UC Office of the President
CDL Staff Publications

Title
Methods to Tame the Madness: A Practitioner's Guide to User Assessment Techniques for Online Finding Aid and Website Design

Permalink
https://escholarship.org/uc/item/1wm2b563

Journal
RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage, 13(2)

Author
Hu, Rachael

Publication Date
2012-11-01

Methods to Tame the Madness: A Practitioner's Guide to User Assessment Techniques for Online Finding Aid and Web Site Design

~ Though this be madness, yet there is method in't. ~
Hamlet, act 2, sc. 2

SHAKESPEARE COULD HARDLY HAVE anticipated the proliferation of Web resources and content available to academic researchers in the twentieth century, yet the common expression "method in the madness" derived from his work is strangely applicable to assessing how users engage with these rich and complicated resources. With the ubiquity of Internet connectivity and electronic devices, searching for information has become an almost exclusively online endeavor. For the special collection libraries and archives that deal in primary source collections and a variety of material formats, the online presentation of their materials gives rise to special challenges due to the complex nature of the descriptive records that provide access to the collections they manage. For archival collections, a formally structured descriptive document known as a finding aid has been used to describe their content and organization. In the 1990s, the traditional paper-based finding aid made its way into the Internet age and was transitioned into the Encoded Archival Description (EAD) standard, a machine-readable XML-based format designed to facilitate its online display and indexing.[1] This new format precisely captured and translated the inherent structure of the finding aid into an online environment. The EAD standard also enabled archival collections to become systematically discoverable on the World Wide Web for the first time and encouraged the growth of Web sites dedicated to the aggregation of these descriptive records as well as providing environments to view digitally scanned facsimiles of objects from these collections.

As the creation and development of Web content—both commercial and academic—grew, an important indicator of an archives' continued value became a user-

---

1. Library of Congress, "EAD: Encoded Archival Description Version 2002 Official Site," available online at www.loc.gov/ead/.

*Please note: URLs for all Web sites referenced herein were valid at the time of article submission.

friendly Web site displaying easy-to-find descriptions of its collections, whether published directly online or available at its physical location. Beginning in the 1990s, user-centered assessment and design practice emerged as a key component in the design and development of Web sites in the commercial world. Recognizing the value and benefits of this assessment-oriented approach, library and archival communities have started to build their own body of practice geared specifically toward addressing the ways in which users interact and engage with archival resources online. Nevertheless, some archives administrators indicate that they continue to struggle to build and maintain a consistent program of assessment activities for their online finding aids and collections. One grant-funded study in 2008 looked at archivists' views of user-based evaluation.[2] Archivists responded that they saw the value in user-based evaluation since they needed to capture and document use of their collections both for onsite users and those who accessed their materials online. Many barriers to conducting such assessments were identified, however, including "money to hire outside experts, time to conduct user-based evaluation research in-house, and expertise."[3]

This article aims to provide archivists and special collection staff with a practitioner-level primer for understanding user evaluation techniques and methods. Strategically applied methods can help guide staff in making intelligent decisions on how to determine and design effective feature sets for their Web resources. This overview will provide archivists and library administrators with an understanding of how to select appropriate assessment methods, determine the resources necessary to implement these methods, and appreciate the inherent strengths and limitations of various techniques.

Assessment Literature
Beginnings of HCI and Usability Disciplines

To better understand usability practices that support Web-based design, one should first understand the context in which this field of practice came into existence. Gilbert Cockton observes in Interaction-Design.org's online *Encyclopedia* that, prior to the mass adoption of personal computers (PCs) in the 1980s, using computers was considered to be the domain of experts and required specialized training.[4] Once PCs became ubiquitous, however, the need for user-friendly computer interfaces stimulated the formation of the Human Computer Interaction (HCI) and usability

---

2. Wendy Duff et al., "Archivists' Views of User-Based Evaluation: Benefits, Barriers, and Requirements," *The American Archivist* 71, no. 1 (2008): 144–66.

