

“DEEPLY DISCOUNTED” USABILITY TEST OF AN INTERFACE FOR A SURVEY DATA EXTRACTION TOOL

John J. Bosley, BLS
Cavan C. Capps, Bureau of Census

Introduction and Background

For over a year, BLS and the Census Bureau have collaborated on the redesign of the user interface to an application that is designed to support finding, defining and extracting (downloading) customized data sets from statistical (survey) databases. This application, or tool, has been named “FERRETT,” which stands for “**F**ederal **E**lectronic **R**esearch [and] **R**eview **E**xtraction **T**abulation **T**ool.” The initial version of FERRETT has been in use for over two years, primarily to allow users to work with data sets from the Current Population Survey (CPS) through the Census Bureau website. . This version uses a series of CGI screens to support user specification of particular survey variables, and even select specific values for those variables, which can then be extracted from an on-line database. Extracted data can be downloaded as ASCII files, or in SAS file format, for local processing. The tool also supports limited on-line processing, e.g. simple descriptive statistics, without the need to download. Data descriptions (“metadata”) are available on demand to help users make sure they are getting the data they want.

The first version of FERRETT is still in use and can be examined at <http://FERRETT.bls.census.gov/cgi-bin/FERRETT>. It was designed so that it can interact with any database so long as the metadata, or descriptions of the database elements, follow a specific format. Over time, more and more survey databases “owned” by the Census Bureau as well as several other federal agencies have adopted FERRETT as a means of enabling user interactions with their survey data.

A small team of software developers at the Census Bureau, led by the second author of this paper, is creating a new version of FERRETT. This version will use Java instead of CGI scripts to increase functionality and flexibility. The first author of this paper, a usability engineering specialist at the Bureau of Labor Statistics, is collaborating with the development team to conduct iterative usability testing of the interface to the FERRETT upgrade. This interagency collaboration is based on the fact that the CPS and its various supplements is a survey jointly “owned” by BLS and Census, so that both agencies have a large stake in ensuring FERRETT’s usability.

The present paper presents and discusses the findings of a preliminary usability test of an interface prototype for the FERRETT upgrade. This prototype reflected the enhancements to functional and navigational control that are built into the new version of the tool in a graphical interface. This interface is substantially different from the CGI interface to the first version; screen images of both versions will be provided later in this paper. (Figure 1: Screen image of FERRETT CGI implementation and Fig. 2: Screen image of new prototype go about here.)

The prototype GUI as tested provided extremely limited functionality to the user, and was primarily a test of the basic navigational schema for the tool. Navigation is a key determinant of the efficiency and effectiveness of tool to find, manipulate and extract data from databases. One of the acknowledged weaknesses in the CGI-based version of FERRETT was its slowness and rigid structuring of the process for searching, selecting, and extracting data. The authors decided to proceed with the test despite foreknowledge of its limitations because they knew that using Java to extend FERRETT’s functionality beyond CGI would place heavy demands on the limited resources of the Census development team. The authors wanted to obtain empirically-based information about the usability of the new design’s navigational capability, no matter how qualitative and imprecise that information might be. Discovery of navigational usability problems of “show stopping” severity was imperative before the team invested scarce time and resources in implementing the functionality to match the prototypical screen design. The fact that the test this paper reports on was undisguisedly qualitative in nature led the authors to term it “deeply discounted” in the title of the paper.

Method

The development team worked with focus groups of users to define a basic set of needed functionalities. This resulted in a kind of “storyboard” allocating a manageable subset of the functions to each of several screens. They then used the Microsoft Paintbrush program to draw this series of prototypical screens on which graphical elements represented the range of functions planned for the Java upgrade of FERRETT. These images were then embedded into HTML to provide basic navigational controls on each screen. Screens were hyperlinked to enable screen-to-screen navigation. There were up to 5 hyperlinks per screen so that different navigation paths could be tested.

The usability test team consisted of the two authors and two additional Census developers. Testing was conducted in the BLS usability laboratory, which provides the capability to observe users and their interaction with the computer interface by TV in a room separate from the room where a standard PC workstation is set up for use by participants. The team recruited ten people to participate. Eight of the 10 were BLS or Census employees who had used the original FERRETT tool on their jobs, to find and extract survey data sets. The other two were graduate students in economics at the University of Maryland, College Park. They had not used FERRETT but were accustomed to using survey data in their academic work assignments.

