

USABILITY TESTING OF DATA ACCESS TOOLS

John J. Bosley and Frederick G. Conrad

BLS, 2 Massachusetts Ave., N.E., Rm. 4915, Washington, DC 20212

ABSTRACT

Thanks to the World Wide Web (WWW), the public now has access to statistical data produced by many public and private organizations. As a result, there are potential data users who have little experience working with statistical information. Non-expert users find it difficult to cope with the great quantity and variety of data that are available on line, as well as with the specialized technical terms (metadata) used to describe the data. Many producers of statistical information make on-line software tools available to support users' efforts to find the information they want. This paper reviews usability test findings for three tools that the U.S. Bureau of Labor Statistics and the U.S. Bureau of Census offer users. These three tools collectively contain a wide range of features and functions, and the usability tests employed a range of methods to cover various aspects of tool use. The paper employs a model of a generic data access task to integrate the findings and help generalize them. It discusses three types of usability concerns that were evident across the tests. First, a tool's usability depends on including sufficient guidance and instructions for using it. Second, the tool should avoid unnecessary complexity and feature "clutter." Third, usable data access tools must enable users to overcome deficiencies in the way data sets are named and give users some understanding of how statistical databases are organized. The paper provides specific instances of each problem class, and proposes some ways that data access tool designers can avoid or correct these types of usability problems.

1. INTRODUCTION

Citizens can search for and (hopefully) retrieve an enormous amount of statistical data through the World Wide Web (WWW) from many public and private organizations. Data vary widely from site to site in content, quality, timeliness, and many other attributes. Web access opens up these rich data resources to many potential users who may know very little about data, and thus are unprepared to cope with the volume and diversity of the information suddenly available. Several federal agencies in the United States have created software "tools" to help persons trying to locate and use data via the Web.

1.1 Scope of this Paper

This paper reports on usability tests that the authors and colleagues at BLS and the Census Bureau performed to evaluate some of these agencies' data access tools. The goal of the research, of course, was not simply to uncover usability problems but to inform the design of more usable versions of the tools.

We tested the usability of three tools by having a sample of users search for and retrieve statistical data as described in several scenarios.

- A Census Bureau HTML tool enables users to find and download survey data on various social, economic and health topics collected by the Census Bureau and other agencies. This tool had been released for public use prior to the usability tests we conducted. The HTML version that was tested has since been replaced by a Java implementation.
- A BLS Java data query tool that gave users access to more than 50,000 labor force statistical time series collected by BLS. "Time series" contain data sets collected periodically (monthly, quarterly, etc.) and cumulated over many periods. Different series include different measures of labor force characteristics, such as unemployment rate or unemployment level. Different sets of demographic factors such as age, gender, and race characterize different time series as well. The test used a prototype.
- A BLS Java data query tool that gives users access to current hourly wage rates (dollar amounts) for a large number of occupations, occupational classes, and skill levels within specific metropolitan areas. This too was a prototype.

The user interfaces to these three tools were quite different, reflecting the different kinds and organizations of the underlying databases. However, from the user's perspective, the task of finding and accessing data is essentially the same. A common conception of the data users' task can help developers achieve more uniform user interfaces that more closely match the users' understanding of the data search and retrieval process. Users could learn to use such data access tools more easily and transfer more of their data access skills from one tool to the next.

1.2 Conceptual Framework: A Generalized Data Access Task

In general, users gain access to data by selecting characteristics of the available data until just the relevant subset of data has been located, and then by submitting a request for these data. More specifically, users perform three subtasks. (Levi and Conrad, 1999). First, they must specify the attributes of the data of interest with the help of the tool, and send this description to the appropriate database. Ideally, the tool then transparently formulates a query from the user's specifications that the "back end" database software can interpret. The database should send the query results back to the tool, which then displays a set of descriptions (labels, names, etc.) of data elements or datasets that match the user's data specifications.