3. Duff et al., "Archivists' Views," 158.

4. Gilbert Cockton, "Usability Evaluation," *Encyclopedia of Human-Computer Interaction*, available online at www.interaction-design.org/encyclopedia/usability_evaluation.html.

testing fields.[5] Since the 1990s, a significant body of literature has developed around the practice of testing and assessing Web-based resources and systems. In *Usability Engineering*, Jakob Nielsen, one of the pioneers of usability testing methodology, states that "so many things sometimes make an interface good and sometimes make it bad that any detailed advice regarding the end product … makes it close to useless … In contrast, the usability engineering *process* is well established and applies equally to all user interface designs."[6] In other words, if a project team charged with creating a Web resource incorporates a sound and consistent usability testing process during the design phase, it will be more likely to succeed in producing an interface that meets the needs of the intended user base and promotes use of the resource. Nielsen outlines some fundamental steps to usability engineering that include getting to know user populations, analyzing any tasks they have to perform, creating prototypes of the interface during the design period, evaluating the interface through a set of proven design principles, and testing the interface with designated users. These basic usability evaluation methods have led to a pervasive body of practical assessment activities that have been adopted by many practitioners of Web design.

Archives and Library Assessment Activities, Pre-2000

Even prior to the advent of Web-based finding aids, archives had been exploring user studies to help monitor, analyze, and inform the services that they provide. In 1986, a seminal article by Paul Conway represented "a first attempt to structure a comprehensive, profession-wide program of user studies."[7] Predating the development of online finding aids, Conway's study set a framework and methodology for understanding users in the physical premises of the archives. Conway asked such foundational questions as "How good are the services?" He went on to assert that the establishment of such a framework "presupposes that service to users is the foundation of archival programs."[8] Conway proposed ways to assess user information needs at various mediated touch points in a user's visit to an archives, such as registration and orientation. With such a framework in mind, we might well ask how an archives would continue to assess its users once they had moved beyond the visit to the physical archives location to visit its virtual environment on the Web.

In the pre-2000 world of information resources, libraries were the first to make substantial advances into assessing the usability of their electronic resources. Many ear-

---

5. Cockton, "Usability Evaluation."

6. Jakob Nielsen, *Usability Engineering* (San Diego: Academic Press, 1993), 16.

7. Paul Conway, "Facts and Frameworks: An Approach to Studying the Users of Archives," *The American Archivist* 49, no. 4 (1986): 394.

8. Conway, "Facts and Frameworks," 398.

ly development projects were guided by outside HCI consultants. One such project involved the Library of Congress (LC) and its National Digital Library Program.[9] In the mid-1990s, LC brought in an outside consulting team from the Human Computer Interaction Laboratory (HCIL) at the University of Maryland, which built a plan to test designs of a Web interface allowing users to search, browse, and view materials from LC collections. Users reviewed these designs in a group setting at three different points during the design phase of the project. The project subsequently served as an example of how HCI assessment techniques could be used to understand how library users behave in an online environment and how to phase assessments in an iterative way to formulate a progressive design process.

Settling the Wild West: Establishing Evaluation Methods for Archives, Early to Middle 2000s

In the late 1990s and early 2000s, archives took their first steps into the new century by adopting the new EAD standard for electronic finding aids. By 2001, a number of articles discussed the implementation and deployment of EAD. For instance, James Roth surveyed a number of institutions to gauge the adoption rate of the new standard. Roth's study included some lightweight user feedback gathered through anecdotal reporting and some Web log tracking; however, the main focus of the study was to ascertain uptake of the new standard rather than formal usability studies.[10]

The first forays into usability testing of online finding aids showed archivists feeling their way through various techniques and learning lessons about what worked and what could be improved for future assessments. One example was the assessment methodology chosen for the Polaris Finding Aid project.[11] The project team drew on practitioner advice from usability pioneer Jakob Nielsen and paid particular attention to testing with "representatives of your audience not colleagues."[12] The test technique involved sending out questionnaires to individual users to complete on their own. Another early method used in this study involved log analysis of Web traffic. Study authors Burt Altman and John Nemmers mention that further assessment activities were planned for post-release of the online resource that would involve more high-touch methods such as interviews with walk-in users.[13]

---

9.   Catherine Plaisant et al., "Bringing Treasures to the Surface: Iterative Design for the Library of Congress National Digital Library Program," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York: Association for Computing Machinery, 1997), 518–25, also available online at http://old.sigchi.org/chi97/proceedings/briefing/cp.htm.

10.   James M. Roth, "Serving Up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids," *The American Archivist* 64, no. 2 (2001): 230–31.

11.   Burt Altman and John R. Nemmers, "The Usability of On-line Archival Resources: The Polaris Project Finding Aid," *The American Archivist* 64, no. 2 (2001): 121–31.

12.   Altman and Nemmers, "Usability of On-line Archival Resources," 125.

13.   Altman and Nemmers, "Usability of On-line Archival Resources," 128.

Christopher Prom, in his *EAD Cookbook: A Survey and Usability Study*, discussed assessment activities centered on archivist-based inspection of several online finding aid Web sites.[14] These early studies represented efforts by the archival community to develop an understanding of which access components were necessary to effectively present their finding aids online. Recommendations from these usability activities identified overarching requirements, such as the need for clear navigation and search functionality.

