

# ***Testing Library Web Sites: ARL Libraries Weigh In***

*Mary Pagliero Popp*

Academic libraries rely on their Web sites and on their Web catalogs to provide users with a variety of information, access to full-text resources and digital library services. This growing reliance on the Web to deliver library services and resources means that academic libraries must pay more attention to how effective their sites are and how easy they are to navigate and to use efficiently. Librarians on the front lines need to know more about the methods of testing and evaluating Web sites to help them improve existing Web resources and plan new ones.

At Indiana University Bloomington, we completed a comprehensive study of our Web site during spring 1999, working with two outside consultants and the usability coordinator of our campus computing service. At the time we were doing our initial work in 1999, information about what other libraries were doing had not yet been met by publications focused on libraries. The library literature has now begun to include articles describing usability testing techniques and programs of evaluation. To date this literature, however, is small and includes few descriptions of what how libraries have tested their own sites.

## **Mining the Library Literature**

Writings in the library literature describing procedures for Web testing and evaluation started appearing with some

regularity in 1998. One of the first major publications detailing a specific program that tested an academic library Web site was the paper presented by Michelle Clairmont, Ruth Dickstein, and Vicki Mills at the Living the Future 2 Conference at the University of Arizona in 1998. The authors described in detail the work they did to evaluate and redesign Sabio, the University of Arizona Web site using paper prototypes and formal usability tests.<sup>1</sup> Other, later articles have provided additional details. Veldorf, Prasse and Mills<sup>2</sup> provided an overview of usability testing methods and research in information retrieval on the Web, then described methods used at Arizona, including a survey, focus groups, card-sorting and usability tests. Mills and Dickstein gave in-depth information about their heuristic testing and design walkthrough.<sup>3</sup>

Librarians at the Pullman and Vancouver campuses of Washington State University used usability tests with students to test their Web-based catalog, and several parts of their Web site, including sections devoted to article indexes, full-text and other library catalogs.<sup>4</sup> Two studies have been done at Cornell using focus groups and interviews. Payette and Rieger studied the Mann Library Gateway for the sciences. Zsuzsa Koltay and Karen Calhoun, in a paper presented at the 9th National ACRL Conference in Detroit, described their focus group study of the Cornell Informa-

---

*Mary Pagliero Popp is the information technology public services librarian at Indiana University Libraries.*

tion Gateway.<sup>5</sup> Danish academic library Web sites were studied by comparing them against a list of Web site evaluation criteria describing effective academic library Web sites. Librarians from Danish academic libraries were asked to review one to five sites from other libraries against these criteria.<sup>6</sup>

### **Web Site Evaluation Methods**

The literature of human computer interaction (HCI) and of usability engineering can help librarians to identify methods for testing Web sites. There are a number of useful evaluation methods that can be used at different times in the development of a Web site. All of these methods require a formal and planned testing process. Methods range from those that can be conducted by the design team without any direct input from users to those that provide extensive empirical data on the ways users interact with the Web site. Testing methods are listed below in order from least to most user involvement. For libraries, use of testing methods that allow the Web site designers to watch users work with the Web site should be a high priority.

#### ***First steps***

The first step in any evaluation process is to decide what needs to be studied and what questions answered. It is crucial to develop a statement of purpose for the evaluation, to identify the audience for the Web site and its information needs, and to determine exactly what will be tested. This planning process will make the choice of a testing method easier and more logical.

#### ***Heuristic evaluation***

This technique is also sometimes called a "usability audit." Jakob Nielsen calls it "one of the main discount usability engineering methods" easy, fast and cheap."<sup>7</sup> A small group (usually three to five people) of evaluators examines the site and compares each element of the site it to a list of recognized usability principles, called "heuristics." The evaluators review the site independently and agree in advance about the categories of problems to note. Each evaluator reviews the site at least twice and prepares a report or presents findings in a meeting. Heuristic evaluation is often done best at the beginning of a project and then repeated at later stages of development of a site.<sup>8</sup>

#### ***Cognitive walk-through***

This is another method that involves expert evaluators. The evaluators design specific task scenarios. The user's goals

and purpose for each task are defined and tasks are broken down into relatively small pieces. The evaluators role play the part of the user working with the site, noting problems, paths, and barriers, essentially reviewing the ease of learning the site.<sup>9</sup> This work can be done at the very beginning of the design process when the system is just a series of hand-sketched screens or after the Web site has been designed.

