USABILITY TESTING OF DATA ACCESS TOOLS[1]

John J. Bosley and Frederick G. Conrad, U.S. Bureau of Labor Statistics
John J. Bosley, BLS, 2 Massachusetts Ave., N.E., Rm. 4915, Washington, DC 20212
bosley_j@bls.gov

ABSTRACT

Government statistical agencies and other organizations publish data on the World Wide Web, making it available to users through browser-based tools. As a result, many data users with little experience working with statistical information now have online access to these data. Findings of usability tests of three Bureau of Labor Statistics and Census Bureau data access tools revealed usability problems in three categories: insufficient guidance or instructions for tool use, unclear data naming and database organization, and unsystematic interface organization. Specific instances of each problem class are provided, and methods for resolving these types of usability problems are proposed.

Key Words: World Wide Web, Access to statistical data, Usability

1. INTRODUCTION

Statistical data collected and stored by many public and private organizations can now be accessed through the World Wide Web (WWW). In the case of official statistics, data users access statistical information at the national level (e.g. FedStats in the U.S., the Statistics Canada site in Canada and the National Statistics site in Great Britain), or at the individual agency level (e.g.. the Bureau of Census or the Bureau of Labor Statistics (BLS) sites in the U.S.) More powerful websites to enable users to access statistics simultaneously from various organizations are beginning to appear. For example “The Dataweb” now lets users access data simultaneously from databases at the Bureau of the Census and the National Center for Health Statistics and other agencies. University-based organizations such as the Inter-university Consortium for Political and Social Research (ICPSR) maintain on-line data repositories. Finally, sites operated by commercial firms offer other gateways to statistics on the Web. (URLs of some of these websites appear in the Appendix.)

Any of these sites offers users access to an enormous amount of data. The data are diverse in their conceptual bases, their content, their quality, timeliness, and many other attributes. Web access opens up all of these complex data resources to many potential users who may know very little about <u>data</u> at all, and are unprepared to cope with the volume and complexity of the information suddenly accessible to them without substantial on-line support and guidance. Several U.S. federal agencies have created software “tools” that aim to provide such on-line support to persons trying to locate and use data via the Web. The centralized FedStats website identifies nine such agencies, and some of these provide several tools.

The authors’ informal examination of the tools on this list and identified numerous usability problems, some of which appeared to pose serious barriers for users. On the bright side, several federal agencies, including the Bureau of Labor Statistics, the Census Bureau, the Energy Information Agency and the National Cancer Institute are currently performing usability engineering activities in order to make the tools at their websites more usable and useful (e.g. Levi and Conrad 1997). Nielsen (1993) characterizes usable software as easy to learn, efficient to use, easy to remember, promotes graceful recovery from errors, and is satisfying to use. This paper focuses on some of this usability activity.

1.1 Scope of this Paper

This paper reviews usability tests that the authors and their colleagues at BLS and the Census Bureau have performed to evaluate the extent to which some of these agencies’ data access tools are usable. The tools are displayed in a web browser and users interact with them in order to obtain agency statistics. The goal of the

---

[1] We thank Cathryn Dippo, Carol Hert and Michael Levi for valuable comments on an earlier draft of this paper.
The opinions expressed here are those of the authors and not of the Bureau of Labor Statistics.

research, of course, was not simply to uncover usability problems but to serve as a basis for improving tool usability through design modifications.

The tools tested were:

1. FERRETT ("Federal Electronic Research Review Extraction & Tabulation Tool")—This tool is hosted by the Census Bureau but enables access to survey data at Census and other agencies. It has some capability for on-line data processing, e.g. tabulation as well as a download function.

2. A Java data query tool—This tool helps users access any of the BLS "time series," i.e. data points for a given sub-population collected periodically (monthly, quarterly, etc.) over a period of time. A prototype version of this tool was tested on BLS' collection of labor force time series. This is the agency's largest database of this type, and contains over 50,000 such series. The labor force database was chosen for test purposes because of its size, on the assumption that a tool shown to be usable with such a large database will be robustly usable when it is scaled down to give access to smaller databases.

3. A "wage finder" tool—This tool is in many ways a variant of the Java data query tool that applies to many different datasets. The "wage finder" gives users access to wage data (point estimates of hourly wages) from BLS' National Compensation Survey. Metropolitan area and occupational types as well as specific occupations are used to categorize these data.

Usability tests on FERRETT began after it had been deployed for some time as a CGI application, and continued with a new Java-based version whose design grew out of the usability test program. This paper draws on test results for the original CGI version. Both of the BLS tools underwent usability testing and design improvement based on the tests while still in prototype form and before public release.

