curators became lost among the concepts of collections, sites, and jobs and the actions of creating, building, and managing.

3. Curators were primarily dealing with sites that only had one capture job and tended to think in terms of “sites” versus “jobs”. All were able to view lists of files and preferred this to searching the contents of captures, largely because they were not familiar enough with the content to search it. (Note: Curators that completed Scenario 4 readily understood their search results.)

Scenario 4 Objectives

1. How might a curator envision using the "comments" metadata option for a file?
2. How valuable do you think the site metadata will be to you?
3. Looking ahead to end user access to collections in the archive, how important do you anticipate the site-level metadata will be?

Scenario 4 Outcomes

Note: Only three curators completed this scenario resulting in limited outcomes.

1. No data was collected for this question.
2. One curator confused site-level metadata with site capture reports and thought the reports would be valuable in “troubleshooting” capture results to improve capture settings.
3. One curator commented that the capture date would be quite important for researchers. A few curators discussed end-user search and discovery of collections, versus sites. For example, one envisioned end users searching across all collections within the repository and thought collection-level descriptions should be created with that in mind.

General Findings

- The WAS has come a long way in the last year and is going in the right direction.
  - The WAS interface design is clean and is easy for curators to explore.
  - The WAS functionality relates well to the activities involved in web collection development.
- Terminology in the interface is confusing to curators who are often not clear on the distinctions between the following terms used in various places on the WAS screens and consequently are not sure what action to take.
  - Site, Collection, Job, Capture, File, and Object
  - Manage and View
  - Create, Build, and Add
- The interface favors people who tend to read the explanatory text on screens and contained within sidebars prior to taking actions. Some of the curators did this almost consistently and some almost never did this. The latter group met more often
with frustration, got lost navigationally, and made incorrect interpretations of actions and concepts.

- Likewise, some curators readily discover the icons on screens and explore both the mouse-over text descriptions and/or the Icon Keys in the sidebar. However, other curators find the action icons rather small and they are not drawn to either discover their meaning or to select them.
- Many curators, from time-to-time, get lost in the application and are unsure either of how they navigated to a location or of how to navigate to the feature or functionality they need. This was most common when curators attempted to add content to a new collection.

**General Recommendations**

- **Terminology**
  - Identify a core set of terms for key actions and concepts within the WAS and use them consistently.
  - Add an indexed and searchable glossary that includes definitions of common terms, concepts, and icons.

- **Navigation**
  - Create an obvious and optimal method curators will use for each major activity in the web collection development workflow.
  - For each major activity ensure that curators can make a single corresponding selection from the main drop-down navigation bar.
  - Simplify the number of avenues available for task completion. Merge related features/content, for example, perhaps the Capture History Tab and View Results screen can be merged and made accessible via Sites ➔ Capture Results.
  - Consider adding breadcrumbs to each screen to indicate where a user has navigated to within the application.
  - Ensure that the browser’s back button works appropriately.

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

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

The bulk of the remainder of this report consists of the detailed usability test findings and includes recommendations for refinement of the WAS. Additionally, some observations and recommendations from the project curators are included.
Scenario 1. Create / Capture a Site

Scenario

Scenario 1: In your role as a librarian, you need to create and define a site you want to capture. But first you have to edit the descriptive metadata for a site you previously defined and also recapture the site.

Task 1: Edit the descriptive data for the 'Crawler Test' site, which contains files used to measure the effect of different heritrix crawl settings on what is captured: (a) edit the description and (b) add a subject.

Task 2: Initiate a capture of the 'Crawler Test' site.

Task 3: Create and define a new site of your choice.

Findings

- Most curators readily edited the descriptive data for an existing site, either by selecting the Manage Sites screen image on the initial WAS screen or by selecting this option from the Sites drop-down menu.
- One curator initially wanted to click on the site name to edit the site, but easily discovered the edit icon.
- One curator was able to edit the descriptive data for the Crawler Test site only after being prompted to explore the icons on the Manage Sites screen, after which they selected the edit site information icon. For this curator the icons on the Manage Sites screen were not intuitive and the curator did not refer to the Icon Key in the sidebar. In contrast, a few curators thought the icons were very intuitive.