#### ***Individual library staff/administrators look at Web site in systematic way***

This is a method that most libraries have used to "test" a Web site, either as part of a Web evaluation plan or as a stand-alone activity. It can provide useful insights, but it cannot take the place of formal evaluation methods. It should not be confused with either heuristic testing or a cognitive walkthrough.

#### ***Analysis of logs for Web pages***

Server logs for Web sites can be set up to review activity on the site and to provide summary data about users, including what IP domain they come from, the browser they are using, the pages that have been accessed most often, etc. This useful summary data can sometimes be supplemented by transaction log data, particularly from Web-based catalogs.

#### ***Online or Web survey***

Many libraries have found that a survey linked from their Web sites can provide useful information. The survey can be set up to automatically generate survey results and compile them into a database. Those who answer the survey are self-selected; they choose to access the survey and to answer it. Such a survey can provide satisfaction data, information about purposes for which the users are connecting to the site, and other similar information.

#### ***Focus group discussions***

Focus groups have been used in libraries for many years to learn about what users want. For Web evaluation, they can be used to learn more about users and their needs, to review an existing or proposed Web page or Web site, to determine user satisfaction or to find problems with a site. Groups typically have eight to twelve participants. They can be conducted in person, via telephone, or online. Transcripts should be made of the discussion. Participants are chosen from among the groups composing the Web site's target audience. Each group is homogeneous in its makeup.<sup>10</sup> Focus groups differ from general group discussions in that

they are structured to gather information and open-ended questions are developed to make sure that the needed information is gathered.

**Usability testing**

In usability testing, evaluators observe users as they work with the Web site, menu, Web catalog or other Web resource to perform a set of typical tasks. The goal of usability testing is to improve the site, to find what parts of the site work well for real users completing representative tasks, and where there are problems with terminology, navigation, the arrangement of content and similar issues.<sup>11</sup> Usability tests can be done in a laboratory setting, with two rooms separated by a one-way mirror and videotape cameras focused on the user and on the computer screen. An equally effective study can be done in a room with a computer, a tape recorder, a facilitator who works with the participant, and an observer who takes notes. Feedback is obtained from participants through the “think-aloud” protocol. The participant keeps up a running commentary as the tasks are completed, articulating thought processes, questions, confusion, and decisions.<sup>12</sup> Evaluators can also measure the number of clicks needed to complete a task and the amount of time the user needed to complete each task.

**Paper mockups/prototyping**

Usability tests can be conducted using a live Web page. Alternatively, you can test using paper versions of screens, with a separate page for every choice on the screen. The user is given a set of tasks to complete and is asked to think aloud, telling the testers how he or she would complete the task.

**Field study**

A field study is a study of users as they work with the resource in their own environment to fulfill their own information needs. Users are chosen from among the groups who make up the target audience for the Web site. Library staff watch (and perhaps videotape) these participants as they use the library Web-based resources where they would most frequently do their research—in their homes or offices or in a library setting. Participants are asked to think aloud about what they are doing and why. Observers take notes about what the participant says and does and the successes and problems the user has with the library Web resources being used.

**ARL Survey: What Libraries Are Doing**

During the fall of 1999, I surveyed the 122 ARL member

libraries about their Web evaluation activities. By the end of January 2000, I had received 86 returns, a response rate of 70 percent. Of these, three were returns from libraries whose policies do not permit them to answer surveys; eighty-three usable responses were evaluated. The survey questions covered three broad areas: 1) whether libraries had Web sites and whether they had conducted formal testing of the sites; 2) what libraries had tested and what methods they used; and 3) how library staff members had learned about testing techniques and what they still needed to know. The survey generated an extensive amount of data. This data is being further supplemented by interviews. The preliminary results are interesting and provide a useful starting point for further discussion.