The functionality of the various current access tools varies considerably. While some of these differences undoubtedly reflect the influence of real differences in the nature of various organizations' data, the authors propose that, from the user's perspective, the task of specifying and accessing data is, essentially, the same irrespective of such variations among different data sources. This forms the basis for conceptualizing problem types across different tools. If there were some convergence in developers' conception of the users' task, as there is for applications like spreadsheet or word processors, users could learn to use data access tools more easily and transfer their data access skills from one tool to the next.

1.2 Conceptual Framework: A Generalized Data Access Task

This sub-section describes the authors' conception of a generalized data access task. (A relevant workshop report can be found in Levi and Conrad 1999.) Users are essentially narrowing down the available data to obtain a sub-set—often just one data point or series—of interest. Users must perform three component tasks sequentially to reach this goal. First, they must specify the attributes of the data in which they are interested to the extent that the data access tool allows, and send this description to the appropriate database. Any usable tool should transparently formulate a query from the user's specifications that the "back end" database software can interpret. The tool then displays for the user a set of descriptions (labels, names, etc.) of data elements or datasets that match the user's data specifications.

Second, the user must evaluate the candidate descriptions (e.g. data series names) on the basis of how well they fit the given specifications, and choose the best match(es) relative to the task goal(s). The user may decide that none of the returned descriptions are acceptable but, if at least one is chosen, the tool lets the user request the chosen item(s), which are returned. If the user does not choose any she can either return to the first step and reformulate her query or abandon the search. Third, the user examines the actual data to determine if they fulfill her goal(s). If unsatisfied with the results, the user must then decide whether to return to the first step and modify specifications in hope of obtaining better results or abandon the search. The tests reported here included little attention to this third step, but future research may focus on evaluating usability from this standpoint.

2. USABILITY TEST METHODS AND PROCEDURES

This paper is based on three usability tests whose findings are summarized in the next section, along with recommendations about how to avoid the kinds of problems identified in the tests. Two of the tests were conducted at a desktop workstation in the BLS Usability Test facility. This facility places the user in a typical office environment with bookshelves, and a modular desk and table. A personal computer offers reasonably high-speed Internet access. The computer enables web browsing with a late version of either Netscape or Internet Explorer. Normally, the user is given an orientation and task instructions, then left to work alone while test personnel observe user actions on a TV monitor in an adjoining room. Cameras and a scan converter in the work room deliver images of the workstation screen, the user's face, and keyboard and mouse actions to the monitor. It is possible to provide additional guidance, ask questions about apparent problems, and so forth via an audio intercom

The test of the "wage finder" prototype tool was conducted on laptop computers at a hotel in a large southern city. This test took the software to the field to capitalize on access to a user group with high motivation to work with the NCS survey data. These were compensation specialists from both private and public businesses and labor unions who were attending a state convention. Test personnel sat beside participating users at a worktable, observing and guiding the users' performance. No videotaping was possible under these conditions. BLS staff wrote down notes about their observations as the users worked on prescribed access tasks.

These tests have all involved small groups of from 5 to 10 users, although the FERRETT tool was tested on more than one occasion so that the total number of user participants grew to about twenty for that tool. Except for "wage finder" test participants were a diverse group recruited from academia, government, the press, and private business. Most volunteers have been at least moderately familiar with using statistical information, but the test team has attempted to include some users in each group who are unfamiliar with procedures for accessing data on the Web. Users were assigned a range of tasks to perform in the various tests. In most cases, specific data targets have been assigned to users so that the test team can measure users' success. In a few cases users were invited to use the tool to look for data that were personally interesting to them.

We report observational and qualitative results, although in most of these studies we quantify the users' performance, for example by counting how often they access the intended data. In some cases, test sessions were videotaped for later review, but often test data were captured primarily as notes taken by observers during the test sessions coupled with user feedback in debriefing interviews following the individual sessions. Test observers met following one or more sessions and literally "compared notes," discussing their observations of signs of difficulty in using the tool until they could agree on a characterization of the problem and a formulation of its source in tool design. In some cases, these qualitative summaries were simply delivered to the software developers, who then came up with design changes aimed at improving usability. In other cases, usability experts on the team worked along with the developers on making design improvements for the interface. In either case, of course, the developers wrote the code to implement design changes.