Participants worked with the FERRETT prototype interface one at a time. They were given a general orientation to the purpose of the test and the procedures to be used. We have already indicated that only the navigational hyperlinks to move from screen to screen were functional in the prototype. (Each screen’s scroll bars were functional, however, as will be seen in images of various interface examples later in this paper.) Thus it was necessary for a test team member to “sit in” with the user. That team member answered questions about how the new application would respond to user actions involving non-functional portions of the GUI, and also indicated to the user where the active hyperlinks were located on the various screens, in case this was not completely obvious. The team was well aware of the degree to which this method deviates from the norm of non-interference in user behavior during usability testing, and so the team member playing this “tutorial” role behaved as neutrally and objectively as possible. For example, the tutor did not volunteer information but instead waited for the participant to exhibit some interactive behavior spontaneously, e.g. clicking on a button or trying to select an item from a list, before describing what that action will cause to happen in the fully functional application.

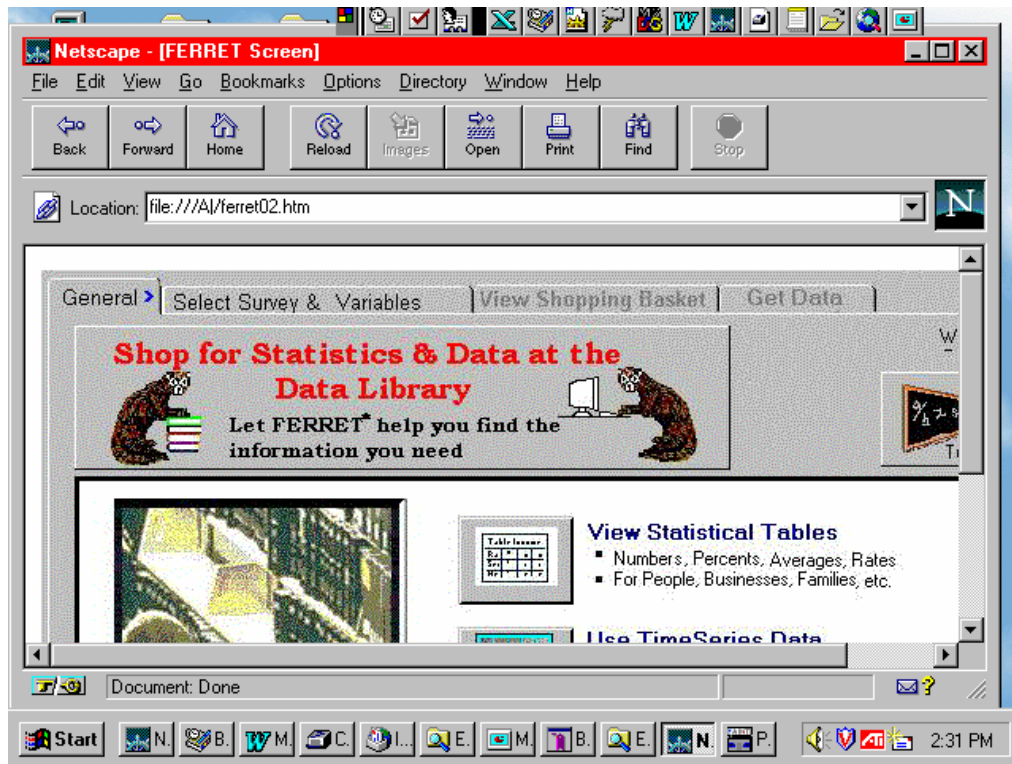
Participants were, however, encouraged to “think aloud” and describe what their intentions were as they simulated interaction with the various parts of the interface. At least two team members in the observation room took extensive notes during the test session, and then debriefed each participant to clarify ambiguous observational points immediately at the end of each session. Since the developers themselves acquired first-hand information about usability problems, the BLS usability engineer did not write an extensive deliverable report on the work. Instead, he wrote a “memorandum to the record” summarizing his own observations. This memorandum serves as the primary basis of this paper.