By 2003, archives-focused user assessments revealed a desire to examine different user groups and their research behavior patterns in relation to archives and their online materials. Various studies focused on such distinct groups as genealogists, historians, K–12 teachers, and novice users. The methods employed by these studies involved deeper engagement with the user, such as gathering qualitative data from hour-long interviews. The desire to learn about a researcher's workflows and expectations would help to shape future design strategies for online finding aids. For example, Wendy Duff and Catherine Johnson's study of genealogists showed that the way archivists organized collections went against how genealogists tended to conduct their research.[15]

As the decade progressed, a solid understanding of the basic features needed for the optimal display of finding aids was established. The methods for user testing employed by archives grew more refined and focused on appreciating users' "pain points" when interacting with the various areas of an online finding aid. By 2004, Prom had captured the benefits of task-based testing methods and how it informed a more detailed understanding of the nuances of online archival description. His methodology was founded on having a very specific scope or objective to testing. One particular test focused on users' ability to find collections and their ability to search folders within a finding aid. The narrow focus of the test allowed for gathering data that would address specific user needs and provide even more granular design recommendations for the various archival Web sites tested in the project. Prom's pioneering study showcased one of the more sophisticated test procedures to date in the archival assessment field. Users were tested both remotely via questionnaire and in person through observational methods. The most distinctive aspect of this project involved having test observers capture many nuances of user behavior such as mouse movement, keystrokes, and even emotional reactions.[16] This was particularly significant because it represented one of the first mentions of

---

14. Christopher J. Prom, "The EAD Cookbook: A Survey and Usability Study," *The American Archivist* 65, no. 2 (2002): 257–75.

15. Wendy M. Duff and Catherine A. Johnson, "Where Is the List with All the Names? Information-Seeking Behavior of Genealogists," *The American Archivist* 66, no. 1 (2003): 79–95.

16. Christopher J. Prom, "User Interactions with Electronic Finding Aids in a Controlled Setting," *The American Archivist* 67, no. 2 (2004): 242.

such detailed observational techniques in relation to an archival usability test.

The Beginnings of the Gold Rush: User Testing and Assessment for Archival Web Resources, 2007–2012

By the mid- to late 2000s, many major institutions and consortia were embarking on regular assessment activities to inform design, development, and ongoing maintenance of their online primary source collections and archival descriptions. These institutions and consortia tended to create collaborative groups or establish core staff to perform these assessment activities. Some examples of this activity can be found at the California Digital Library (CDL) and the Northwest Digital Archives (NWDA). CDL, the digital library for the University of California system, has created and maintained two primary source Web resources: the Online Archive of California (OAC), the main repository for electronic finding aids from more than 200 institutions throughout California; and Calisphere, a repository for many of the digital objects that belong to collections described by OAC finding aids. Both Web resources were informed by and designed with substantial user study and testing. Many internal reports have been published on the CDL Web site for the benefit of the larger archival community.[17] NWDA, a consortial online resource for accessing archival descriptions of primary source and special collection materials across five Northwestern states (Alaska, Idaho, Montana, Oregon, and Washington) also formed a collaborative usability group composed of staff from many of its participating institutions. This group has maintained an online wiki that contains many of the internal reports documenting the assessment work that has been conducted to continually enhance the usability of the online service.[18] NWDA continues to explore new and more usable ways of displaying and discovering their archival and manuscript collections with their upcoming Cross-Search and Context Utility (XCU) interface.[19]

As commercial services have continued to make their way online in the new century, a growing industry of online assessment and Web site monitoring tools have come into existence. The old labor-intensive methods for capturing and analyzing Web logs have become transformed with a new generation of Web analytics tools that automatically conduct the analysis and reporting of usage for the Web resource owner. Archivists have started to experiment with these Web analytics tools to aid in tracking usage of their Web sites and reveal user behavior patterns

---

17. The Regents of the University of California, California Digital Library, "OAC Redesign Project," available online at www.cdlib.org/services/dsc/projects/oac_redesign.html.

18. Orbis Cascade Alliance, Northwest Digital Archives, "NWDA Interface and Usability Design Working Group," available online at www.orbiscascade.org/index/nwda-interface-and-usability-design-working-group.