3. SUMMARY OF USABILITY TEST FINDINGS

The authors reviewed the summary reports on all three usability tests and abstracted from them a small number of "cross-cutting" findings for inclusion and discussion here. There were numerous relatively minor usability problems that were specific to a given tool, but these were left out of consideration in order to highlight three major usability problem <u>types</u> that were plainly evident in all of the tests. These cross-cutting problem types are:

1. Insufficient guidance or instructions for tool use
2. Unclear data naming and database organization and
3. Lack of organization and feature prioritization.

A brief sub-section discusses each of these types of problems and offers suggestions for overcoming each type. The authors wish to emphasize that the tool designers have welcomed the insights the tests gave into problem areas, and have used the test findings in every case to rework the tool so that usability is now very significantly better for all of them. Sometimes the improvement was confirmed by a repeat test, but often improvement is shown by inspection.

3.1 Problems of Insufficient Guidance or Instructions for Tool Use

Given that the applications tested are characterized as “tools,” it was surprising and disappointing to the test team to find how little guidance the interface gave users on how to apply the tool to the data access task. This problem was particularly evident at the level of overall orienting guidance. Often the interface had clearly recognizable “widgets” such as drop-down lists, radio buttons, etc. and the specification alternatives these represented were shown clearly. The tests showed that inadequate description of the general context within which the tool is embedded was a major source of this problem type. Answers to questions like “What sorts of data can be accessed using this tool?” “How do I describe the data I’m looking for?” or “Where do I start on this interface in order to use the tool most effectively?” were often lacking. Sometimes important instructions were available in the “help” function, but de-emphasized. The link to some useful instructions for using FERRET (Figure 1) is labeled “Hint.” By embedding this link in a terse, inadequate instruction, the design implies that the brief instruction should be sufficient. In fact, the “Hint” clearly offers the user better, clearer guidance.

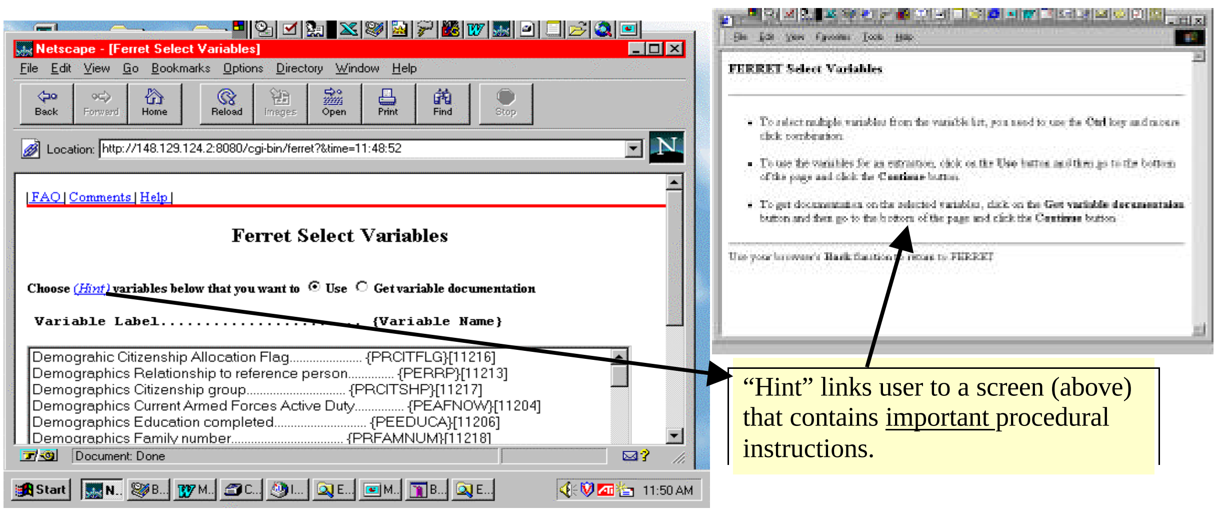

Figure 1. Example—Problems of Inadequate Tool Use Guidance

This type of problem can be alleviated if tool developers adhere to guidelines such as:

- Don’t assume that all users who find the site know much about the statistics available there. Provide high-level orienting information about the data the tool is linked to, especially contextual information such as who collects these data and for what purposes. Tell them the topical domain of the data – if the data are about mortality, for example, then say so. Don’t make users guess.
- Provide adequate specific instructions as well. Often tools have a surprising amount of unoccupied “screen real estate” available for more extensive guidance. On the other hand, something as simple as numbering the steps in a procedural sequence with “1,” “2,” “3” and so on takes little room on the screen—but can be highly effective. One of the versions of FERRETT used this simple, explicit “hand holding” and it improved usability, although eventually this feature became unnecessary after additional tool redesign.
- Well designed “help” screens can greatly enhance a tool’s usability. They should not however substitute for putting as much guidance on the primary interface as possible. Also, “help” should be provided in small units closely tied to specific contexts, not lengthy and irrelevant discourses full of general information. The tools included substantial “help” that users could access, especially the two BLS tools, but perhaps did not indicate clearly enough that help was available and how to access it.
- Some tools attempt to compensate for limited usability by offering extensive tutorials to help novice users acquire necessary expertise. Unfortunately users who need the tutorial are unlikely to spend the time to work through it, preferring instead to “jump in” right away in order to obtain real results This is especially true for web users who expect results in a small number of clicks, and are very unlikely to follow extended performance-based practice exercises or to engage in word-by-word reading of large chunks of text, according to Nielsen (2000).

3.2 Unclear Data Naming and Database Organization

This problem area also has two aspects, as the title indicates. One aspect involves unfamiliar terms from specialized vocabularies that <u>data</u> experts or <u>computer</u> experts embed in the tool. The other aspect involves failure to communicate to users in their terms how data names or descriptions are organized in databases. As an example of the first problem aspect, the data series that are accessible using the Java data query tool often include unfamiliar, specialized terms in the names of the series. These terms often make distinctions between data series that are meaningful only to expert users. A prime example in the labor force data is a distinction between two measures based on the same concept, e.g. "unemployment," where one is expressed as a <u>rate</u> and the other as a <u>level.</u> To the non-expert user, the most meaningful part of the name in either case is the familiar term "unemployment." The technical rate-level distinction may not be noticed at all.

Some distinctions may not only be difficult for non-experts to understand, but they may represent entities that are of little interest or value to broad segments of the total user community. An example of this would be age or income data series which involve unusual, perhaps unequal intervals instead of more familiar 5– or 10-year age groups, or income categories other than those using increments of, say, 5 or 10 thousand dollars to define their boundaries. The unusual categories may have some utility for a few specialist users, but for most users their inclusion in the set simply complicates the selection process by inflating list length.

The other aspect of this problem type, data organization, stems from the fact that the databases to which access tools apply are often organized and hence displayed in a manner that makes it difficult for users to process information (metadata) about database content. Some examples of how data organization can lead to designs that are hard to use are noted in Figure 2, a screen from BLS' Java data query tool. This tool supports winnowing long lists of time series, over 50,000 in the case of labor force statistics, to find the series the user is interested in. The user does this by selecting various values (e.g. the seasonally adjusted number of men employed in agriculture) in text boxes, each containing either a list of <u>statistics</u> which differentiate among series as classes, or all possible values of one of the <u>variables</u> that series may include.