**Are ARL libraries testing their Web sites?**

Fifty-two (63%) of the eighty-three libraries responding stated that they were doing some sort of evaluation of their Web sites. Of these libraries, two used only individual staff reviews and analysis of Web server logs. All other libraries used at least one testing method from the literature of HCI. In addition, two libraries described firm plans for evaluation in the year 2000.

**What parts of ARL library Web sites are being tested?**

Most libraries tested the entire site or the core pages and some other part of the Web site, such as the catalog. 46% of the libraries indicating a testing regimen (24 libraries) had tested the entire Web site in some way. Another 46% of the libraries involved in testing had tested core Web pages alone or in combination with an additional section of their sites. Nine libraries (11%) had tested their Web catalogs. Five libraries had tested their electronic indexes pages.

**What testing methods are being used?**

The most frequently used methods are shown in table 1. A few libraries had used alternative methods. One had studied the information architecture of the site. Two others had used card-sorting techniques.

Test Method	No. of libraries	% of libraries
Cognitive walkthrough	27	52%
Individual staff review	26	50%
Focus groups	23	44%
Online or Web survey	21	40%
Usability testing	19	37%
Web server log analysis	19	37%

Only seven libraries had used only one testing method. Fifteen libraries had used two testing methods. Ten libraries had used three methods. Twenty libraries had used four or more testing methods.

#### **Who conducted the tests?**

Nearly all libraries used library staff to conduct the tests, usually a Web oversight committee. Three libraries (New York Public Library, Yale, and Indiana) had used outside consultants to do some part of their testing. Two libraries used staff from other parts of campus to lead focus groups.

#### **When were the tests conducted?**

Libraries involved in testing often conducted different tests in different years. Two tests were done in 1996 and fifteen in 1997. Thirty-six libraries were involved in testing in 1998. 1999 brought a huge jump in the number of participating libraries. Eighty-two libraries did some Web tests in 1999. Of these, eighteen libraries began usability testing for the first time. Libraries using staff-based methods, such as heuristic testing, cognitive walkthroughs, staff reviews, and Web server log analysis often indicated that these tests were done regularly.

#### **Comments and conclusions**

Interest and involvement in the use of formal methods to test and evaluate library Web sites grew exponentially, with an explosion of work done in 1999. Although many libraries based their testing on methods that only involved staff, more than 40% of libraries testing used methods that directly involved users. Primarily library staff did the work of evaluation, making the issue of staff training and continuing education important.

#### **ARL Survey: How Librarians Learn about Web Testing**

Respondents to the ARL survey were asked to choose from a list of ways they and their colleagues had learned to evaluate their Web sites and to add any additional methods they had used. They were also asked to recommend three resources others might find useful in learning about Web testing. Finally, respondents were asked to list topics related to Web site development and evaluation that they would like to learn about in more depth.

#### **Learning Methods**

Most libraries (39) used two to four methods to learn about Web evaluation. Reading books and articles (44 libraries) and finding information on the World Wide Web (37 li-

braries) were the most frequently cited. Twenty-five respondents had attended conferences and workshops. Seven respondents indicated they had taken formal courses. Training from consultants or the local computer center was available to staff from four libraries.

#### **Recommended Resources for Learning More**

There was little consensus about the most useful resources to learn more about Web evaluation. Nineteen respondents (37% of those who are involved in testing) either gave no answer to this question or stated that they did not know enough to recommend resources. Twenty-three of the recommended resources were mentioned only by one library; six resources were recommended twice. Only three resources were named 3 or more times:

1. Jakob Nielsen's Web site or its Alertbox section: <http://www.useit.com>—11 recommendations.
2. Jared Spool, et. al. *Web Site Usability: A Designer's Guide*. San Francisco: Morgan Kaufman, 1998. (1997 edition from North Andover, Mass.: User Interface Engineering contains the same content.)—8 recommendations.
3. Review other library Web sites—3 recommendations.

#### **Continuing Education Needs**

Answers to the question about what respondents would like to learn in more depth produced a similar lack of interest. Twenty-eight respondents (54%) gave no answer to this question. Of those who did answer the question, there was little consensus about what respondents wanted to learn. Twenty-three suggestions were listed only one time. Usability methods were mentioned seven times. Heuristic methods and procedures for conducting focus groups were each mentioned three times.

