Data Dissemination to a Web-Based Audience: Managing Usability Testing during the Development Cycle

Michael D. Levi
Project Manager, Data Dissemination Systems
U.S. Bureau of Labor Statistics
Levi_m@bls.gov

Prepared for the United Nations Economic Commission for Europe
Meeting on the Management of Statistical Information Technology
Geneva, Switzerland; February 2001

Introduction

There have been many changes in statistical survey methodology and implementation over the years. Everything from questionnaire design and respondent interviewing procedures to statistical sampling, aggregation and estimation techniques has been revised. Certainly one of the areas where the most rapid change has taken place is in automated data processing, where an evolution from crude automation, to massive centralized mainframe computing, to distributed client-server or three-tier architectures has given everyone from data collectors to analysts to end users unprecedented personalized computing power.

Impact of the World Wide Web

Perhaps no infrastructure change has had as significant an impact, however, as the growth of the Internet and the accompanying quantum increase in the number of data customers who now expect and demand full access to an almost unlimited selection of statistical information. Statistical agencies have, of course, responded to the opportunities and challenges presented by this new data dissemination medium. Whereas in the early 1990's only a trickle of economic statistics were available "on-line" through dial-in bulletin boards and a handful of FTP sites, today every agency of note has a Web site full of the most recent numbers.

The World Wide Web has brought some existing challenges into sharper focus. Long-standing requirements for accurate, timely, and reliable numbers face more demanding scrutiny than ever before:

- When a relatively small number of users obtained data, issuing a correction was an possibly embarrassing but tractable task. When many thousands of anonymous users obtain data, issuing a correction that will reach them all is impossible.

- If most data is disseminated through printed publications distributed by the postal service, a delay of a few hours (or even days) may barely be noticed. When data dissemination is essentially instantaneous – the interval between posting fresh information and the first users accessing it is typically measured in seconds – even the smallest delay gives rise to vocal complaints.

- Instant availability has also brought with it the raised expectation of continual availability. Whereas in the past data users might have had some patience for a busy telephone line, current Web users have no patience at all for an agency Web site that is overloaded, down for maintenance, or otherwise not available all the time. And since the Web audience is global, there are few, if any, times when an interruption of service will not inconvenience some user.

Human-Computer Interaction and Usability Engineering

In addition to the requirements for accuracy, timeliness, and reliability, there is one more essential requirement for data dissemination: The information presented must be intelligible to the intended audience. This, too, poses a significant challenge to all producers of economic statistics. As the audience has grown larger it has also grown much more diverse. In the past, data dissemination vehicles typically assumed a certain level of economic and statistical sophistication. Frequently the user base for a given set of statistics was as knowledgeable of the subject matter as the producers themselves. This is clearly no longer the case. The audience for the economic statistics such as the U.S.

Consumer Price Index or Local Area Unemployment Statistics range from professional economists and policy makers, to interested members of the lay population, to 13-year old students working on school assignments. Economic and statistical literacy can no longer be taken for granted. Yet at the same time no agency would want to "dumb down" its data presentations and alienate the sophisticated users who have been their traditional user base.

Presenting complex data in a form that can meet the differing needs of a highly diverse population is a non-trivial task. What is at stake here are the human factors, or *usability*, of a particular screen design or sequence of screens.

Usability can be defined as the degree to which a given piece of software – including presentation systems such as a Web site – is an effective and helpful tool for the computer user who is trying to accomplish a task, as opposed to being an additional impediment that must be overcome before the task can be successfully completed. The broad goal of usable systems is often assessed using several criteria:

- Ease of learning
- Retention of learning over time
- Speed of task completion
- Error rate
- Subjective user satisfaction

Methodologies for building usable systems have been introduced and refined over the past twenty or so years under the discipline of Human-Computer Interaction (HCI), or usability engineering. HCI principles include an early and consistent focus on end users and their tasks, empirical measurements of system usage, and iterative development.

Like other software engineering methodologies, usability engineering includes requirements gathering, design, implementation, and testing phases. Books and guidelines for building more useful and usable Web sites have begun to appear, and the Web community at large is beginning to accept that flashy graphics and fancy fonts do not necessarily lead to more productive visits by more satisfied users.