19. Orbis Cascade Alliance, Northwest Digital Archives, "IMLS National Leadership Grant," available online at www.orbiscascade.org/index/imls-national-leadership-grant.

that can provide insights into design and development decisions for their online collections. In 2011, Christopher Prom reported on the University of Illinois Archives' implementation of Google Analytics and their subsequent findings. Prom was particularly interested in tracking how users entered their Web site. While staff had assumed that users entered the Web site from the homepage, subsequent tracking through Google Analytics revealed that the majority of users entered via a search engine's results link and therefore arrived directly at resources embedded in the middle of the site.[20] Prom's study demonstrated how intricate levels of detail gleaned from Web analytics reports can help determine the kinds of contextualizing information that users need when they land on an archives' Web site.

Assessment Methods

The following assessment methods showcase a variety of techniques and activities that can be integrated throughout the lifecycle of an online archival resource, from the requirements gathering period, through the design and development phase, and culminating in postrelease monitoring. Assessment methods generally fall into two camps: summative and formative. Summative methods are used after a project has been completed or a service decommissioned. The majority of evaluation methods covered here are formative: that is, methods used to generate critical information during the life of a project to inform its further development.[21] The strengths and challenges to each method will be discussed as well as resources needed to conduct some of the activities. Recommendations will be geared toward practical implementation efforts that maximize cost savings, efficiency, and effectiveness. Recommendations are not meant to address research methods found in academic user studies.

On Your Mark, Get Your Goals Ready, Go!

An important starting point for any Web-related assessment project or activity should be defined by setting goals and objectives. Goals and objectives can be developed by brainstorming all of the questions that the implementation team has in regard to the online Web resource project, whether it involves building a new system from the ground up, redesigning a system, or making smaller changes to enhance an existing system. The types of questions will determine which type of assessment method to use or whether a combination of methods is necessary to provide enough evidence to answer all project complexities.

Defining the target audience of a Web resource allows an archives' staff members

---

20. Christopher J. Prom, "Using Web Analytics to Improve Online Access to Archival Resources," *The American Archivist* 74, no. 1 (2011): 170.

21. Wikipedia, "Evaluation," available online at http://en.wikipedia.org/wiki/Evaluation.

to take a more focused approach to their assessment activities. The assessment efforts will be more effective when background information and reactions are gathered from the intended end-users of the Web resource. Deciding to design a Web site for a particular group or groups of users can help determine which features will meet the explicit needs of the designated users. For example, for a user study conducted by the Planets project, an initiative to ensure long-term access to European scientific and cultural digital content, Kellie Snow et al. explain that identifying their user groups and understanding the research interests, context, and motivation of these groups is an essential component in developing content usage and communication channels that fit naturally into the academic users' research lives.[22]

Methods for Gathering Requirements

Once overall goals and objectives have been determined and a target audience defined for a project, requirements for features should be captured, including major functional or display elements. Some projects, such as the Finding Aid Template Redesign Subcommittee for the L. Tom Perry Special Collections at Brigham Young University, begin with a feature wish list driven by the team's interests that is then validated by a follow-up user analysis activity.[23] In other cases, successful requirements of gathering methods focus first on the user's preferences to uncover background context on their workflows and scenarios for using a particular system. A variety of methods can be used to understand a user's preferences and workflow context.

One such method is the **user survey**. Surveys can offer both a quantitative and qualitative glimpse into a user's preferences and priorities. One important strength of a survey is that it can be distributed widely and potentially garner data from a large number of users without having to engage with each user individually. A good survey can gather basic demographic information for a user as well as information about how that user ranks or prioritizes certain features. There are a number of online survey tools available that are easy to use and either free or very economically priced. These tools make it easy to create, distribute, and analyze the results of a survey.

There are, however, also challenges to creating a sound survey. Survey design is an important factor in being able to shape the user data gathered and can become

22. Kellie Snow et al., "Considering the User's Perspective: Research into Usage and Communication of Digital Information," *D-Lib Magazine* 14, nos. 5/6 (May/June 2008), available online at www.dlib.org/dlib/may08/ross/05ross.html.