1 Several screen images from the WAS are included in the report. The findings related to a screen image are listed beneath the image.
Most curators readily edited and created sites either by accessing the ‘Manage Sites’ screen image on the initial WAS screen or by selecting this option from the Sites drop-down menu.

- Seed URLs
  - It is not clear if seed URLs need to include the leading “http://”. The examples in the Tips sidebar do not include this.
  - Some curators open a new browser tab and go to the live site they want to capture. They copy the URL from the site and see this as a way of avoiding a typing error. When this is done, the “http://” text is always included in the URL.

- Get Subdomains
  - Some curators correctly understand the concept of subdomains. Others have a range of understanding that informs their selection of the ‘get subdomains’ checkbox. Among these are:
    - All web pages whose URLs begin with the seed URL (i.e., all pages within a site)
    - Sites that are related in some manner
    - Domains for a site
    - No clear idea

- Scope
  - Some curators expressed a hesitancy to select the “Host + linked pages” setting. From experience these curators remembered that a very large number of files are captured with this setting and many of the files are irrelevant.
  - However, one curator selected “Host + linked pages” “just to be sure” and another curator remembered that ”Host + linked pages” was the best setting to capture their sites.
  - The difference between “Host” and “Host + linked pages” is not intuitive to all curators. If a curator neither refers to nor fully comprehends the definitions, then, for example, “linked pages” might mean either or both linked pages and files within the seed URL host site as well as linked pages from other sites.
- **Save (all tabs) → URL Format is Invalid Error Message**
  - This error message is generated when the WAS cannot validate the format of a URL. When some curators neglected to include the leading “http://”, they received an error message indicating they needed to include it.
  - One curator found the requirement to enter the leading protocol information to be an annoyance as well as inconsistent with current browser requirements. However, this curator found the feedback about how to fix the problem clear and easily corrected the URL.
  - The screen containing the error message appears to deselect the ‘get subdomains’ checkbox if the curator had selected it. The curator may or may not notice this, which could result in their desired option to include subdomains being omitted when they re-save their site.

- **Save (all tabs) → URL Could Not be Validated Error Message**
The error message is quite long to cover all possibilities, including that the URL may be valid.

One curator got this error message after cutting and pasting the URL from the live web site to the Seed URL field and could not understand why the WAS could not validate it on the live Internet. After looking at the Examples in the Tips sidebar, the curator removed the leading http:// and the ending backslash. This resulted in the WAS formatting error message for the URL. After adding the http:// (but not the ending backslash), the site was saved.

Some curators commented that descriptive data would likely reflect local cataloging practices, both in terms of controlled vocabularies and degree of detail.

**Description**

- One curator wondered what level of detail would be expected or desirable for site description? Suggestions details included:
  - Why the curator thought a site was important to capture
  - For an organizational site, what the organization was about

- One curator saw the description as an important historical record for future curators that documented the thinking of a curator at the point in time that a site was captured.
- Another curator routinely views a site’s source code to see if there is metadata they can copy and add to the Site Description field.

**Creator**

- Several curators assumed ‘creator’ referred to the curator who was adding the new site to the WAS as opposed to the creator of the site itself.
- In contrast, a few thought it referred to the site creator.

**Subjects**

- Some curators stated they would use controlled vocabularies, for example, LC subject headings.
- One curator copies subject headings (LCSH) from their existing catalog records if available.
- It was not clear to some curators how to separate multiple subject terms (e.g. using commas or semi-colons.)
Usability Test Results for WAS R4

- **Coverage**
  - Coverage was meaningful to many but not to all.
  - One curator appreciated the modifier of “geographic” appended to Coverage, commenting that coverage could mean time.
  - One curator stated they would use a controlled vocabulary (e.g., LCSH) for coverage.
  - One curator assumed that adding geographic coverage data would enable sorting captured sites by the data values in this field.

  ![Web Archiving Service screenshot]

- One curator interpreted the Max Time value (1 hour) to be the time required to capture the site and did not seem to know that this setting was one of several that a curator could select.