Good design guidelines by themselves, however, do not guarantee a usable end product. Usability testing is the process by which the human-computer interaction characteristics of a system are measured, and weaknesses are identified for correction. Such testing can range from rigorously structured to highly informal, from quite expensive to virtually free, and from time-consuming to quick. While the amount of system improvement is often related to the effort invested in usability testing, all of these approaches lead to better systems.

There are two quite different aspects to Web design – (1) the overall site structure and navigation flow between pages, and (2) the design of each individual page. Correspondingly, there are different usability tests to elicit information on these aspects.

Testing Site Structure

The objective in evaluating Web site structure is to determine whether the site mirrors the way users mentally partition the information space so that they can rapidly gain familiarity with the site and find the information they require with relative ease. Two complementary methods are useful for this.

The first is known as a "Card Sort" (Mele *at al* [1997], Levi and Conrad [1996-2]). A group of end users is given a set of randomly ordered index cards. Each card is labeled with a concept from the task domain such as "Consumer Price Index News Release" or "Employment and Earnings Statistical Methodology" or "BLS Contacts", one title per card. The stack of cards is given to a group of users who are asked to sort them into meaningful piles and put a label on each pile.

Several statistical packages have a cluster analysis procedure which can take the sorted cards and aggregate them into a summary hierarchy. If a relatively small sample of subjects is tested, then a simple visual scan and mental aggregation can give much the same insight as a formal cluster analysis. In practice, many people do just this.

A second technique, the "Category Membership Expectations" test (Mele *et al* [1997]), analyses user expectations in the reverse direction. This exercise begins with a set of broad category names, such as "Data", "Programs", or "Publications." Users are then queried regarding what sort of information they would expect to find within each category. Instead of beginning with the details and combining them into a coherent whole, category membership operates from the

higher levels to explore what should reasonably be grouped within each.

Used in conjunction one with the other, a Card Sort and a Category Membership Expectations test becomes much more powerful than when either is used alone.

These two techniques are commonly identified as analysis and design tools, and, in fact, that may well be when they can be most effectively employed. But they can also be quite useful as evaluation tools to validate a design after it has been completed and in determining whether the site creators successfully met their users' expectations.

Testing Page Design

Page design consists of formatting an individual Web page, including page content along with all graphic elements, headers, footers, and embedded links. The design objectives include developing consistent, readable, understandable pages, with common elements identified in a uniform manner in predictable locations, so that data users can find information rapidly and reliably (or determine that the information they seek does not exist on the page and can make a reasonable choice concerning what to do next).

One effective method to test page design is called a "Heuristic Evaluation" (Nielsen and Mack [1994], Levi and Conrad [1996-1]). Heuristic evaluation is an "inspection method" rather than an end-user oriented test. It consists of HCI experts exploring a system, identifying usability problems, and classifying each problem found as a violation of one or more usability principles. Heuristic evaluations readily identify likely usability problems due to inconsistent titles and labels, unintelligible jargon, and confusing layout. They tend to be weak in identifying system-wide structural problems.

Nielsen and Mack [1994] describe seven other inspection methods in addition to heuristic evaluation. These include cognitive walkthroughs, guideline reviews, pluralistic walkthroughs, consistency inspections, standards inspections, formal usability inspections, and features inspections. What all have in common is having HCI experts, rather than end users, go though a structured process to identify usability weaknesses.

Empirical testing of end-users can also be very effective in evaluating page design. Users are shown sample HTML pages, either online or as a paper prototype, and asked to find specific information or navigation elements. This can be particularly effective when the text on a given page is replaced with random alphabetical letters or symbols that preserve the page formatting but are meaningless in themselves. This removes one major clue as far as the user is concerned, and highlights how effectively the layout and graphics communicate without accompanying language.

One final examination of graphical elements is an "Icon Recognition" test (Mele *at al* [1997]). Here the evaluators produce a number of different possible icons or graphics to represent portions of the site and ask the user to match an icon with a category. Icons with high recognition (where users consistently identify the graphic with the desired category) and low interference (where the icon is not identified with more than one category) are required for effective communication with the user population.

Testing Site Usage

The above set of structural and page evaluation methods examine elements of the site in isolation and try to determine optimal solutions for each element. There are also usability testing methods that examine a site as a whole (or specified subsite) to evaluate how well everything works together.