23. Cory J. Nimer and J. Gordon Daines III, "What Do You Mean It Doesn't Make Sense? Redesigning Finding Aids from the User's Perspective," *Journal of Archival Organization* 6, no. 4 (2008): 223.

quite involved. Laurie Bridges and Tiah Edmunson-Morton describe how a mixture of open-ended and close-ended questions help to gather a well-rounded view of an undergraduate's image-searching behavior.[24] Open-ended questions can serve to uncover user behavior and motivations in their own words, while close-ended questions can offer users a chance to express how they would prioritize or rate a list of preferred features. Thus, survey design requires thought, preparation, familiarity with question formats, and time to compose the questions as well as time to interpret the responses. Another challenge is getting users to respond to surveys and overcoming what many have called "survey fatigue." One UK study has indicated that it may be difficult to get responses to a survey, especially if the survey is sent via mail or e-mail.[25] Reminders and incentives may need to be incorporated and planned in the survey distribution process. Survey responses, even to the open-ended questions, may sometimes leave the project team with only a part of the picture. Following up with one-on-one interviews may be necessary to supplement survey findings and provide a "deep dive" into user motivation and thus validate or elucidate more cursory findings from a survey. Nevertheless, with sufficient time and skill, effective surveys can be conducted and they can be a good way to gather feature requirements from a large body of potential users. CDL conducted such a survey with both open-ended and close-ended questions during the requirements gathering period for its Web site redesign project. The survey enabled the project team to canvass users from all UC campuses without having to travel to each individual campus.

Structured interviews are another method that can be extremely helpful in gathering requirements for a Web design project, especially if the interview has an observational component incorporated within the interview structure. Observations have the potential to provide important insight into a user's habits and behaviors. Conducting a structured interview requires generating a list of questions to guide the process as well as recruiting users who are willing to spend a dedicated period of time to the process. The questions generally do not require the same level of craftsmanship as the survey instrument, since an interviewer conducting the interaction can always follow up immediately for clarification or additional information. Another component of running an interview is capturing the content of a session, whether that involves having another person take notes or using a recording device. Interviews can also be a very flexible method, with some being conducted in person or remotely via phone or other synchronous communication tool such as video chat. Some challenges involved in conducting interviews can be recruiting participants who are willing to spend a half hour or more of their time. This is another

---

24.   Laurie M. Bridges and Tiah Edmunson-Morton, "Image-Seeking Preferences among Undergraduate Novice Researchers," *Evidence Based Library and Information Practice* 6, no. 1 (2011): 27–28.

25.   Ian G. Anderson, "Are You Being Served? Historians and the Search for Primary Sources," *Archivaria* 58 (2004): 81–129.

case where it may be useful to budget for incentives to recruit participants. The number of participants needed for this method can be small, especially by comparison to survey techniques. Nielsen suggests that as few as five user interviews can gather enough evidence to identify a pattern of behavior.[26] Thus, in considering the pros and cons of this approach, the strengths far outweigh any weaknesses. Interviews are a very effective and efficient way to gather rich data to inform a solid requirements gathering effort.

Focus groups are yet another method that is helpful for requirements gathering. Focus groups usually involve a small group of six to ten users involved in a group discussion revolving around a particular topic. A trained facilitator usually leads the discussion. Rosalie Lack has noted the ability of focus groups to bring user expectations of a service to light since it "is useful for learning what participants think about a particular topic and for uncovering why they think as they do."[27] This method is particularly effective for gathering group impressions on high-level topic areas in a relatively short period of time, usually an hour. While focus groups can help to broaden and explore certain topic areas, Wendy Duff et al. mention that the group setting can influence the thoughts and feelings that group participants express.[28] CDL used focus groups in 2007 during the early stages of the OAC redesign project. The project team visited several UC campuses to review the OAC Web site with groups of archival staff to understand their pain points in using various components of the site, such as the search and finding-aid page navigation. This method proved to be an effective information-gathering technique, because the team was able to see similar feedback patterns across the various focus groups. The focus groups from the OAC project involved groups of coworkers who were familiar with each other; thus, the conversation and discussion flowed easily. However, this method relies on a certain amount of serendipity and dependence on the chemistry of the participants that could impact the quality and depth of the user information gathered. Furthermore, for some Web design projects, when more detailed or nuanced user feedback is needed, individual interviews can provide a depth of concentrated information that focus groups cannot.

Another user-based method to gain context for a user's behaviors is a **cultural probe or diary**. In 2002, Elaine Toms and Wendy Duff experimented with a diary method for understanding how archives users researched material.[29] Overall, they

---

26.   Jakob Nielsen, "How Many Test Users in a Usability Study?" *Jakob Nielsen's Alertbox* blog, June 4, 2012, availability online at www.useit.com/alertbox/number-of-test-users.html.

27.   Rosalie Lack, "The Importance of User-centered Design: Exploring Findings and Methods," *Journal of Archival Organization* 4, nos. 1–2 (2006): 72.