- One curator would like to have the option to schedule capture jobs for particular times to accommodate time zone differences, particularly when sites are on servers in other parts of the world.

  ![Web Archiving Service screenshot]

- From the Site Summary screen, one curator could not determine how to edit the descriptive data for a site they had created and for which they entered no descriptive data prior to saving the site. This curator simply never discovered the edit icon on the Descriptive Data tab. Subsequently, this curator was able to edit the site’s descriptive data for their site by selecting the edit site information icon from the Manage Sites screen.

  ![Web Archiving Service screenshot]
Recommendations

- In the Tips sidebar, clarify the need to include http://. If possible accommodate curators who omit this.
- Allow curators to cut-and-paste seed URLs from live web sites. If there are known problems with this (for example an ending backslash) include this in the sidebar information.
- Consider capturing subdomains for all sites as the default.
- Clarify the semantics for the creator field.
- State how multiple subject terms/phrases should be delineated in the subject field.
- Add the capability to schedule capture jobs at future times (versus Capture Now.
- For some curators it might be helpful to add a text label to the icons.
Scenario 2. Create a Collection

Scenario

Scenario 2: You are a librarian interested in California water issues, in particular about web-based materials related to droughts in the state. You have already captured a few key web sites:

- California Department of Water Resources
- California State Water Resources Control Board
- Bureau of Reclamation Mid-Pacific Region

Now you are going to create a new collection using materials from these sites.

Task 1 - Create a new collection.

Task 2 - Add the capture results for one of your three key web sites to the collection.

Task 3 - Add the capture results for another of your three key web sites to the collection.

Task 4 - Add a single file from the capture results of your third key web site to the collection.

Findings

- The conceptual model for the content of collections includes:
  1. Entire sites
     - Multiple captures of the same site
     - Single captures of a site
  2. Discrete files
     - Selected from captured sites
     - Captured as discrete files
     - Multiple captures or single captures of the same file
  3. Both sites and files

- This model is fairly clear to curators and seems to meet their needs. However, this understanding does not transfer easily to the navigational options and terms in the WAS interface.
- In general, curators found it easy to add a site and to add a collection. The confusion emerged when curators attempted to add the capture results for a site to a collection. (Does one do this from drop-down options on the Sites menu or on the Collections menu?)
- The ‘Tips: Collections’ text on the ‘Add Collection’ screen provides one illustration of the terminology that might cause confusion:
  o “After you add a collection, you will be able to build your collection from captured sites and files. You will build collections by adding entire capture jobs or individual search results to it.”
  - Note the terms ‘captured sites’ and ‘files’ in the first sentence and the terms ‘capture jobs’ and ‘individual search results’ in the second sentence.
For some curators, captured sites and capture jobs were equivalent concepts. The understanding that specific capture jobs for a site are a fundamental building block for collections appears to not be clearly understood.

Building a collection that includes files, the other fundamental building block of collections, requires a clear understanding of the hierarchical relationship between files in the individual search results of capture jobs for sites.

There are several methods/paths to add content to a collection (e.g., Site Summary: Job History tab; Build Collections; View Results: Results button) and this may contribute to navigational confusion.

Most curators were able to discover and successfully create a new collection using the ‘Add Collection’ option from the Collections drop-down menu.

One curator read the sidebar information on the Add Collection screen and erroneously decided that creating a new collection should be done from Build Collection menu option, not from the Add Collection screen they initially selected.
Curators had different interpretations of who the audience for the Collection Description was. Some would use it to inform other curators about the collection, others would target the information for end users of the collection, and one thought it best to address the information needs of both other curators and end users in the descriptive content.

One curator wondered if a spell checking tool was available.

The confirmation screen returned after a curator successfully added a collection directed them to “Browse your sights to find jobs or files to add to this collection.” Selecting this option led to the “Manage Sites” screen being presented. However, from the Manage Sites screen, most curators could not determine how to add a site to their collection.

The sidebar content is not pertinent to this screen.
Many curators attempted at first to add content to their collections by selecting Sites → Manage Sites. As previously stated, most curators could not determine how to add a site to their collection from this screen.