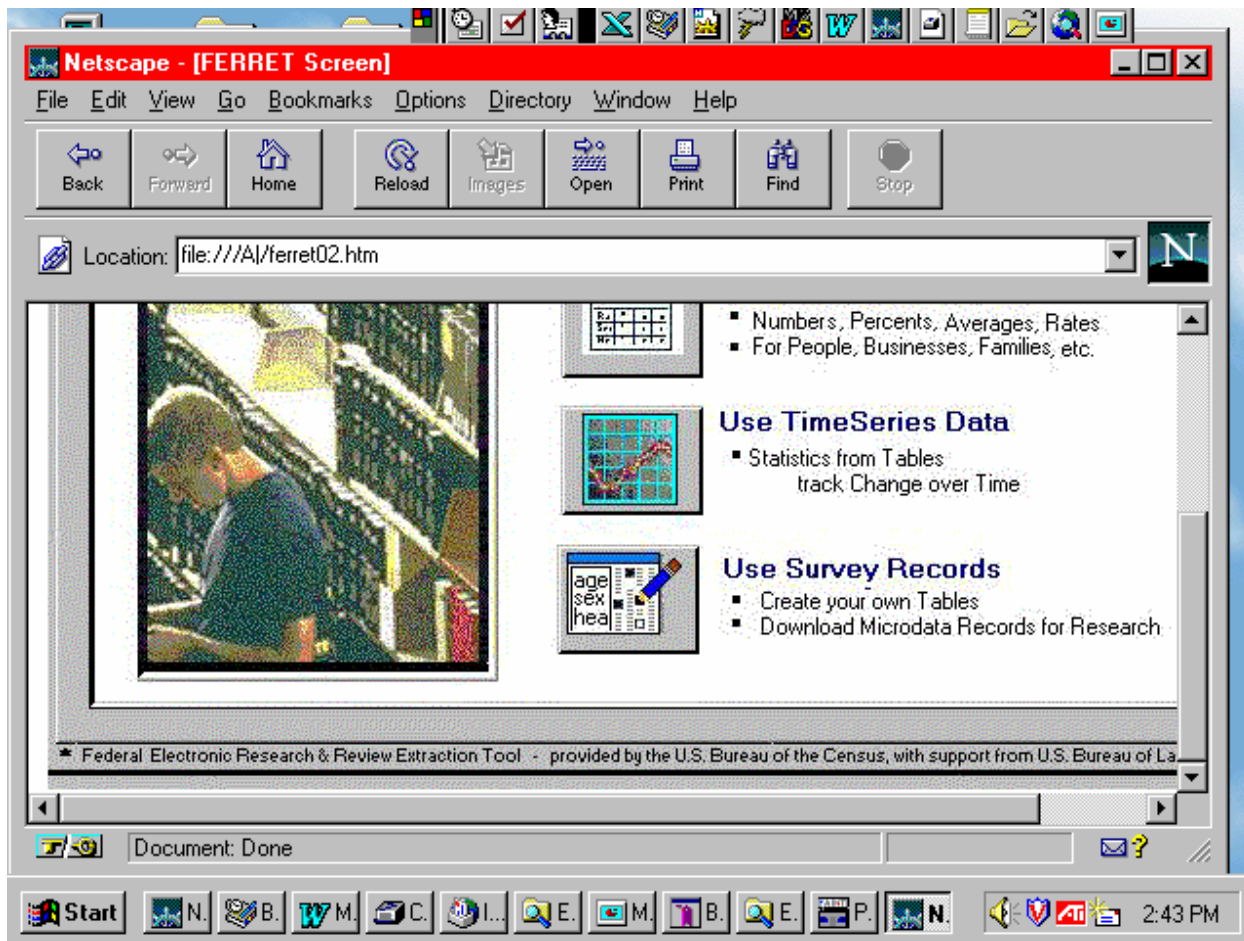
Results

Given the “deeply discounted” approach used in this study, the results to be reported here might better be characterized as “usability issues identified,” rather than as reliable and definitive characterizations of specific usability problems. The presentation of the results will be organized on a screen-by-screen basis, since the individual FERRETT screens are intended to support a “natural,” intuitive series of steps leading up to extraction of data from one or more survey databases. A graphic reproduction of each screen will be inserted in this paper in the order the FERRETT developers envision them being used, and the principal usability issues uncovered for that screen will be described and discussed.

The first screen (Fig. 1) is intended to provide an overview of data elements (variables) included in each particular survey that is accessible. By scrolling down on this screen, the user will be able to pre-select the type of data to access, be it tabular, time series, or “microdata,” the latter defined as the actual unprocessed survey data on file. The FERRETT redesign plan calls for

implementation of microdata access as first priority, and so this study used the particular set of interfaces that will support user interaction with microdata data sets.





**Figs. 1 and 1A: Startup Screen for FERRETT Upgrade
(Two separate figures are required due to limited scrolling range in Windows)**

Thus, none of the three prominent buttons, offering fundamentally different ways of looking at data, was enabled for this screen. In the fully functional application, microdata access would be gained by clicking on “View Survey Data.” The only active control for this test was the tab labeled “General” at the top of the screen, and since the other tabs are obviously deactivated (grayed out), no significant usability problems were posed by the design of this startup screen.

The test participant moved to the next screen (Fig. 2) by clicking the “General “ tab. The next screen is shown below.