Second, the user evaluates the candidate descriptions (e.g. data series names) on the basis of how well they fit her or his expectations, and chooses the items that best match the user's concept of the data most relevant to the user's goal(s). None of the descriptions returned may appear sufficiently relevant but if one or more does, the user requests the data thus described. If he or she makes a request, the system retrieves data from the database. Third, the user examines the actual data to determine if they fulfill her goal(s). If unsatisfied with either the candidate descriptions or the actual data, the user may return to the first step to submit a different query. If the user gets no better results, he or she will eventually quit. The tests reported here included little attention to users' evaluations of actual data but future research may focus on tool usability from this standpoint.

2. METHODOLOGY

Two of the three tests covered in this paper were conducted in BLS' usability laboratory, with a browser installed on a standard desktop workstation. The prototype wage locator was tested in the field at a state convention of corporate and union compensation specialists who will be frequent users of the production version of this tool. The prototype was loaded on laptop computers for this remote test.

These tests involved small groups of from 5 to 10 users, although the Census tool was tested repeatedly and so more than 15 users were involved. Except for the wage locator test, participants were recruited from academia, government, the press, and private business. Most were somewhat familiar with using statistical information, but the test team intentionally included some users who were unfamiliar with getting data online using the WWW. Users performed scripted tasks with specific data targets so that the task could be judged on success or failure. In a few cases users were invited to use the tool to look for personally interesting data, so that success or failure in these cases depended on the participant's self-report, not an objective determination.

Although it was usually possible to quantify the users' performance in terms of successful completion of a task script, usability problems were also identified by observing test sessions and asking for user feedback after the sessions. Multiple test observers discussed their perceptions of user problems with each tool to reach consensus on the nature of the problem and its source in tool design. They then communicated a summary of observed problems to tool developers, and in some cases worked with the developers to improve the design.

3. RESULTS

Three major types of problems showed up in all three usability tests. There were numerous additional minor usability problems specific to a given tool, but this section focuses on just the three major types that were evident in all of the tests. These pervasive problem types are:

1. Insufficient guidance and instructions for using the tool
2. Complex or "cluttered" screen layout
3. Deficiencies in data sets labeling and lack of information indicating how data are organized in a database

A brief sub-section is devoted to illustration and discussion of each of these types of problems and to some possible remedies for each.

3.1 Insufficient Guidance or Instructions for Tool Use

The primary interfaces of all three tools gave users too little guidance on how to use the tool, especially top-level orienting guidance. In all three cases, there was ample room on the screen to display additional instructions for use. The interfaces used common and recognizable "widgets" such as drop-down lists, radio buttons, etc, but lacked adequate guidance about the general context within which the tool is intended for use. The interfaces often left users wondering about basic questions like "What sorts of data does this tool help me find?" "How should I describe the data I'm looking for?" or "Where do I start on this interface in order to use the tool effectively?" Sometimes important instructions were provided through a "help" link, but this de-emphasizes the instructions and makes the user follow the "help" link instead of having immediate access to the information.

Here are some ways to improve the effectiveness of guidance and instructions for using a tool.

- Provide high-level orienting information about the data that the tool works with, such as who collected the data and for what purposes. Don't make users guess or assume that they know much about the data already. Tell them the data topics; if the data are about mortality and age, say so.
- Provide adequate procedural instructions. Something as simple as numbering the necessary steps in a procedural sequence ("1," "2," "3" and so on) takes little space but can greatly increase usability.
- Show as much key guidance on the primary user interface as possible. Provide "help" in small units tied to specific contexts, not in long text segments (Nielsen, 2000). Indicate clearly how to access separate "help" facilities or pages.