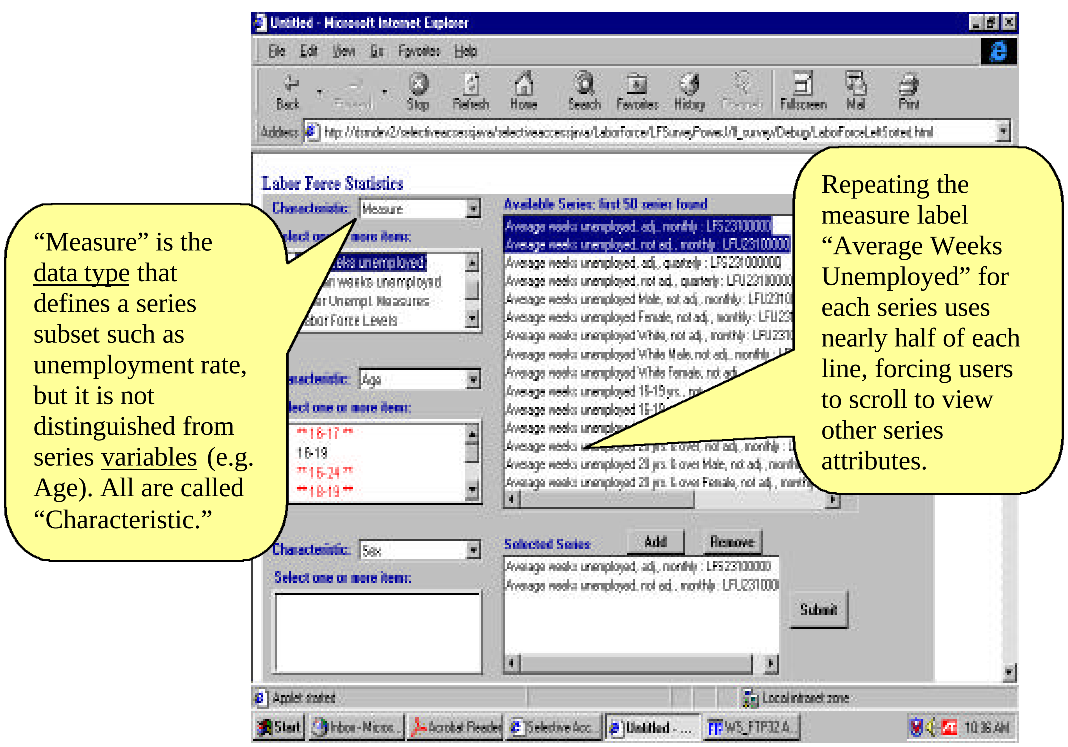

Figure 2. Example—Problems of Data Naming and Organization

Each labor force data series is identified in part by the name of a labor force <u>measure</u>, e.g. unemployment rate. Each measure identifies a broad <u>statistical</u> class of series. This name appears first in the list of defining descriptors for <u>every</u> series involving the measure, but it is not distinguished in any way from the additional descriptors that specify attributes or values of <u>variables</u> used in the series. This has two consequences for the tool designer. First, it requires the <u>tool</u> to indicate clearly to users that they <u>must</u> choose the <u>correct</u> statistical measure—that this is <u>the</u> crucial selection criterion. The design shown in Figure 2 failed to provide such an indication clearly enough. This led some users to pay more attention to specifying choices of the variable components like age and sex than to picking the basic measure defining the series type they were looking for.

Second, repeating the measure in every label for all its class of series can use quite a bit of screen area. The fact that this redundant descriptive term falls at the beginning of each series label makes it more difficult for users to scan the list of series names, since they have to skip over the measure name to find the terms that really distinguish one series from another.

These series names may contain a large number of descriptive terms. From the standpoint of clear labeling, a data access tool should enable the user to view the full name of each series in the currently selected set so that she can determine whether the results the system is returning are satisfactory. In Figure 2, the tool tries to do this, and furthermore show the list of selected items on the same screen. As the figure also shows, however, the text box cannot be wide enough to show long series names because the recurring measure label plus an arbitrary alphanumeric identifier for a series causes parts of long labels to be hidden. Thus the user must scroll left and right to view all information for some series, and this diminishes usability.

Tool developers should adopt practices such as the following to reduce the seriousness of data naming and organization problems. It should be noted, however, that as long as the data to which a tool applies continue to be named and/or organized in ways opaque to users, there is a limit to how much design can compensate.

- A usable tool should provide easy access to definitions of technical, unfamiliar descriptive terms. Such terms might be hyperlinked to a glossary (with easy return to the working interface of the tool), or mouse-over definitions could be used.
- The tool should inform or indicate to the user if a specifying term, such as the <u>measure</u> for labor force series, is a unique identifier for a data <u>type</u> (major sub-classification category). This kind of term should be distinguished from secondary attributes that are used to refine selections within the major data type.
- A usable tool will clearly indicate when users try to specify for missing or never-collected data. A tool can do this by showing missing data types or values in gray or in a different color from valid data specifications.
- A usable tool should make it easy for users to review the results of their specifications, to verify that they are getting the kinds of data they want. The Java data query tool shown in Figure 2 attempted to do this but its success was limited because of how series labels are constructed.
- Including features in a tool that give users the means to combine specification terms, such as "and-ing," may seem an attractive way to gain efficiency in data specification tasks. Research shows, however, that many users have difficulty constructing queries that require Boolean logic, as in SQL (e.g. Greene and Devlin 1990) beyond simple "and-ing, (Sewell and Teitelbaum 1988). At one point, the FERRETT designers included a feature intended to support a kind of graphical construction of SQL-type specifications by pointing and clicking data names and operator buttons, instead of using text entry. This feature did not help users do any better and was dropped from the design. Even if logical operations are performed implicitly, e.g. by "and-ing" individual terms chosen from different lists, users need clear feedback about the result of the operation. More generally, the interface should inform users whether their choices are narrowing ("and-ing") or broadening ("or-ing") the set of data specified.