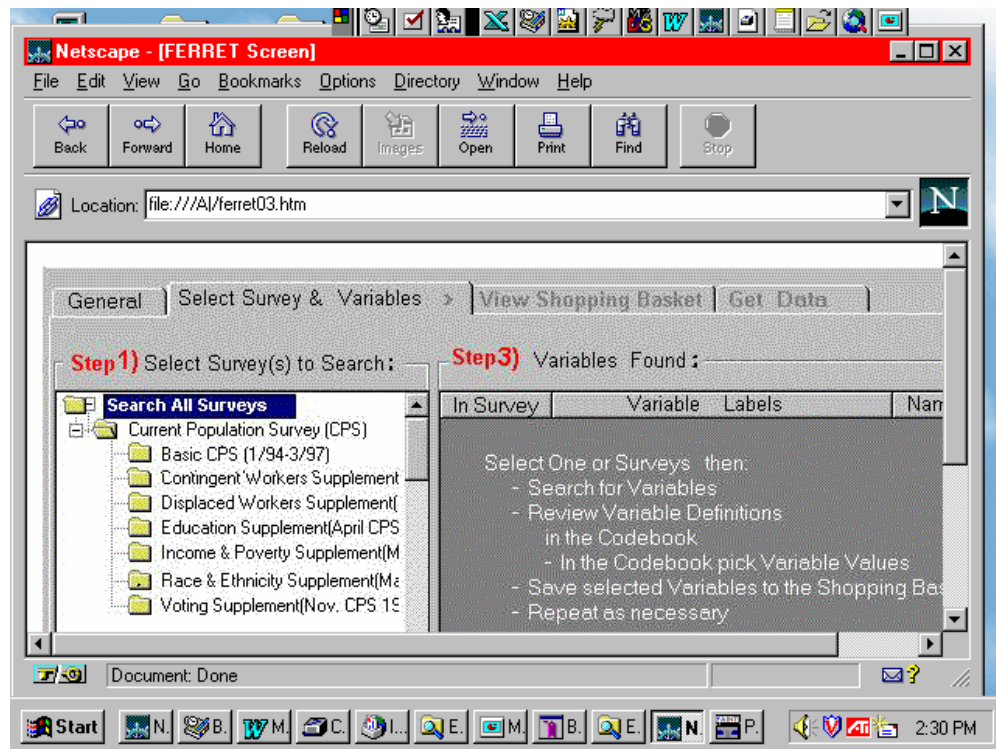


Fig. 2: "Getting Started" Screen

Note that this brings forward and makes active the screen designated "Select Survey and Variables" in the tab set at the top of the screen. In the scroll box at the left of this screen, the user sees a list of individual surveys accessible by using FERRETT. Those shown in this example are the Current Population Survey (CPS) Basic Survey for the years 1994 to March, 1997, along with six of the latest versions of monthly CPS supplemental surveys. This display uses the familiar Windows file folder metaphor. The default selection (shown) is "Search All Surveys." The area at the right, another scroll box, contains instructions for how to use this and following screens.

Here we encounter an obvious usability problem in the way this screen is designed. Note that the designer has provided "stepwise" navigation guidance in red text. The screen, however, is not large enough to display "Step 2," which is in fact a search functionality; it is hidden below the lower screen border, and test participants were confused by this. The next figure (Fig. 3) shows Step 2—but now note that the context-providing row of tabs at the top of the screen have scrolled out of sight.