3.2 Complex or "Cluttered" Screen Layout

All three tests indicated that data access tools should prominently offer users a small set of "core" features that facilitate quick access to not more than a few data elements. For example, the field usability test of the wage locator tool also showed that a key potential user group—wage analysts—mainly looked for an average hourly wage for just one occupation and one geographic region at a time. The test participants stated that this was not just true for the usability test; it was in fact the way they intended to use the tool in performing actual work. The tool should not display less useful features with equal prominence. Usability testing can also point to unwanted features that can be eliminated to further simplify the screen.

Figure 1 shows the wage locator prototype that was tested. The test results indicated some of its features reduced usability. The list box titled "Select a Level" (lower left) made it possible for users to get a wage rate for different skill levels within an occupation. This field was blank when the application opened. Where a default is lacking users must explore the feature to discover what choices it offers. For this tool, most users did not want to select a level, but rather wanted the "Overall Average Wage" rate. Making that phrase the default choice speeds up data access for many users. Generally defaults should be inclusive, for example, a default for gender data should include both men and women.

The control labeled "Don't know what Level?" actually leads to a "help" screen, and the label should indicate this more explicitly and simply, as discussed in Section 3.1. Finally, the test showed that these users did not need to "Remove" data choices--or indeed to make multiple data choices at all--before retrieving the data. Thus the bottom text box and "Remove" button appear to be unnecessary.

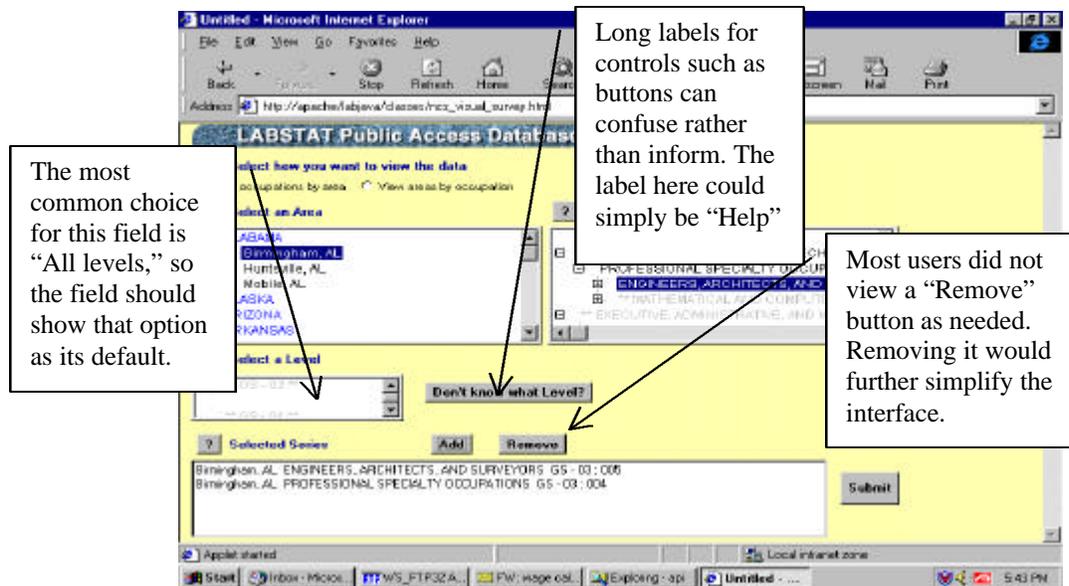


Figure 1—Interface complexity and "clutter"

In Figure 2 in the next section, the different time series labels in the large list box include a common string at the beginning of in every label. This repetition interferes with scanning a list, since all useful distinguishing information is toward the end of the label. It also contributes to making the labels so long that the user must scroll horizontally to view pertinent descriptors, even in a large text box This is a clear usability flaw.

Avoid features that depend on constructing Boolean specifications beyond simple “and-ing”(Sewell and Teitelbaum, 1988). Research shows that more complex Boolean expressions defeat many users (Greene and Devlin, 1990; Spink, Wolfram, Jansen, and Saracevic, 2001). If the tool implicitly performs logical operations, such as automatically “and-ing” individual terms chosen from different lists, users need clear feedback about the result of the operation. It is especially important that the interface shows users whether they are narrowing (“and-ing”) or broadening (“or-ing”) the set of data specified.