28.   Duff et al., "Archivists' Views," 149.

29.   Elaine G. Toms and Wendy Duff, "'I Spent 1 1/2 Hours Sifting Through One Large Box …': Diaries as Information Behavior of the Archives User: Lessons Learned," *Journal of the American Society for Information Science and Technology* 53, no. 14 (Dec. 2002): 1232–38.

reported that the method was challenging to execute, since recruiting users willing to participate in such a time-consuming activity was difficult. Moreover, even among the small number of participants who signed up for the study, some still did not complete the assignment. Diary methods, while interesting and creative as an assessment technique, tend not to contribute consistently enough for the specific and sometimes time-constrained needs of a Web design project.

A **competitive or comparative analysis** is a requirements-gathering technique that can be used by project teams to formulate initial project lists without having to recruit and engage with users. The analysis usually involves a close inspection of a number of Web sites that serve the same goals and audience needs as the project that is in progress. For their finding aid redesign project, Cory Nimer et al. used this technique to generate a list of functional features that are available on similar finding aid Web sites, such as the OAC, NWDA, the Bentley Historical Library's Polar Bear Expedition Digital Collection, and Denver Public Library's EAD repository.[30] If limited resources and time prohibit more extensive user-based assessments, this method easily generates a list of features that have been vetted and developed on other sites. One can also compare a variety of designs to see how common features have been implemented. For example, if a project team were attempting to redesign a search results page, an informative comparative analysis would be to look across a number of search engines to ascertain which elements were most common in the search results display. However, a limitation of this method could be when there are no comparative design solutions available for a given feature or less than desirable design solutions for a given Web site interaction. In this situation, the project team will have to generate their design solutions with other user-centered methods, such as those covered in the next section.

Methods for Generating Design Solutions

Once a list of features and general requirements has been gathered, there are a number of additional activities that can be conducted to generate design specifications. One oft-used assessment method that generates data suitable for designing the information architecture of a site is an activity known as **card sorting**. This method involves having a user sort a number of cards with various components of the Web site's content or features into groupings. The results of the sorting exercise enable the project team to understand how the user would expect to find content arranged on the site. These data can help the design team create an effective structure for the site with appropriate content "buckets" that would be meaningful to users and allow them to navigate the site according to their natural preferences. This technique involves some preparation time to create the categories

---

30.   Nimer and Daines, "What Do You Mean," 225.

that will be shown on the cards. Also, Nielsen suggests that, to obtain a valid result set, card-sorting exercises should involve at least 15 participants.[31] However, the time invested in preparing and conducting this activity is well spent, since it is an extremely effective way to get concrete user guidance on design specifications for a Web site. These days, card sorting does not need to be a solely in-person endeavor; subscription-based online tools are currently available that allow a user to follow a link and conduct a card sort virtually through a Web interface. Such tools are easy to use and produce the same results as the traditional execution of this method.

Another method that engages the user in the design process is known as **participatory design**. In participatory design, users are given the opportunity to create informal sketches of a page with all of their desired features. This technique allows the project team to gain insight on how users prioritize certain features based on placement and how the user is likely to name a particular feature. Though the design and development team may not translate the user-generated designs literally into the final product, this method still provides useful insights into user-preferred page layouts and meaningful labels while offering another way to validate the overall objectives and requirements of the project design. Yet, if the project team has limited time or resources to spare during the design phase of the project, this technique may not result in conclusive or actionable user data.

Methods for Validating and Fixing Design

Once an initial design has been created for a Web project, some form of user-based evaluation is necessary to test the usability of the site. Sometimes this **usability testing** can occur with static prototypes or simple line drawings of the interface known as wireframes. Programming effort can be costly, so testing prototype interfaces before actual development begins will save time and money over the entire run of the project. The most common form of user testing involves creating a number of tasks that users will perform on the system and having users interact with either a static or interactive prototype of the proposed interface. The data gathered from these tests will help the project team pinpoint where the design is successfully supporting the user's experience on the site and where the design is failing to meet the user's expectations.

Usability testing involves a certain amount of preparation time. Team members will have to prepare a list of tasks that they would like the user to perform. If the testing is being conducted synchronously with the user, then a facilitator will need to be designated to lead the user through the series of tasks. Successfully leading a usability test requires practice and preparation. The test itself can be conducted in a number of ways, with users being tested in person in a usability lab setup or

---

31. Nielsen, "How Many Test Users."

remotely on their own computer through a screen-sharing application. With the advent of screen-sharing applications, remote testing has opened the door for development teams to conduct testing with users in a more cost-effective way. In remote testing, no travel is involved for either user or test facilitator; and, beyond the cost of the application, no additional equipment needs to be purchased. Many of these screen-sharing applications are part of a telephone or Web conferencing system and require only an Internet connection and browser with appropriate plug-in. Additionally, screen-sharing applications usually include a recording mechanism that can capture a user's audio response in addition to video captures of the user's mouse movements. Provided permission has been requested from and granted by the user, recordings can be shown to the entire project team so that all team members have a firsthand understanding of the user's reaction to the interface. Live and recorded usability tests are a highly effective way to evaluate interface design. Any time spent preparing for this method is far outweighed by the benefits of being able to test and refine an interface before it is committed to code.