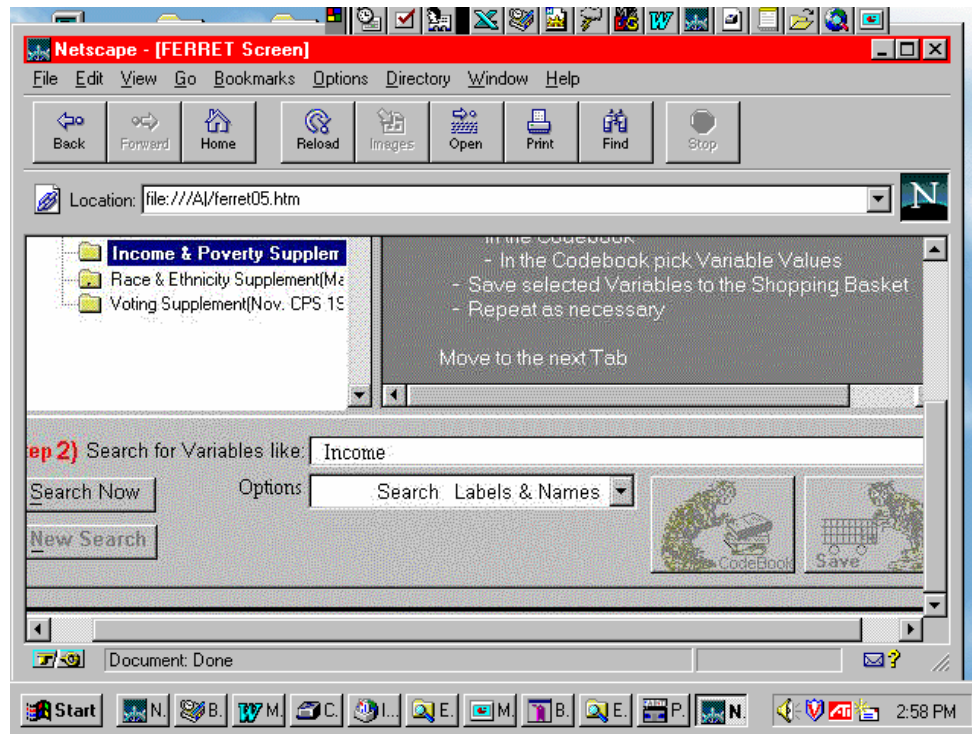


Fig. 3: Search Function Implementation

Now in this figure, the “missing” Step 2 (in terms of the screen shown in Fig.2) is visible—although in the prototype the screen width had been sized so that horizontal scrolling was necessary (a minor problem easily remedied.) More serious concerns are raised by the following observations:

- The fact that the (lengthy) instructions in the gray field to the right appear much less important than they are. They appear in a kind of subdued “reverse video” whereas the selection box to the left is much more perceptually salient. Also, note that the user needs to scroll in order to read the entire text of the instructions.
- The label on the drop-down list associated with the search text entry field is the single word “Options.” This did not convey any information about the nature of the option set—and the very use of a drop-down list implies that a number of options (choices) is going to be offered.
- There is a similar lack of clear description of the two buttons on the right of the screen, although for expert survey data users the label “Codebook” will be meaningful. The “Save” instruction on the button depicting a FERRETT with a shopping cart is insufficient to clearly indicate that this button is used to temporarily store a satisfactory search outcome, and not to “save” data in the usual sense.

Figure 4 shows the result of clicking on the “Codebook” button, although the screen title “Code Book and Value Restriction” is slightly at odds with the button label, introducing as it does the notion of recoding variables (see the “Recode” button at the right of the screen) but without using that particular term, calling it instead “Value Restriction.”

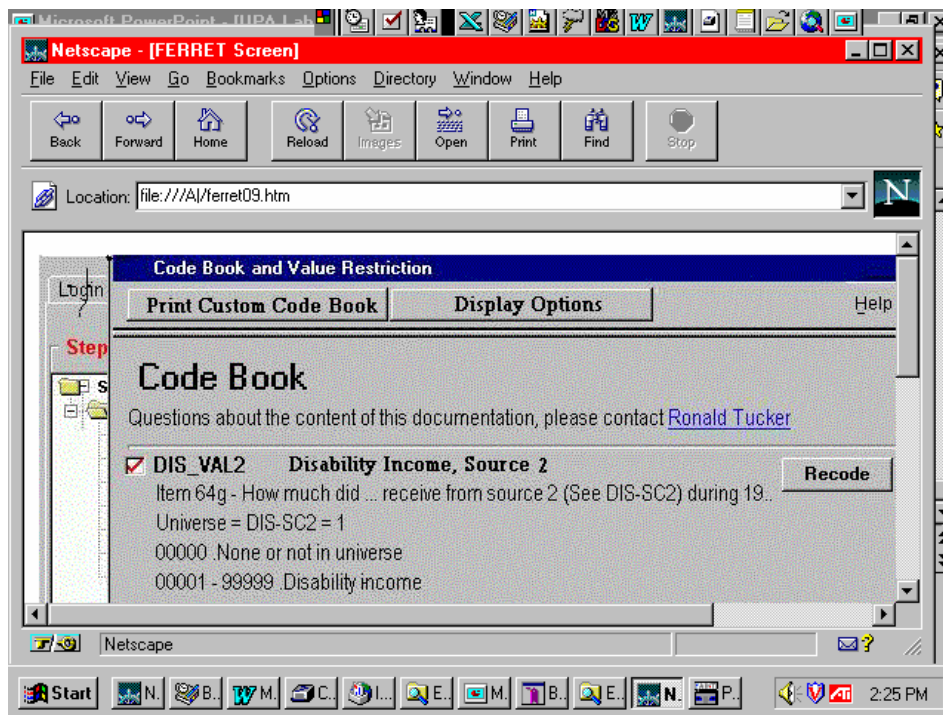


Fig. 4: Code Book (Metadata or Data Dictionary)

Usability problems on this screen were associated with the following aspects of this screen.

- For the first time, a “help” function is introduced into the interface. Furthermore, “help” is offered in two forms. On-screen (presumably) help is offered by the button at the top of the screen. Personalized but delayed help is offered by the link to Ronald Tucker by (presumably) e-mail.
- The function of the check box appearing next to the mnemonic variable name (in this case DIS_VAL2) is not indicated to the user. This actually was intended to enable the user to select the variable for later use by checking the box. Taking an action such as this while working with “documentation” seemed a foreign idea to many participants.
- The nature of the application’s response to clicking “Recode” is not clear. Users might well be reluctant to click this button, thinking that some automatic data transformation would be initiated, and the interface does not offer any indication of a means of stopping or reversing the process.

