

End User Interview Questionnaire

Notes to Interviewer

Start the interview by (a) creating a comfortable atmosphere in which people feel valued for their participation and (b) establishing the context for the discussion. Be prepared to offer definitions of key concepts and examples as outlined in the boxes below.

Opening

“When a library creates a web archive it is important to ensure that user or patron needs inform their plans. The purpose of this discussion is to elicit your needs and thoughts regarding web archives. Our discussion will address needs and issues in the major phases of web archive creation with a particular emphasis on the selection of web-published materials and on users’ needs for access to these materials.”

Key Concepts

Web-Published Materials

Web-published materials are accessed and presented via the World Wide Web. The materials span the cultural heritage spectrum and include a range of material types from text documents to streaming video to interactive experiences. Web-published materials are both dynamic and transient. They are at risk of disappearing. Web archives preserve web-published materials.

Web Archive

A web archive is a collection of digitized or born-digital materials, many of which are initially made available on the web. Some of these materials may also exist in other forms but the web archive captures the digital versions for posterity and may provide users access to the materials long after the web sites themselves are no longer in existence.

Digital Archive Examples

CyberCemetery

<http://govinfo.library.unt.edu/>

“The University of North Texas Libraries and the U.S. Government Printing Office, as part of the Federal Depository Library Program, created a partnership to provide permanent public access to the Web sites and publications of defunct U.S. government agencies and commissions. This collection was named the “CyberCemetery” by early users of the site.”

UCLA Online Campaign Literature Archive
<http://digital.library.ucla.edu/campaign/>

“Every American election produces thousands of campaign flyers, pamphlets, posters, and bumper stickers, generally called "campaign literature." These documents provide an important record of the campaign, its participants, issues, and tactics. The UCLA Online Campaign Literature Archive presents a subset of the materials in the complete Campaign Literature Collection. It contains copies of all archived websites plus scanned images of selected print materials. (Currently, this includes all elections from 1908 to 1939, and some materials from 1940 and 2000.) These are made available on the web for the use of scholars and students across the world.”

Opening. Participant Background & Need for Digital Archives

Facilitator:

Indicate which category best describes where the participant works/studies.

- a. _____ K-12 School
- b. _____ College or University
- c. _____ Federally Funded Institution
- d. _____ State Government Institution
- e. _____ Local Government Institution
- f. _____ Non-Profit Organization
- g. _____ Corporate Institution
- h. _____ Specify Other: _____

1. What is the name of your department?

2. What is your current position, academic status, or title?

3. How many years have you been in this position? _____

4. Do you use web-published materials in your:

- a. Research activities? _____ Yes _____ No
- b. Professional activities? _____ Yes _____ No

5. Have you ever tried to retrieve a critical document or a file from the Web that was no longer there? How often has this happened?

6. Think about one of those incidents and describe the circumstances? How severe was the loss?

Topic 1. Selection of Materials for a Digital Archive

- Which of the following information sources in your discipline are accessible on the web? How important are these web-published information sources in your discipline, for either research or professional information?

	Web Accessible?	Importance		
		High	Medium	Low
Journals & Periodicals				
Books, Brochures				
Databases				
Newspapers				
Videos				
Audio files				
Technical & Research Reports				
Proceedings of Meetings & Symposia				
Doctoral Dissertations & Master's Theses				
Government Records or Documents				
Unpublished Work & Publications of Limited Circulation				

- Are there other web-accessible information sources that are important in your discipline?

- For each type of web-accessible information in your discipline, how long does it provide significant value after publication?

	# Years				
	< 1	1-3	5	10	>10
Journals & Periodicals					
Books, Brochures					
Databases					
Newspapers					
Videos					
Audio files					
Technical & Research Reports					
Proceedings of Meetings & Symposia					
Doctoral Dissertations & Master's Theses					
Government Records or Documents					

The Web at Risk: Needs Assessment tool kit

	# Years				
	< 1	1-3	5	10	>10
Unpublished Work & Publications of Limited Circulation					
Other:					
Other:					
Other:					
Other:					

Topic 2. Authenticity of Materials in the Archive

4. Suppose that source materials originally published on the web are no longer available except as 'digital copies' in an archive. What issues arise if you cite an archive as the material source?