Over the past decade, a number of commercial tools and services have emerged to aid Web designers in testing their designs. These services can range from free to an annual fee-based subscription. Online usability tools are used in a way similar to surveys where the experience is mainly unmediated and users can generally take the test and interact with the interface on their own. Besides being able to set user tasks and ask questions, some tools have specialty functions such as heat maps that display where users are clicking on the Web page. One of the weaknesses of such unmediated methods, similar to survey methods, is a lack of ability to follow up with the user to ascertain motivation or emotional reaction to the design.

Iterative design and user testing is a highly productive variation on traditional usability testing. An iterative process usually involves a smaller number of initial tests with design adjustments made between those tests to evaluate the effectiveness of those adjustments. One method involving iterative design and testing is called the **RITE method** (for "Rapid Iterative Testing and Evaluation"). Following the RITE method, changes are made to the prototype after testing with one or two users. Once these changes have been made, the revised prototypes are tested with another one or two users. According to Michael Medlock et al., the RITE method allows a project team to make a high number of fixes immediately and to validate those initial fixes with a follow-up round of testing. Some weaknesses of this method include the small sampling of users. Such a small sampling cannot be said to be a true representative of the user population.[32] However, as noted earlier, Nielsen has

---

32.   Michael C. Medlock et al., "Using the RITE Method to Improve Products: A Definition and a Case Study," Usability Professionals Association, Orlando, Fla., July 2002, available online at www.microsoft. com/en-us/download/details.aspx?id=20940.

suggested that a pattern of use can be established after sampling only five users.[33] Other weaknesses of this method surface when the project team does not have a sufficient understanding of the user's mental model for the interactive situation at hand, and the faster usability testing cycle only reveals surface usability issues with the interface.[34] Conducting a deeper user investigation prior to development of the new interface could mitigate the drawbacks of this method.

Inspection methods offer another easy design validation process that both is lightweight and does not involve user recruitment. An example of an inspection method is a heuristic evaluation that involves one person walking through and evaluating a Web site based on a series of design principles. Heuristic principles developed by Jakob Nielsen are heavily used. Some examples of these principles include visibility of system status, match between system and the real world, user control and freedom, consistency and standards, recognition rather than recall, and flexibility and efficiency of use.[35] Gerhardt-Powals is another set of heuristic principles that can be used to great effect during the design process. The principles include such concepts as automating unwanted workload, reducing uncertainty, fusing data, presenting new information in meaningful aids to interpretation, use names that are conceptually related to function, and limit data-driven tasks.[36] A heuristic review does not require much preparation or cost to conduct; however, the person conducting the inspection must have some familiarity with Web design concepts and be able to analyze and recognize whether an interface meets the guidelines put forth in any set of heuristic principles. This method is also good at unveiling major problem areas that need to be addressed. Thus, if a project team has a tight delivery timeline and cannot spend time on recruiting users and conducting tests, an inspection method is one of the most efficient ways to conduct some lightweight design validation of the interface.

Ongoing User Feedback Channels

For Web resources that are already in production, ongoing assessment can be conducted through a variety of user feedback channels that are relatively low cost. Some examples of user feedback channels include usage statistics, customer support data, Web 2.0 comments, and remote reference inquiries.[37] Customer support

---

33.   Nielsen, "How Many Test Users."

34.   Matthew Sharritt, "What's Wrong with the RITE method?" *Situated Research* blog, April 30, 2010, available online at www.situatedresearch.com/blog/2010/04/whats-wrong-with-the-rite-method/.

35.   Jakob Nielsen, "Ten Usability Heuristics," available online at www.useit.com/papers/heuristic/heuristic_list.html.

36.   Usability.gov, "Heuristic Evaluations," available online at www.usability.gov/methods/test_refine/heuristic.html.

37.   Molly E. Bragg, "There Is Always More That Can Be Done: A Survey to Investigate Libraries' Measurement of Digitized Primary Source Use" (master's thesis, University of North Carolina at Chapel Hill, 2011), 25.

data, remote reference inquiries, as well as Web 2.0 comments tend to be more anecdotal; but, if quite a few requests come in pertaining to similar user problems, then an identifiable pattern can be established without user testing. Overall, these anecdotal methods are good at filling in the picture of a user's major problems with an interface, but they do not provide a comprehensive foundational body of user data regarding a Web site's usefulness and value.