The next figure, Figure 5, illustrates a “pop-up” calculator that appears when the “Recode” button in Figure 4 is clicked.

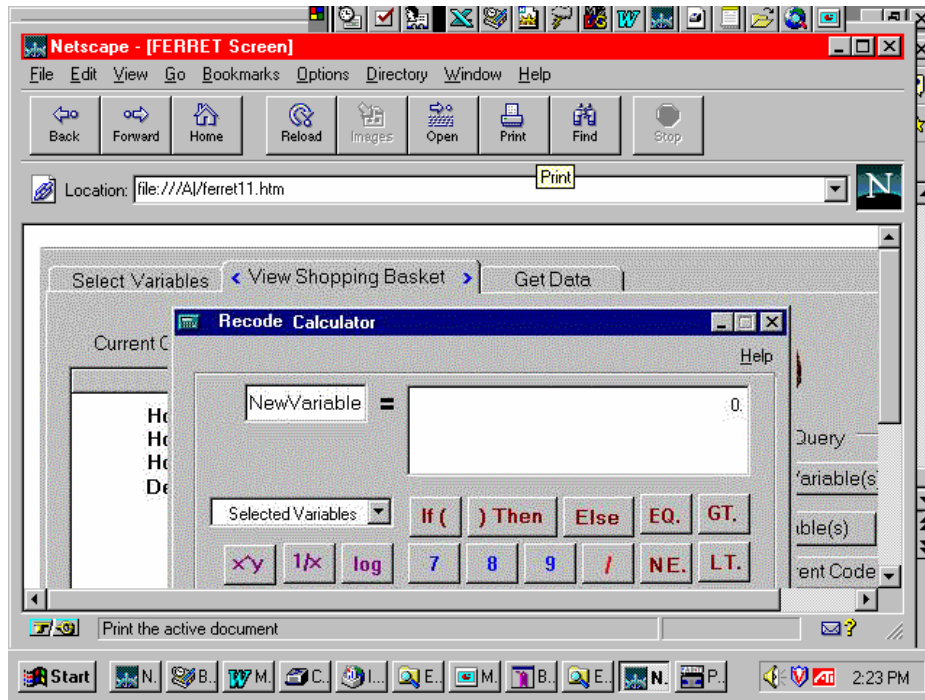


Fig. 5: “Pop-up calculator” –An Aid to Formulating New Variables

The FERRETT development team created a “calculator” modeled on the one that is supplied as an accessory to the Windows operating system. This is intended to enable the user to write code including conditional specifications, Boolean relations, and the like, to create new variables from sub-elements of the existing array of variables in a survey. The calculator was not functional during this usability test, so the only usability information the team could gather consisted of verbal evaluations of this tool’s perceived utility. In general, the test participants’ descriptions of their perceptions of how this tool would function were so diverse as to indicate that they would have problems using the tool if this layout and design were implemented. Their comments also indicated that they thought of recoding categorical variables (selecting values from a value domain) differently than from working with continuous variables such as age or income, where they might wish to segment the continuous variable into disjoint sub-ranges, select only part of the total range, and so forth. This casts some doubt on the capability to design a single interface that would support these qualitatively different processes.

The next figure, Figure 6, shows how the FERRETT prototype implemented one of the key abstract concepts in the system as an interface. This is the concept of a “shopping basket” of variables. Certainly this metaphor is used widely in e-commerce on the World Wide Web. How well does it serve the needs of statistical data users?