- A typical curator reaction after reviewing the available actions was: “I don’t want any of that, hmmm?”
- One curator selected the deactivate site icon.
- Another curator wondered if selecting the Capture button would move the site into the collection.

Terminology confusion

- Icon Key: text in side bar for magnifying glass is “view capture history”
-Hover text for magnifying glass icon is “view job history”
- The confirmation screen returned after adding a collection states: “Browse your sites to find jobs or files to add to this collection.” Selecting this returns the “Manage Sites” screen, which does not offer an obvious “add content” ability.
- Some curators wanted to “see their capture results” to select content for their collection.

When desired actions/functions were not found, curators employed a variety of discovery methods:

- Exploring screens from other drop-down menus/options
- Reading screen text
- Reading Tips
- Hovering over icons

On the Manage Sites screen, the terms used to describe the most recent capture job status are explained in the tips sidebar. However, the concepts of “active” and “inactive” apply to all capture jobs, not the most recent capture job, and the distinction between these terms in the drop-down menu and the status terms is confusing. One curator found a job status of “Finished” inconsistent with the meaning implied by a limit setting of “all active sites”.

Curators did not generally discover and explore the Job History tab, either from the Site Summary or via the Manage Sites screen. One curator stated they did not know exactly what Job History meant.

One curator who did navigate to the Site Summary: Job History tab from the Manage Sites page decided this screen was not appropriate for the task of adding content.
Conceptually this curator had wandered “off task” from the test script, but was “on task” for the focus of their planned collection, which will include only selected files from capture jobs and not entire jobs. This curator concluded that individual files could not be selected and added to a collection from the Job History tab. Subsequently, the curator reassessed the scripted task and discovered that she could complete the assigned task of adding a job from this tab.

- From the View Results screen, curators could not discover how to add a job to their collection.
- If curators selected a specific site name, they went to the Site Summary screen, which provided no clear indication of how to add the site to the collection.
- It was not always clear to curators that Jobs were related to Available Results.
- When a curator did select the drop-down date options for Available Results for the Crawler Test site:
  - It was not clear if any of the dates had already been added to the collection
  - However, it was clear that the dates referred to multiple captures of the site
- If the workflow logically and actually proceeds from adding sites, to capturing sites, to evaluating site results, to building collections, then perhaps when curators are building collections they have already evaluated their capture results and will not logically be looking at sites and capture results as they build their collections.
Many curators did discover the Build Collections screen. Some found it from the Manage Collections screen and others from the Collections drop-down menu.

It was readily obvious to most curators how to add jobs/sites from the Build Collections screen.

The heading ‘Jobs’ lists captured sites. This is an example of the jobs versus sites terminology confusion.

There was no indication that a job/site had already been added to a collection(s).

Most curators readily understood and interacted with the [+-] icons to expand and contract a site, which was necessary to view the check box and add a job to a collection. However, some never experimented with this. It seems these curators were attempting to sites to their collection but there were no checkboxes next to site
names. They were unsuccessful in their initial attempts to add a job/site or a file to their collection.

- The checkbox in front of a job was sometimes needlessly selected prior to selecting “Search for files” or “List files”.
- Until they are familiar with a site’s content, curators expressed a preference for browsing to discover files of interest, rather than searching.
- The load time for listing files is very long. One curator commented that the hourglass was good as an indicator that the files list was being generated.

After being guided to the Build Collections screen to add contents to their collection, one curator immediately clicked on a site name but was then stumped regarding how to add a job to their collection. For this curator, the process of checking a job’s checkbox, selecting a collection to add the site to, and selecting the Add button was not obvious. Here is how this curator proceeded:

- First they clicked on the collections drop-down options next to the Add button, perhaps looking for a site to add there.
- Next they selected the Collections top-level menu options and wondered if they should go to Manage Collections?
- When prompted that they were attempting to add a site to the collection they were building, the curator selected Site → Add Site and proceeded to create a new site, presumably thinking this would add the site to their collection.
The confirmation screen returned after adding contents to a collection provides three navigational text phrases. After adding contents to a collection, it was not always clear which one curators should select to add additional contents to their collection.