A scenario-based usability test (Dumas and Redish [1994], Rubin [1994], Levi and Conrad [1996-2]) involves presenting representative end-users with scenarios, or specific tasks, designed to cover the major functionality of the software system and to simulate expected real-life usage patterns. Such scenarios should be formulated by knowledgeable task experts in consultation with the system designers. Results are then tabulated using such measures as whether the participants correctly accomplished the tasks, the time taken for each task, and the number of pages accessed for each task.

There is a variety of ways that the interesting information can be saved for analysis: think aloud protocols, the Web server logs, audio or video taped sessions. The key is to observe and record what representative users are doing on the system.

Many large organizations have invested heavily in fully equipped usability labs staffed by experienced professionals. Companies such as Apple and Microsoft routinely subject new software to a battery of usability tests. Even smaller organizations and government agencies can equip a usability lab at modest expense. This might include a one-way mirror for observation of subjects and video equipment to capture user sessions for later analysis or presentation. Usability testing need not involve a laboratory, however, nor need it be expensive nor require an army of usability professionals. Meaningful tests be effectively organized and run by educated lay people; results can be captured quite well with paper and pencil.

A case in point involves running usability tests at remote sites. Conducting tests at end users' work sites, or even setting up a booth at a conference or trade show, can be more realistic than the sometimes artificial setting of a lab. Several companies offer "portable usability labs" for this purpose, but it can also be quite effective to simply set up a workstation and run interested passers-by through a few scenarios on a prototype system. If the right venue is chosen, these subjects often better represent the target population than any test users recruited to the developers home location. What's more, workers in their offices or conference attendees are commonly curious about the development process and motivated to help determine improved system features and approaches.

In addition to evaluating 'hard' measures like task speed and error rates, it is extremely useful to investigate the less observable aspects of interface design that cumulatively (and often subtly) contribute to users' subjective feelings of satisfaction or frustration.

Some instruments that have been developed to meet this need are the Software Usability Measurement Inventory (SUMI) and Web-site Analysis and MeasureMent Inventory (WAMMI) , developed by the Human Factors Research Group within University College Cork, Ireland, and the Questionnaire for User Interaction Satisfaction (QUIS), developed by the Human-Computer Interaction Laboratory at the University of Maryland in the United States. Designed to provide reliable and consistent cross-platform and cross-application satisfaction measures, these questionnaires ask participants about a variety of factors that assess user satisfaction.

Finally, usability evaluation need not end with a system's public release. Standard Web server logs are an invaluable source of information about usage patterns once a Web site has gone live. At this point the testers need not find usability experts or representative users; real users' sessions are captured in great detail and are available for analysis.

For example, Web logs typically give full details of every text string entered into a site's search engine. An analysis might separate the user sessions which begin with a search from those sessions where the search comes only after many pages have been accessed. The latter category might represent a failure of the site's organization – users can not find what they are looking for by traversing the hyperlinks, and so fall back on a search capability. When the logs show consistent patterns of this nature, it may be time to rethink the page hierarchy.

The advantage of using Web logs is that they capture real users going about their tasks. The weaknesses of using these logs is that they pose a huge data reduction task, the users' goals can usually only be guessed (though search strings may provide strong clues), and there is typically no way to query the users as to what they really were looking for.

Management concerns

Software project managers typically have a great deal on their minds. Adding usability engineering to the mix simplifies some of the back-end activities (a well-designed site will need to be revised less often, will generate fewer inquiries to the help desk, may place a reduced load on servers and networks, and generates fewer irate complaints to upper management) but does add up-front complexity and cost to development.

The resources required to effectively implement usability engineering into a Web site development effort fall into three main categories: staff, time, and money.

Clearly a capable, trained staff is necessary to properly implement an HCI approach to development. Ideally every developer will be knowledgeable of design guidelines and experienced in implementing them, and a group of usability professionals will be available to conduct meaningful evaluations. Some organizations may already have a pool of evaluators to draw upon. Any statistical agency that is accustomed to evaluating

and field testing questionnaire design has the necessary resources and culture to perform systems usability analysis.

Failing this, however, the experience of many organizations is that usability engineering can be implemented gradually, starting small with a core of interested personnel and expanding as the efforts demonstrate their usefulness. There are sufficient reference materials available, as well as industry-based courses and academic programs, to begin.