3.3 Problems of Screen Organization and Layout

One of the authors reviewed about 200 "help desk" messages to the FERRETT support staff in order to get an idea of what kinds of tasks to use in usability tests of this tool. He found that a majority of FERRETT users who contacted the help desk have simple data access goals. For example, most are looking for data involving one or a few variables, often specific to a geographic area such as a state or metropolitan area. They may wish to perform some basic manipulation of these data, such as selecting only one gender category, restricting age values to a certain range, and the like; rarely do they wish to do more complex data transformations. Similarly, the field usability test of the "wage finder" tool indicated that a key set of potential users—wage analysts—usually want to access an average hourly wage for just one occupation and one geographic region at a time. These observations seem to indicate that usable data access tools will offer a set of "core" features to support simple, very common data access tasks and clearly identify these on the tool's main interface. Even if a tool offers some on-line analytic support, as FERRETT does, that too should be simple. Experience with FERRETT shows that most users will be satisfied with simple and straightforward on-line data tabulations, rather than applying any more powerful techniques such as regression analysis. Those users who do complex analyses of the data are normally experts who will download the desired sub-set of data to the desktop for such analyses, because of the convenience and flexibility of working off-line.

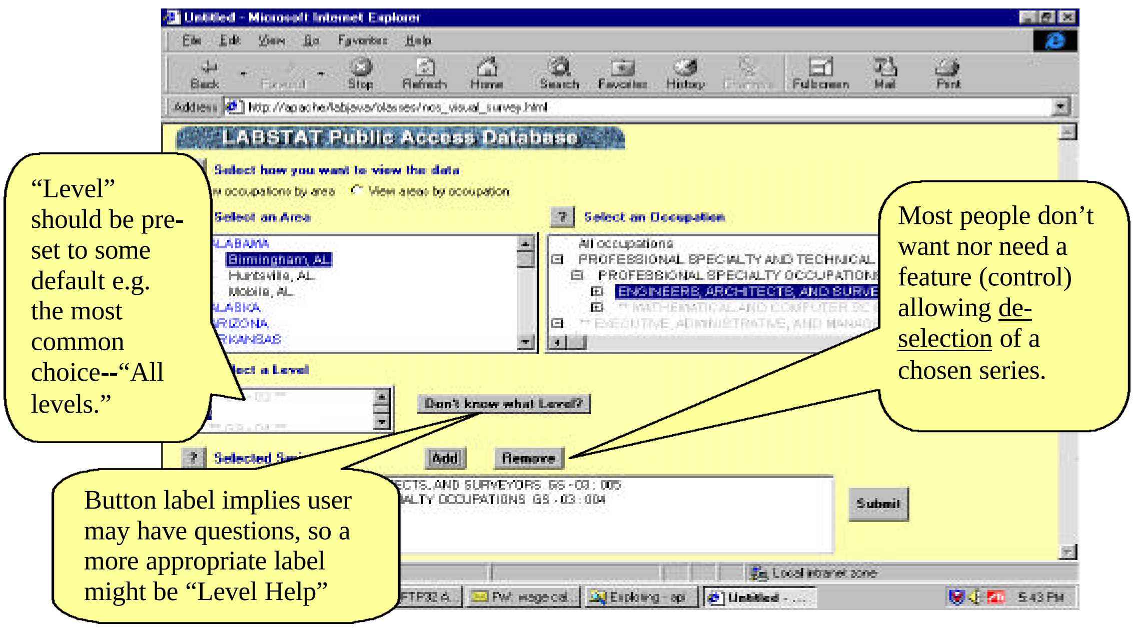

Figure 3: Example—Problems of Screen Organization and "Clutter"

Another way in which screen layout can facilitate users' task performance is by using default specifications on the interface. In many cases, our usability testing clearly indicates that one selection is so popular that setting it as a default choice in the tool eliminates the need to change the setting for most users. In general, it seems plausible for most selection options to set the default selection inclusively, e.g. for "Sex," data for both men and women. In some cases, one of the tools provided no default value for a specification, e.g. in Figure 2 the list box for "Sex" is empty. In such cases users hesitate to manipulate that interface feature. They may feel constrained to explore the feature, e.g. the blank list box, just to discover what choices it offers, even if this serves no purpose in reaching their task goal.

Figure 3 shows a screen shot of the prototype "wage finder" tool that BLS tested with a set of known wage data users at a statewide convention of these professionals. The list box titled "Select a Level" (lower left portion of the screen) permits users to select a federal government occupational (GS) level as one of the specifications for finding an hourly wage. When users opened the tool this box had no default selection. In fact, most users were not interested in getting wage data by level, instead opting for "Overall Average Wage." The latter is an obvious default choice. Note also that a control labeled "Don't know what Level?" proves to be a means of accessing some kind of "help." A label making this specific purpose clearer seems more appropriate. Finally, the test showed that these users had little or no interest in a feature to support unselecting choices from a cumulative list. Indeed, these users typically had no need for any facility to make multiple specifications before accessing data, which the tool supports with the lower-right text box and associated controls in Figure 3. Users said that on the job they would usually look for only one wage at a time.