- “Am I managing collections or managing sites?” In fact this curator was building a collection and had just added content to the collection.
- “Back to Previous” was an obvious choice to some curators for continuing to add content to a collection.

When a second job was added to a collection, the confirmation screen one curator received listed another job in the collection. However:

- Such a list was not included on the confirmation screen presented to another curator after they successfully added a second job to their collection.
- The confirmation screen returned after a file was added did not list the other contents (jobs) in the collection.

**Recommendations**

- Consider decreasing the number of user options on the Manage Sites screen. Perhaps from the Sites drop-down menu, there could be more navigational options related to smaller chunks of functionality.
- On the Collection Updated confirmation screen, consider:
  - Removing the navigational text “Manage Collections” and “Manage Sites”.
  - Changing “Back to Previous” to text indicating a curator could continue adding content to a collection. (This latter might not be necessary and could be tested further.)
  - Consistently listing the existing contents (jobs and files) in the collection or not listing them. Not certain how helpful this list will be to curators over time.
  - Having two action icons or text obviously present on screen: “Add more content” (in lieu of Back to Previous) and “View content” (in lieu of a list).
- From confirmation screen returned after a curator successfully adds a collection add text to guide curators in the process of adding jobs or files to add to the collection as well as a link to “Build Collections” screen
- From the Build Collections screen and possibly the Manage Sites screen, provide indications regarding the collections in which a job is included.
- Adding files to collections needs to be more obvious in the interface.
- If it does not thwart the work flow, consider having one consistent method by which jobs and files are added to collections. From that screen, provide:
  - An obvious indication of how to add:
    - Capture jobs for a site
    - Files from capture jobs
- Consider adding breadcrumbs to screens to identify screen location within the WAS.

Scenario 3. Manage Collections & Capture Results

Scenario 3: Using the collection related to drought in California that you created in Scenario 2, you now want to manage the materials in the collection.

Task 1 - Delete a capture job from the collection.

Task 2 – Add a PDF file to the collection.

Task 3 - Delete a file from the collection.

Findings

- One curator selected Sites → View Results when attempting to delete content from their collection. (Note: This functionality was not available.)

- Many curators selected Collections → Manage Collections when attempting to view their collection’s contents but could not readily see a means to delete.
  - Some discovered the icon to view the contents.
  - Others neither discovered this icon on the page nor read the Icon Key in the side bar.
- One remarked upon the ‘????’ under number of Files: “What’s that?”
- One curator found the icons quite small on the screen and presumably was less drawn to them to accomplish the task of viewing contents prior to deleting a job.
Selecting the view collection content icon from the Manage Collections screen, presented curators with the screen above. Upon initial viewing, it was not readily apparent to some curators how to delete a site listed in the Contents from this screen.

As stated before, most curators readily understood and interacted with the [+/-] tree expansion to view the contents for a site. However, once again, some curators never experimented with this capability and concluded they had arrived in the “wrong” functional area to delete a site. Selecting other options from the drop-down menus was often their next navigational move.

The first item listed under site contents was a folder with a date/time title and a trash can icon. Some questioned what the date/time title referred to and were hesitant to select the trashcan icon. “Do I want to delete this?”

In addition to their uncertainty regarding the date/time title, it was not always clear to everyone what “Entire job” referred to.

The trash can icons were at two levels: date/time title (job) and Entire job. Curators wondered at what level they should delete a job and wondered about the impact of their choice. One curator wondered if two icons were necessary when only one job for a site had been added. Would not the date/time title suffice for “Entire job”?
The use of the term “job” was confusing to some curators.

- For the first site in the list, the curator had only added one file to the collection but the mouse over text for the trash can icon next to the date and time stated: “Remove job from collection.”
- One suggested that “job” referred to a crawl in queue or in process while a “capture” was the result of a completed crawl.
- Another decided that deleting a “capture job” meant deleting the “entire site”.

One curator noted that the trash can icon text next to Entire job stated something slightly different (“Remove entire job . . .”) from the same icon next to the date and time (“Remove job . . .”).