#### **Comments**

These results are somewhat surprising. The literature of human computer interfaces (HCI) and usability engineering is rich. Most of the major practical testing handbooks were published in the mid-1990s. It seems that many librarians are either unfamiliar with this literature or do not find it useful enough to recommend to others. Librarians also have few suggestions for what they would like to learn. We can speculate that this is because they feel they already know what they need to do the testing they conduct. More research is needed to learn whether it is important that the library literature and library organizations devote more time to discussion of Web evaluation methods in the library setting and what topics are important.

**What Next?**

It is clear that libraries are rapidly moving forward in testing the effectiveness of their Web sites. At the same time, the library literature is not keeping up. Librarians need to talk more about what they are doing, to share their results, and to set an agenda for future research. Is this an area that will continue to grow? Only time will tell.

**Acknowledgement**

The author would like to thank the Indiana University Librarians Association (InULA) for its support grant to cover part of the costs of the ARL study.

**Notes**

1. Michelle Clairmont, Ruth Dickstein, and Vicki Mills, "Testing for Usability in the Design of a New Information Gateway." Paper presented at the Living the Future 2 Conference, University of Arizona. 1998. <<http://www.library.arizona.edu/library/teams/access9798/lft2paper.htm>> Accessed January 16 2001.
2. Jerilyn R. Veldorf, Michael J. Prasse, and Victoria A. Mills, "Chauffeured by the User: Usability in the Electronic Library." *Journal of Library Administration* 26.3/4 (1999): 115-40.
3. Vicki Mills and Ruth Dickstein, "Usability Testing at the University of Arizona Library: How to Let Users in on the Design." *Information Technology and Libraries* 19 (2000): 144-51.
4. Two articles have been written about this work. The first: Janet Chisman, Karen Diller, and Sharon Walbridge, "Usability Testing: A Case Study." *College and Research Libraries* 60 (1999): 552-69, provides detail about the study itself. The second interviews the participants to learn how well the study worked and what they would do differently: Nicole Campbell, Sharon Walbridge,

Janet Chisman, and Karen R. Diller, "Discovering the User: A Practical Glance at Usability Testing." *Electronic Library* 17 (1999): 307-11.

5. The first study only looked at the Mann Library Gateway: Sandra D. Payette and Oya Y. Rieger, "Supporting Scholarly Inquiry: Incorporating Users in the Design of the Digital Library." *Journal of Academic Librarianship* 24 (1998): 121-29. The second looked at the common gateway for networked resources at Cornell: Zsuzsa Koltay and Karen Calhoun, "Designing for WOW!: The Optimal Information Gateway." In *Racing Toward Tomorrow: Proceedings of the Ninth National Conference of the Association of College and Research Libraries, April 8-11, 1999*, 154-60. (Chicago: ACRL, 1999).
6. Helge Clausen, "Evaluation of Library Web Sites: The Danish Case." *Electronic Library* 17 (1999): 83-87.
7. Jakob Nielsen and Robert L. Mack, eds., *Usability Inspection Methods*. (New York: Wiley, 1994), 25.
8. J. T. Hom, "Usability Methods Toolbox: Heuristic Evaluation." (2000) <http://www.best.com/jthom/usability/heuristic.htm> Accessed January 19, 2001.
9. Cathleen Wharton, John Rieman, Clayton Lewis, and Peter Polson, "The Cognitive Walkthrough Method: A Practitioners' Guide." In *Usability Inspection Methods*, Jakob Nielsen and Robert L. Mack, eds. (New York: Wiley, 1994), 105-40.
10. Lynn Silipigni Connaway, "Focus Group Interviews: A Data Collection Methodology for Decision-Making." *Library Administration & Management* 10 (1995): 231-39.
11. Joseph S. Dumas and Janice C. Redish, *A Practical Guide to Usability Testing*. (Norwood, N.J.: Ablex, 1993), 22.
12. Jeffrey Rubin, *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests* (New York: Wiley, 1994), 217-19.