Some design guidelines that can help developers avoid creating tools with this type of problem are:

- Consider providing alternative versions of tools for different groups of users who differ in their capabilities and goals. BLS' "Most Requested Series" and "Selective Access" are to some extent examples of this. "Most Requested Series" is a kind of "front door" to "Selective Access." It may satisfy the needs of many users, and also teaches the non-expert user by example about the kinds of data series that are comprehensively accessible through "Selective Access." This method of user familiarization with data can be superior to a lengthy tutorial that merely simulates working with the data.
- Most users will not have a clear mental picture of the totality of data that may be accessible using a given tool, especially if the tool allows access to many different data sets or series, each with many dimensions. Where needed to help manage complexity, the tool should make the structure of accessible data explicit by more effective use of graphical design in the interface. Instead of simply showing an alphabetical listing of available data series, the designer should cluster related series by (labeled) topic. This is consistent with Donald Norman's (1988, pp. 54-80) broad design principle that suggests capitalizing on "knowledge in the world," i.e. the interface, not just knowledge in the mind of the user.
- Tool design should fully support data specification if possible, i.e. offer users a chance to specify <u>all</u> data classes, variables, or variable values. To do this may require multiple screens if the data can be specified on numerous dimensions.
- Give users clear feedback on the interface when specified data are <u>not</u> available at the site. The tool should inform users of "dead ends" as early in the specification process as possible.

4. DISCUSSION OF FINDINGS

The most telling effect of tool designers' apparent lack of knowledge about the needs and skills of less-than-expert users was a pervasive assumption that all (or most) users know much more about the data than they actually do. The influence of this error is visible in all three of the cross-cutting problem types. If tool designers believe that users generally know a great deal about the data, both structure and content, then they will not be highly motivated to include much orienting information or procedural guidance in the a data access tool. The tool will require users to work with ill-formed names and unwieldy database structures if designers believe that, in general, these factors will not hinder user performance. And designers who believe that users are skilled in gaining access to data, and that they attach great importance to doing complicated operations with the data they get, will do little to logically organize interface features so that they highlight the key steps in simple data access tasks.

The assumption that users know more about the data than they actually do is particularly detrimental to usability because the current generation of data access software tools is usually retrofitted to legacy databases in which the data are organized to optimize storage efficiency, not accessibility. This organization may not reflect how users conceptualize interrelationships among different variables. The data are often described in specialized terms quite unfamiliar to non-experts. Users are expected to understand these terms, e.g. the difference between "employment rate" and "employment level". Users may need to understand the subtle differences between statistics that sound very similar but are collected in different surveys by different methods, e.g. wages published by occupation from the Occupational Employment Statistics (OES) program and from the National Compensation Survey (NCS). For example, OES data are reported at the national level only, while NCS reports down to the level of metropolitan areas within states.

Finally, legacy databases were until recently used only by experts. Non-experts got processed data products, such as tables and reports. So data access tools often still show a bias toward serving "power users," while features supporting simple and rapid access to a few data points or series are relatively de-emphasized in the design.

Designing for all potential users may well be impossible. Nonetheless, if this is the goal – as one could reasonably argue should be the case for government statistical agencies who serve the general public– it may be useful to describe some broad strategies that could lead to greater general usability if applied consistently.

One approach to tool design advocated by BLS consultants who have been studying web data dissemination from a combined perspective of information science and computer-human interaction recommends creating multiple alternative interfaces for different user classes (Hert and Marchionini, 1997; Marchionini, 1998). This design approach in turn requires first identifying a set of user classes, say those classes who are most likely to want or need the data in question, and then implementing different interface designs for each such group. An alternative approach to designing for all users is to design a single interface geared to users with the lowest levels of knowledge and skill. This will frustrate users with higher levels of knowledge and skill but will not leave anyone behind.

A variant of this approach is to design multiple data access tools for multiple well-defined sub-populations of users, making the consequences of this decision explicit in the interface of each tool. A given user's skill and knowledge levels can much more safely be assumed to match one of these tools. BLS has implemented this approach to some degree. For example, LABSTAT users who are seeking a specific, well-publicized number can use an extremely simple but inflexible "Economy at a Glance" tool to access a pre-formatted table containing current data on various major economic indicators produced by BLS, such as the current unemployment rate. The data are accessed with a single click--but are limited to those the designers (with the blessing of BLS subject-matter experts) chose to include. These data are also point estimates rather than time series.