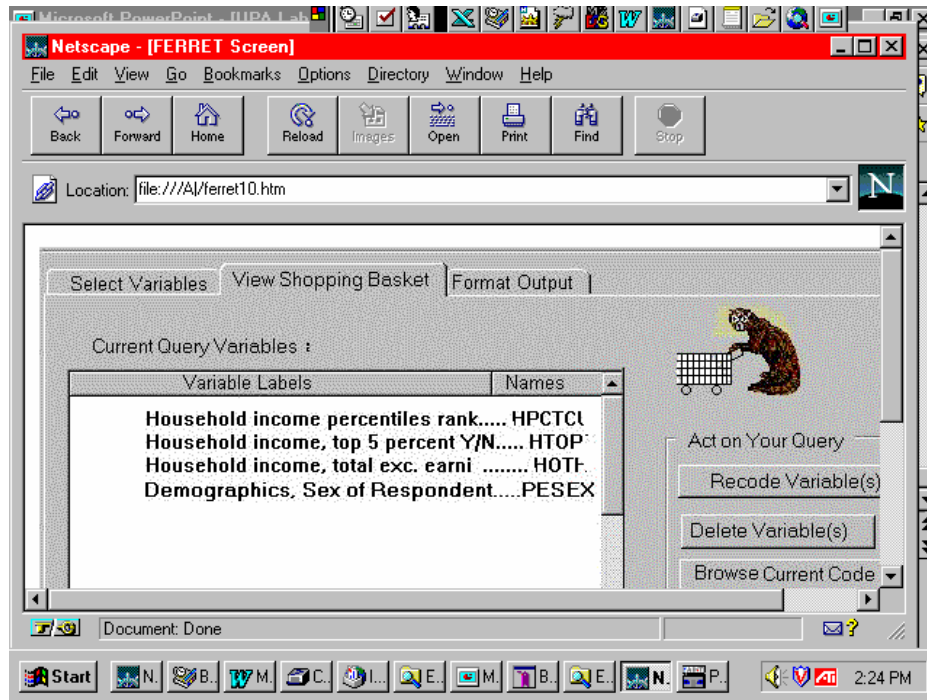


Fig. 6: “Shopping Basket” Containing Data Selections

Note first that clicking on the “View Shopping Basket” tab at the top activates this screen. Also note in Figure 1 that this view is disabled until a survey and at least one variable have been selected. The names of survey variables selected up to this point appear in the scrollable list on the left of this screen. The buttons on the right—“Recode Variable(s),” “Delete Variable(s),” and so forth, provide the user with the means to perform a number of operations on the variables listed that were introduced in earlier screens. This then is a key screen for checking “work in process,” and making changes or corrections to that work from a single interface. Test participants generally understood how this screen was intended to function, and appreciated the convenience represented by the multiple functions accessible by means of the controls available.

In Figure 5, the tab to the right of this one is labeled “Get Data” whereas in this figure that tab’s label is “Format Output.” At a superficial level, this difference boils down to a minor inconsistency in the prototype interface, overlooked in proof-reading and easily remedied. At a deeper level, this difference is symptomatic of the lack of well-established conventions for designating necessary functions in statistical interfaces. We will return to this topic in the final section of this paper.

Figure 7, the last screen in the FERRETT redesign prototype, is the “payoff” screen. The functions on this screen enable users to get the data they have selected, specified, and in some cases created by using functions on the previous screens.

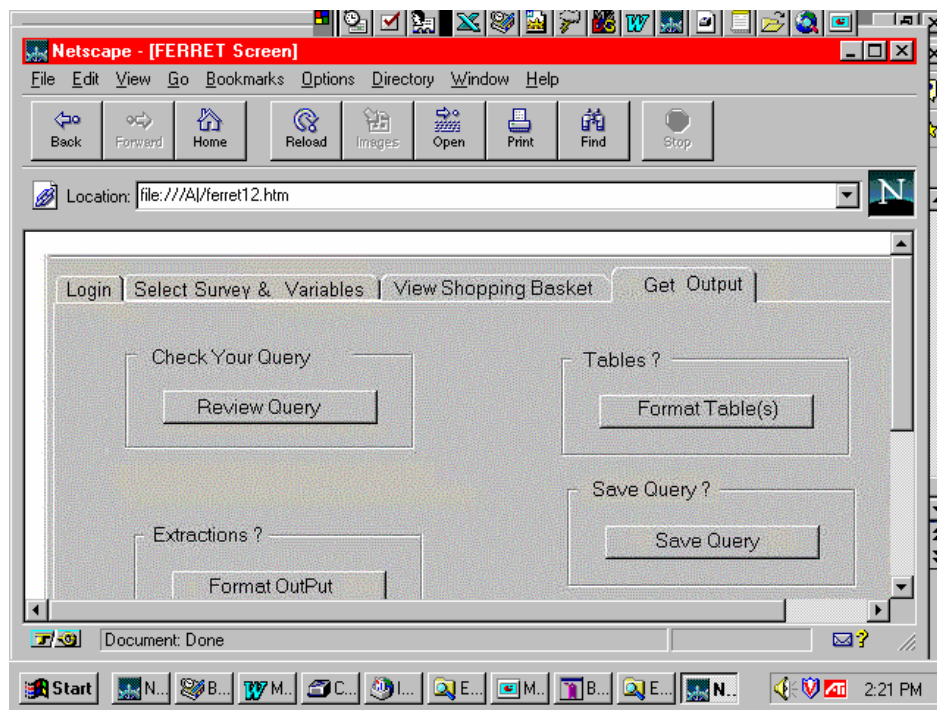


Fig. 7: “Get Output:” Formatting and Acquiring the Data

It was extremely difficult to get useful information about the test participant’s reactions to this screen. For one thing, there was relatively little instructional or informational material on the screen, and for another, the participant had to imagine the functionality represented by each of the buttons, based on a verbal account by a member of the test team. In fact probably the most useful information we obtained related to the scarcity of explanatory or instructional material on the page. Obviously the controls take up only a modest proportion of the total screen real estate. So it would be easy to expand on such cryptic notations as “Extractions?” or explain briefly the difference between the “Tables” function (which implies a kind of formatting), and the “Format Output” label on the “Extractions?” control.