Using design guidelines such as the following can help developers avoid creating tools with this type of problem

- Provide primary access to a simple tool with a small set of features that satisfy the needs of the great majority of users prominently displayed. Give less prominence to rarely-used features, and remove features that are virtually unused
- Avoid asking users for Boolean specifications beyond very simple “and-ing.”
- Show defaults in such features as text boxes or pick lists. Use defaults that many users will probably choose.
- Indicate to users that their data specifications will return no results as early as possible.

3.3 Unclear Data Labels and Lack of Cues about Organization of Databases

This problem area has two aspects that are intertwined. (1) Tool users are often expected to understand and use unfamiliar terms from the specialized vocabularies of data or computer experts. (2) Tools fail to give users any cues about how the data are organized in databases. Concerning the first aspect, BLS time series labels often include unfamiliar, specialized terms that make subtle distinctions between data series only expert users will recognize. A prime example is offered by the distinction between data series that include unemployment measures expressed as rates (proportions), and those that are based on unemployment levels (numbers of people). To the non-expert user, the most meaningful part of either data set name is “unemployment.” This user may not even notice the technical distinction between “rate” and “level” and may accept and use either, when only one of the two is the appropriate choice for that user. In the time series query tool shown in Figure 2, this vital distinction is made in the top-most text field on the left of the screen. However, the tool treats these defining measures as simply another “characteristic” of series.

Some data labels make distinctions that are understandable but of little value to many users. Data sets may use unusual or irregular groupings of continuous variables like age or income that are of interest to just a few specialists. For most users, these irregularities simply adds confusion and unneeded complexity to the data access task.

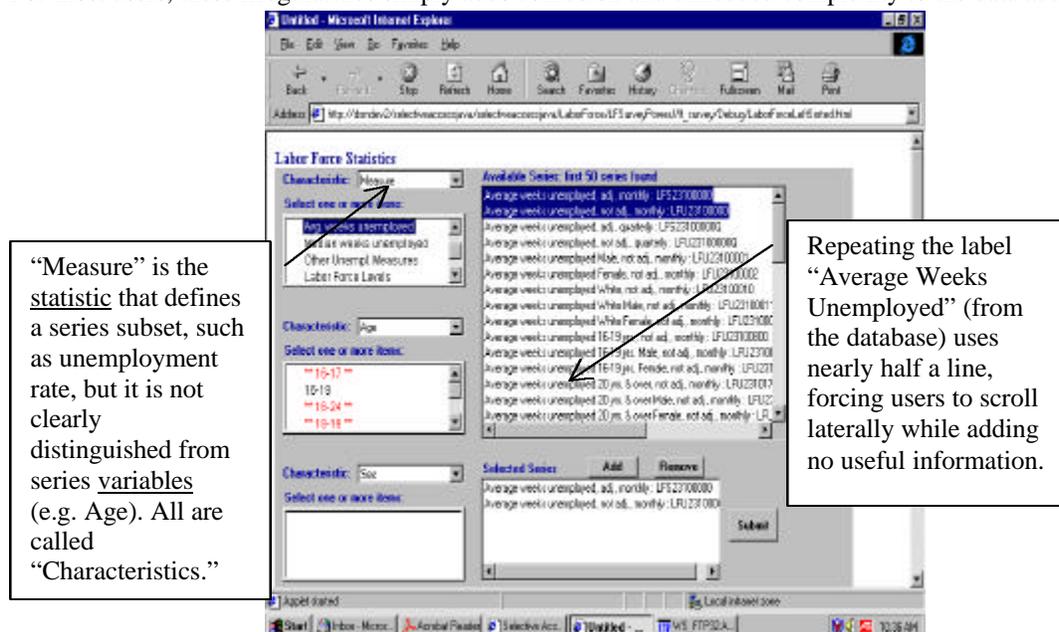


Figure 2. Example—Problems of Data Naming and Organization

Because most existing databases were constructed primarily to serve specialist users inside the producing organization, their structure is opaque to new, relatively unsophisticated users. These users need some cues about the