Other users seeking a popular, often-cited data series can use the "Most Requested Series" tool. This tool is less restrictive, but also assumes some familiarity with BLS time series instead of single data points. It allows users to select from a list of frequently accessed BLS time series. Users can select desired date ranges and other output options prior to executing the query. This makes it relatively easy to obtain entire data series – it requires a few choices and clicks – but the user is limited to a small subset of all possible series and can distinguish among them only the basis of date ranges and output formats. Finally expert users looking for a less well-known, possibly complex data series can use the "Selective Access" tool, which empowers (and requires) users to formulate and execute more extensive specifications of the series being sought.

The Census Bureau also takes a multi-level, multi-tool approach. That agency also offers a range of tools, from simple to more flexible and complex. Like those in LABSTAT, these tools provide data from the same data sources but vary in the flexibility they afford the user and the degree of user knowledge they assume.

In some cases, e.g. FERRETT, the Census Bureau's tool for accessing Current Population Survey and other survey data, users are warned explicitly that the tool is for experts. In other cases they are not; for example, the "Data Analysis System" associated with criminal justice datasets stored at the University of Michigan claims that its on-line analysis process is easy enough for "first-time users," when in fact it is requires extensive analytic skill. It is our impression that the kind of warnings about required expertise that are found with certain tools generally understate the degree of user knowledge of the domain assumed in the design of the tool.

5. CLOSING COMMENTS

All three types of cross-cutting problems described in this paper can be traced to a fundamentally flawed process for gathering requirements prior to initial tool design. In the course of these studies, the authors came to realize that the tools' designers lacked the most 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 as agency staff primarily guided the design of the tools. To be sure, some of this guidance was informed by staff perceptions of the needs of (usually) equally expert data users outside of government, such as academics, policy analysts and other elites. Little or no information was available about data use and needs of members of the general public, who gained the means to become data users once agency data were made available on the Web.

Web data access tools must be designed to meet the needs of the entire broad spectrum of data users on the Web, including first-time users. In order to accommodate this diversity, a process to collect these diverse requirements and communicate them to tool developers is badly needed. The diversity of web users is so great that it is probably impossible to get a complete profile of all possible data users' needs. Establishing even more limited needs assessment process is a daunting challenge, demanding creativity to design and substantial resources to implement. 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 a greater variety of data users who visit our statistical websites.

The authors hope that this paper will persuade its readers in the statistical community (and beyond) that usability engineering activities, including usability testing, can help assure that future data access tools will be much more usable and useful than the "first generation" now available on the web. In order to make future tools more usable, agencies urgently need to collect information about their entire user base's priorities in the kinds of data they want, their familiarity with data in general and those data to which the tool will apply in particular, and other relevant information. Only then can the design process be grounded, as it should, in a clear definition of the characteristics of the group or groups of users for whom the tool is primarily intended. The difficult work of collecting information about user needs will support a much more effective process of designing usable data access tools.

Frequent usability testing as the tool 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 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, less authoritative—but more usable—sources.

6. REFERENCES

Carroll, J. M. and Rossen, M. B. (1987). "Paradox of the Active User," in J. M. Carroll (ed.) *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction,* Cambridge, MA: MIT Press, pp. 80-111.

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

Levi, M. D. and Conrad, F.G. (1996, July and August). "A Heuristic Evaluation of a World Wide Web Prototype." *interactions,* **3**, 50-61.

Levi, M.D. and Conrad, F. G. (1999). "Interacting with Statistics: Report from a Workshop at CHI 99." *SIGCHI Bulletin,* **31**, 31-35.

Marchionini, G. (1998). "Advanced Interface Designs for the BLS Website. Final Report to the Bureau of Labor Statistics." http://ils.unc.edu/~march/blsreport98/final_report.html

Nielsen, J. (1993). "*Usability Engineering*," Boston: Academic Press.

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.

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

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.

7. APPENDIX

Various websites afford access to statistical information. As of July, 2000, statistics in the United States can be found at FedStats at http://www.fedstats.gov. This is a portal site for over 100 federal agencies with statistics online. For Statistics Canada, the link is http://www.statcan.ca. For the National Statistics site in Great Britain the URL is http://www.statistics.gov.uk. At a sub-national level an individual agency may maintain its own website. For example, the US Bureau of Labor Statistics site is http://stats.bls.gov. An example of a university-based data repository is found at http://www.icpsr.umich.edu/index.html for The Inter-university Consortium for Political and Social Research at the University of Michigan. Finally many commercial organizations such as Berinstein Research at http://www.berinsteinresearch.com/stats.htm offer other gateways to statistics on the Web.