5. How will you judge the authenticity of materials retrieved from a web archive? For example: "Is this item 'really' a transcript of the 1986 US House hearing on gun control?"

6. What is the impact to you of the removal of some parts of a web page from the source material before it is archived, for example, a particular image from a page?

- a. Is the authenticity of the source material compromised?

- b. What if it was removed in accord with university or organizational policy?

7. What do you think about an archive altering or tagging web pages in some way to alert archive users to a modification of the original page?

The Web at Risk: Needs Assessment tool kit

8. What can the archive do in a web page or a file to build your confidence and trust in the authenticity and credibility of materials?

9. Changes to materials can be expected due to copyright requirements or software migration. Discuss how keeping records of changes and the reason for the changes might help build trust and confidence in archived materials.

Topic 3. Interacting with Materials in the Archive

10. How do you expect to interact with the web archive? Is it important that the archive interaction mirror your experience of the original 'live' website?

11. Some web pages include active elements such as hyperlinks and interactive forms that are no longer active in archived materials. How should the archive handle each of the following previously active elements? How should they be presented to users?

- a. Email Links

- b. Links to Materials Accessible in or from the Archive

- c. Data Collection Forms

The Web at Risk: Needs Assessment tool kit

d. Can you think of others?

12. Some web sites store personal information about their visitors in order to customize pages for the user when they visit the site. How do you expect customized web content to be handled in a web archive?

13. Some web pages are generated upon request either programmatically or using information retrieved from a database. How do you expect this dynamic web content to be handled in a web archive?

Topic 4. Searching the Archive

14. What information do you expect to find included in a summary of results from a search of a web archive? For example, what are the minimum attributes you would expect? What additional attributes would be nice to have?

15. How important are each of the following to you in locating and selecting web materials using a search engine?

	Importance		
	High	Medium	Low
Topic or subject			
Title			
Author			
Original URL			
Publication date			
Organizational name			
Project name			
Format			
Full-text using any keyword			
Other: _____			
Other: _____			

16. Many web-published materials are frequently modified and a web archive may capture different versions of the same source materials over time. Considering your needs, how important would it be for you to locate different instances of the same item harvested at different points in time?

17. Should there be a separate summary for each version of an item in the archive or should multiple versions be listed in a single summary?

When a web archive is created, it is predictable that some objects will be archived in multiple formats or media types. For example, archives of testimony before a commission might include a video file, an audio file, and a text file.

18. How about each format of an item? Should there be a separate record for each format?

Topic 5. Preservation of Archived Materials

19. Considering cost factors such as the hardware and software required for presentation and storage of materials, how important is it to you that a web archive retains multiple formats of original materials, for example, a video file, an audio file, and a text file with the same content.

Not Important	A Little Important	Somewhat Important	Very Important	Extremely Important
1	2	3	4	5

20. What implications do you see if the original formats of an item are not saved? For example, if a video recording of testimony is created and then transcribed to a .pdf file, what are the implications if only the .pdf version is saved? Does it make a difference if the original video recording is of poor quality?

21. What implications can you anticipate for current or future researchers if only certain file types are retained, for example the text but not the audio or video files?

In question 22, material format is different from material type. Examples of material types are documents, images, audio files, or video files. A certain type of material could be formatted using one of many encoding methods. For example a document might be encoded using html or xml standards and an image might be encoded using jpeg or gif standards.

Additionally, multiple versions of a single material type, formatted using the same encoding standard, may exist within an archive. For example, multiple copies of a document encoded in PDF format could be captured over time.

22. Retaining or removing materials from an archive will involve trade-offs related to costs. In making these types of decisions, how much consideration should be given to:

- a. Frequency of access in the archive

The Web at Risk: Needs Assessment tool kit

b. Material format

c. Multiple formats

d. Multiple versions

e. Length of time in the archive

f. Other factor