After deleting the Entire job by selecting the trash can icon next to that text, the content structure still showed the site name, the job date/time with an icon, and Included. This was confusing:

- Was the job deleted?
- At which level in the tree structure is the job really?
When curators added a file to their collection and the entire capture job was already in the collection, confusion sometimes resulted:
  - One curator was unable to locate the file in the Contents tree and concluded that they were in the wrong location to delete a file they had added.

Terminology
  - A few curators were not certain how to delete content from and add content to an existing collection and wondered if this would be considered "managing" or "building" the collection?
  - A few curators offered that "manage" implied working with a collection that already existing and "build" implied creating a new collection.
  - When a curator selected "List files" from the Build Collections screen, the Tips text caused some confusion:
    - The use of the term "filter" in the first paragraph of the Tips text and the label "Show" for the drop-down menu of mimetypes was confusing for one curator.
    - The heading text (Mimetype) could not be clicked to sort the files by mimetype as stated in the second paragraph.
  - One curator did not understand the meaning of the Update button.
Results → Search
- One curator wondered if the search was of the text in the URL and/or of the contents?
- One curator thought the search results should indicate in what collections an item was already included.

Recommendations
- Make it more obvious how to view the content of collections, for example, in addition to an icon to view the content of collections, some curators may need descriptive text to accommodate curators’ different interaction styles.
- Notify a curator when they attempt to add a file from a capture job to a collection and the parent capture job is already in the collection.
- After an entire job has been added to a collection, allow curators to remove individual files from the job.
- On View Results → Files Tab screen, consider changing "Update" button label. Perhaps "go" or an "→" might be more commonly understood.
- Ensure that the trash can icon consistently refers to the same action throughout the WAS. Currently it appears to refer both to deleting and to deactivating.
- Collection: Overview Tab-Contents
  - If only one job for a site is included in the collection, only include the job title (date/time in the tested WAS release).
  - Consider using the "Included" folder only if specific files have been added to the collection.
  - Consider labeling jobs in the list of collection contents more meaningfully, for example, including the word "Captured" before the date/time. (Note: This may not be necessary. Resolution of the site-job terminology confusion and some experience with the WAS may resolve this concern.)
- When all content for a site is deleted from a collection, do not list the site name in the collection Contents on the Overview tab for the collection.
Scenario 4. View and Add Metadata

Scenario 4: Earlier you edited the descriptive data for the 'Crawler Test' site. Now you would like to view files within the site and add comments to specific files.

Task 1 - Display any 'html' file from the results for the 'Crawler Test' site by searching the results using the keyword "hop".

Task 2 – View the detailed metadata record.

Task 3 - Add a comment about the contents of this file.

Findings

Note: Only three curators completed the tasks in Scenario 4.

- In general, curators were successful in locating and displaying a file, viewing the detailed record, appending comments to the metadata, and displaying the file.

- One curator took this path to display an html file from the Crawler Test site: Sites → Manage Sites → View Capture History icon.
  - Curator was uncertain if this was the path previously taken to get a list of files.
  - After initial evaluation of the screen, the curator concluded that this was not the correct location to view files.

- It appears that the Results button does not initially attract curators or alert them as the means to interact with content and display files.

- Curators bring their own frame of reference, including experiences with other web capture tools and the collections they intend to build, to the WAS.
  - One curator did not understand why there were so many jobs for the Crawler Test site and suggested that it was because they only ran for one hour each. The inference seemed to be that more than one 1-hour job was needed to capture the entire site.
While working in an earlier scenario, another curator thought that only one job (or file) for a site would ever be included in a collection, otherwise it would “just [be a] duplicate”.

From this screen, a curator can select a site or the results button for a specific job. When the curator selects a site name, they are presented with the Site Summary screen and cannot find a way to search for and display files.

One curator suggested that Results be added to the top navigation bar.

One curator selected the site name and from the Site Summary: Job Settings tab selected the seed URL, which:
- Displayed the active web site, not the archived site.
- Predictably, provided no functionality to search the site for files.

One curator noted that the search term was in either or both the URL text and in the Abstract. This curator found the highlighted search term in the Abstract helpful in seeing why the file is listed in the search results.