structures of databases whose content is accessible via the Web. The tools under discussion provided users with minimal information about the databases to which they linked. Lacking such cues, users can become disoriented and use the tools less efficiently. For example, instead of simply listing available data series alphabetically, the designer should cluster related series by topic and provide topic labels. This is consistent with Donald Norman's broad design principle to capitalize on "knowledge in the world," i.e. the interface, not just knowledge in the mind of the user (Norman, 1988, pp. 54-80.)

While there is a limit to how much design can overcome poor data labels and provide good clues to database structure, tool developers can take steps such as these to increase usability.

- Provide cues that enable the user to envision data structures; for example, a labeled tree structure
- Provide easy access to definitions of technical terms, through hyperlinks or mouse-overs
- Indicate clearly what data are missing from the data set. "Gray" names of missing elements, or display them in a different color
- Make it easy to (pre)view the results of data specifications, to verify that these results are appropriate and relevant

4. CONCLUSIONS

The three problem types described in this paper grow out of a flawed process for gathering requirements prior to initial tool design. These tests showed clearly that the tools' designers lacked basic information about the kinds of data access tasks many types of users wished to perform, and what features should therefore be embodied in their tool. Instead, expert data analysts employed by the two agencies had largely set the tools' design specifications. Some specifications were based on staff perceptions of the needs of usually equally expert data users outside of government. However, tool designers gathered little or no information on which to base the design of a more widely usable tool from members of the public who can now find agency data on the Web.

The diversity of web users is so great that it is probably impossible to get a complete profile of all possible data users' needs. But the research reported here indicates an urgent need for greater efforts to learn more about what features data access tools should offer to a greater variety of data users. There is every reason to do the best we can, even though we cannot do everything, in gathering more information about the diverse population of users who visit our statistical websites. At a minimum, designers need to know about what kinds of data this diverse user base is most interested in, and how familiar they are with statistical data generally (Hert and Marchionini, 1997).

Frequent usability testing as the tool design evolves can keep the design relevant to user needs. Usability evaluation of a tool when the tool is ready for deployment can increase the chance that users successfully locate and access the data they are seeking. A tool that is hard to use is more likely to lead users to look for data from alternative, possibly less authoritative—but more usable—sources.

REFERENCES

- Hert, C. A. and Marchionini, G. (1997). "Seeking Statistical Information in Federal Websites: Users, Tasks, Strategies, and Design Recommendations. A Final Report to the Bureau of Labor Statistics." <http://ils.unc.edu/~march/blsreport/mainbls.html>
- Greene, S. L. and Devlin, S. J. "No Ifs, ANDs, or Ors: A Study of Database Querying," *International Journal of Man-Machine Studies*, **32**, 303-326
- Levi, M.D. and Conrad, F. G. (1999). "Interacting with Statistics: Report from a Workshop at CHI 99." *SIGCHI Bulletin*, **31**, 31-35.
- Nielsen, J. (2000). "*Designing Web Usability: The Practice of Simplicity*," Indianapolis, IN: New Riders Publishing.
- Norman, D. (1988). "*The Design of Everyday Things*," New York: Doubleday/Currency.
- Sewell, W. and Teitelbaum, S. "Observations of End-User Online Searching Behavior Over Eleven Years," *Journal of the American Society of Information Science*, **37**(4), 234-245.
- Spink, A., Wolfram, D., Jansen, M.B.J., and Saracevic, T. (2001). "Searching the Web: The Public and Their Queries," *Journal of the American Society of Information Science and Technology*, **52**(3), 226-234.