One growing trend with respect to comprehensive, user-generated data is the automated analysis of Web site usage statistics. Currently, there are many Web analytics packages available, including many free services, such as Google Analytics, that provide detailed data about users and their visits to a Web site. Analytics tools provide page counts as well as visitor information on browsers used, screen resolution, user location, and much more. Analytics provides a window into the path a user takes on the site. It shows if they entered the site through an item- or collection-level page or the home page. One thing that analytics cannot disclose is the user's role—whether, for example, the user is a student or professor—nor does it reveal user motivation and or explain user behaviors on a Web site. For that additional layer of information, design teams will need to conduct one of the higher-touch user engagement methods and potentially add a survey or registration mechanism to the Web site so that user identity or profile characteristics can be connected to Web site usage patterns. These analytics-related activities take varying degrees of programming resources and time to implement. For instance, Google Analytics requires only that a snippet of code be appended to Web pages for tracking purposes. Likewise, varying amounts of time are required to decipher analytics data and to analyze and interpret it. Nevertheless, analytics as a method of information gathering can be very comprehensive and provides a level of detail and accuracy that cannot be conveyed with any of the other methods.

Complex Objectives and Multiple Methods

For the majority of Web design projects, one assessment method alone cannot provide enough information to design and validate user interactions with a complicated access and discovery Web site. The best projects involve a combination of different methods to generate a more comprehensive set of interface design recommendations. For projects whose goals cannot be streamlined to a smaller number of discrete objectives, or if research questions range across several different user groups, a number of different methods will need to be used to zero in on the optimal configuration of project features and design options. During the OAC redesign project, three different user groups were targeted for the redesign efforts: archivists who needed to search quickly over collections belonging to their own institution, expert researchers who needed to discover collections that were pertinent to their research interests, and novice researchers who were not familiar with the

concept of online finding aids and archival research. A combination of assessment methods including focus groups with archivists, one-on-one interviews with expert and novice researchers, and a comparative analysis of similar aggregated finding aid databases was used to form a comprehensive view of all user groups' behaviors and preferences.

The best assessment practices sample at least one method from each of the major assessment categories reviewed above. For instance, one requirements gathering method coupled with a design exercise and a usability test will go far toward providing a well-rounded assessment process. Stephanie Rosenbaum comments that, from a "practical standpoint, it is simply not realistic to usability test every element of every product in all of its contexts. A sensible model is to include both inspection and empirical research in a product development program, and move back and forth among the methods in a star pattern … with its alternating waves of creative and structuring activities."[38]

Concluding Remarks

The assessment techniques and methods discussed in this article are geared toward providing practical evidence that can be the basis for sound decision-making when selecting new features for development and refining interactions for existing features in the display of complicated and layered content, such as those found in finding aids and other archival and library resources.

In general, most of the costs of implementing the methods discussed center primarily on human resources. A staff member who has strong observational and analytical skills as well as an inclination for learning and assimilating new software applications can be an essential part of any Web design project. Many of the skills necessary to conduct assessment activities can be learned outside formal degree programs; however, preparatory time must still be dedicated to the usability process. Many currently available tools can aid in streamlining some of the more complicated quantitative analysis work that has been part of usability testing processes in the past. Some of the tools also help to mitigate some of the previously mundane aspects of the work, such as note taking to capture what is happening during an assessment.

Given the rate at which new technologies, devices, and applications are being built, we should expect that user behavior, preference, and expectation will likewise

---

38. Stephanie Rosenbaum, Commentary on Gilbert Cockton, "Usability Evaluation," in *Encyclopedia of Human-Computer Interaction*, eds. Mads Soegaard and Rikke Friis Dam (Aarhus, Denmark: The Interaction Design Foundation, 2012), available online at www.interaction-design.org/encyclopedia/usability_evaluation.html.

change over similarly brief intervals of time. Today's finding aid design or object display tool may not meet future needs. To stay on top of these sometimes small and sometimes dramatic shifts in user habits, a continual flow of small-scale assessment activities can be usefully maintained to monitor this progression and help archivists ensure that the means of accessing their primary source materials remains relevant to their user communities.