Initially, usability engineering efforts will slow down development. Time must be factored into the schedule for another series of tests (in addition to traditional systems-oriented tests), and time must also be factored in to make corrections to the system based on test results. But just as rigorous systems testing before deployment saves defect removal time after deployment, so thorough usability testing early on will point to improvements that can avoid costly future re-engineering efforts.

There is also a great deal of flexibility in scheduling. Many tests can be run in parallel to other development activities or to each other. Sometimes a relatively quick test will produce results sufficient to identify and correct the worst usability defects without requiring a longer and more comprehensive evaluation. Often a developer can spot problems after one or two subjects have been tested, and begin to correct these even before formal test results are generated.

The cost of a usability program can range from nothing except staff time to tens or hundreds of thousands of dollars for a fully equipped, sophisticated usability laboratory. Again, the experience of many organizations is that a usability engineering physical plant can be implemented gradually. Many tests require a only a meeting room, paper, and pencil. Over time an organization might add recording gear and observation areas.

Relevant to the expenses of outfitting a laboratory should be the potential cost savings that can be realized through usability engineering. The purpose of most Web sites is to attract users and distribute information or products. Losing users because of a poor design could be catastrophic for a commercial venture. Even in the absence of direct monetary considerations, an organization will find the cost of user support – such as calls or e-mail to a help desk –declines as ease of use is enhanced.

When a project manager decides to go forward with a usability engineering or testing program, one final hurdle typically remains: how to get interest, motivation, and buy-in from project sponsors and system developers. The author's experience, along with anecdotal evidence from colleagues in other organizations, suggests that there is no substitute for direct experience. A verbal description of the benefits of usability testing will intrigue some people. But watching – either live or on video tape – actual users make unanticipated mistakes while using a system is an eye-opening experience. After their first exposure to this style of testing, a large percentage of developers wonder how they ever built software without it.

Conclusion

For many, if not most, statistical agencies, the World Wide Web has become the major vehicle to disseminate economic data to its customers, and has become the primary points of contact between a given organization and its user base. For many users this system will be the only grounds on which they can judge the organization. Hundreds of thousands of users will obtain mission-critical data from this source. Ease of learning, ease of use, and general user satisfaction, along with quality and comprehensiveness of content and functional capabilities, will determine the success or failure of the effort.

There are many methods for usability testing. This paper has certainly not exhausted the list of possible methods; developers and researchers continue to experiment with new techniques. Many of these methods are reasonably easy, reasonably fast, and reasonably cheap. Best of all, they are not intimidating for either participants or testers.

Ultimately, the only real way to begin usability analysis in an organization is to take a deep breath and simply start, trusting that the details will fall into place over time.

Some Useful Resources

A Practical Guide to Usability Testing. Joseph Dumas and Janice Redish. (1994) Ablex Publishing Corp.

A Heuristic Evaluation of a World Wide Web Prototype. Michael Levi and Frederick Conrad. July/August 1996 *interactions Magazine;* also http://stats.bls.gov/ore/htm_papers/st960160.htm

Usability Testing of World Wide Web Sites. Michael Levi and Frederick Conrad. (1996) http://stats.bls.gov/ore/htm_papers/st960150.htm

Evaluating Web Site Structure: A Set of Techniques. Kathy Frederickson-Mele, Michael Levi, and Frederick Conrad. (1997) Conference Proceedings of the Usability Professionals' Association; also http://stats.bls.gov/ore/htm_papers/st970070.htm

Usability Inspection Methods. Jakob Nielsen and Robert Mack, eds. (1994) John Wiley and Sons, Inc.

Usability Engineering. Jakob Nielsen. (1993) Academic Press, Inc.

Designing Web Usability. Jakob Nielsen. (2000) New Riders Publishing.

Handbook of Usability Testing. Jeffrey Rubin. (1994) John Wiley and Sons, Inc.

Designing the User Interface, 3rd Edition. Ben Shneiderman. (1998) Addison Wesley Longman, Inc.

Questionnaire for User Interaction Satisfaction Home Page. University of Maryland at College Park. http://lap.umd.edu/q7/quis.html

Software Usability Measurement Inventory. University College Cork. http://www.ucc.ie/hfrg/questionnaires/sumi/index.html