From this screen one curator easily displayed the detailed record for a file.

One curator suggested that a capability to filter search results, for example, by mimetype, might be nice.
One curator commented that the site-level metadata would be valuable for troubleshooting.

Adding comments to a file was very intuitive. One curator wondered what the length of the comment field was and if text entered would wrap on the screen.

**Recommendations**

- If possible, it may promote more intuitive navigation for both the Site name and the Results button on the View Results screen to lead to the Results screen for the selected capture job (with its 4 tabs: Overview, Search, Files, and Reports). Otherwise, it might be best to eliminate the hyperlink for the Site name on the View Results screen.
Additional Comments

Findings

- Assessing a web site’s structure is a very challenging task. One curator thought that any tools that could be provided to assist in this task would be very helpful in determining the scope settings for sites.
  - Guidelines for pre-capture assessment and post-capture assessment would be very helpful. For example, “Prior to defining your site, assess the site on the web by doing the following three tasks.”
- One curator noted that sites of interest in their collection are sometimes available in English, Kurdish, and Arabic versions. For researchers, it is of interest to compare and contrast the content and look-and-feel of the different versions. Additionally, it is critical for researchers and those viewing the archived collection to have an obvious indication that they are not viewing a live site but rather a site captured at a certain date and time and presented from the archive as an exact duplicate of the original site.
- One curator wondered if Unicode was supported. This would enable other language versions of the descriptive data for sites and collections.
- One curator stated that the comparison of different captures of the same site would be valuable in determining if the frequency of capturing the site needed to be modified. This curator expected that the terms (i.e., changed, new, etc.) on the Compare Sites screen would be explained in the Tips sidebar.

Recommendations

- Some of the curators’ web collection plans organize their content into categories. Organizing a WAS collection’s content in this manner would be useful.
- It would be helpful to have the ability to easily sub-divide an existing collection without having to delete content from the existing collection and add it to a new collection.
- Once collections are accessible to end users, provide the ability for them to search the repository with the purpose of identifying any collections in the repository that match their search criteria.
- One curator discussed the idea of curators having the ability to add any site or file within the entire repository to a collection. This person envisioned sites, regardless of which curator initially added them to the repository, as being part of a number of different collections curated by various people. This poses several challenges, one of which is responsibility for maintenance of the descriptive metadata in such a manner that it meets the needs of the various collections in which the site is a member.
Test Script

Scenario 1: Create / Capture a Site

In your role as a librarian, you need to create and define a site you want to capture. But first you have to edit the descriptive metadata for a site you previously defined and also recapture the site.

Please do the following:

1. Edit the descriptive data for the ‘Crawler Test’ site, which contains files used to measure the effect of different heritrix crawl settings on what is captured:
   a. Edit the description
   b. Add a subject
2. Initiate a capture of the ‘Crawler Test’ site.
3. Create and define a new site of your choice. Caution: When you are finished creating your site, please do not capture it.

Scenario 2: Create a Collection

You are a librarian interested in California water issues, in particular about web-based materials related to droughts in the state. You have already captured a few key web sites:

- California Department of Water Resources
- California State Water Resources Control Board
- Bureau of Reclamation Mid-Pacific Region

Now you are going to create a new collection using materials from these sites.

Please do the following:

1. Create a new collection.
2. Add the capture results for one of your three key web sites to the collection.
3. Add the capture results for another of your three key web sites to the collection.
4. Add a single file from the capture results of your third key web site to the collection.
Scenario 3: Manage Collections & Capture Results

Using the collection related to drought in California that you created in Scenario 2, you now want to manage the materials in the collection.

Please do the following:

1. Delete a capture job from the collection.
2. Add a PDF file to the collection.
3. Delete a file from the collection.

Scenario 4: View and Add Metadata

Earlier you edited the descriptive data for the 'Crawler Test' site. Now you would like to view files within the site and add comments to specific files.

Please do the following:

1. Display any 'html' file from the results for the 'Crawler Test' site by searching the results using the keyword "hop".
2. View the detailed metadata record.
3. Add a comment about the contents of this file.

This is the end of the session. Thanks very much for your help!