Participants noted also that the more or less random positioning of the various controls on the interface failed to take advantage of cues from relative positioning to indicate anything about the relative importance or popularity of the functions each button controls.

This completes a “micro” level review of how the test participants worked their way through the FERRET prototype series of interfaces, and some screen-by-screen indications of potential usability problems. In reality, of course, the usability of each of the screens and its functions depends on how well the entire set of screens works together to enable users to find, define or refine, and gain access to on-line data. We turn next to what we see as the major issues this test uncovered at this more “macro” level of usability.

General Concerns and Issues Drawn from the Test

We noted earlier that one key concept guiding the creation of this prototype upgrade (and for that matter the original design of FERRET interfaces) was the idea of a linear process, which the typical user would follow toward the goal of getting the data she wanted and needed. This strategic assumption shows up clearly in the prototype. The navigational tab sequence, left to right, maps a process of initial selection of a survey or surveys, followed by selection of data from that selection. This is followed by a possible intermediate stage of checking the metadata to gain better understanding of questionable choices, or a possible manipulation of the data as it is stored to recode it to conform more closely to the user’s interests. Next, the interface suggests that the user review and confirm all her work up to this point by checking the “shopping basket,” and finally, when satisfied, the process culminates in formatting and actually acquiring the data.

FERRET is one of the pioneering tools for gaining access to statistical data, and as such there were practically no precedents to guide its creation. Under these circumstances, the notion of guiding users

through a set of logical steps by the interface(s)' design was as sensible. This test, as well as inputs from other FERRETT users, raises the question—to be explored through further testing—of how well this model of user behavior, and its representation in interface design, actually works. We should note that FERRETT developers do not have plans to put any a priori constraints on the order in which the user navigates from screen to screen, although of course navigation that doesn't make sense will not be enabled. An example would be viewing the shopping basket before a survey or any variables have been selected. Thus a knowledgeable user will realize that she can work back and forth through the various interfaces at will. Still, the design as tested suggests that there is a “preferred,” if not “correct” way to go about acquiring data. This view may prove to be not only incorrect but counterproductive in many situations.

A related issue is raised by some of the data obtained from test participants. This focuses on the question of how best to organize and allocate the functionality needed to support effective interaction with statistical databases onto the “optimal” number of screens. Even this limited test was able to provide fairly strong indications that one or two screens would not be adequate to support the range of functions the average user should have available. On the other hand, it provided some indication that the number of screens in the prototype may have been at or beyond the upper limit of usefulness. If the screen set had not included the “View Market Basket” progress assessment capability, even this rather simplistic test may have shown serious user problems connected with the need to keep track of the whole action series involving multiple screens.

The FERRETT developers are responding to statements of requirements of sophisticated data users in attempting to provide a “job aid” to facilitate the range of data manipulations referred to in their interface design as “Recode” tasks. Such users are an important part of the customer base for tools like FERRETT. In the authors' opinion, such a capability will prove to be a fairly standard requirement for any application designed to support effective interactions with microdata databases. This usability test, however, provided fairly strong evidence that the attempt to adapt a calculator metaphor for this purpose was not effective. The metaphor was not well understood and as designed participants thought it would probably be difficult to use. The “point and click” aspect of the calculator metaphor, considered in abstraction from how it is implemented, is certainly an appealing, rapid and easily understood mode of computer-human interaction. The problem with the calculator metaphor seemed to be the limitations it imposed on immediate access to such metadata elements as response codes or definitions of variable ranges. This imposes either a substantial cognitive burden to hold these values in active memory, or the need to “shuttle” back and forth between the calculator and the metadata. Perhaps a “point and click” (or “drag and drop”) approach that is embedded in the metadata itself—the “Code Book”—will be more usable.

On the other hand, the test indicated that exercising control through the medium of the metadata documentation is a novel idea and if it is provided, the interface design must provide clear guidance about how it works. As we noted in connection with Figure 4, which showed the Code Book (metadata) screen, participants didn't immediately grasp that a checkbox on the screen along with the data description gave them an alternative way of selecting the variable described. “Documentation,” it seems, is not generally perceived as an environment in which computer-human interaction is expected to take place, although as the spread of World Wide Web usage provides wider exposure to hypertext links, this perception may change.

Finally, the test repeatedly demonstrated how important it is to provide clear and sometimes extensive information and instructions about what various interface components will enable the user to achieve and how to use them. This is reflected in participant's divergent conjectures about what the probable outcome of some action would be, as well as an even more fundamental failure to interpret terse one-word labels or instructions accurately.

To summarize and conclude, these generalizations collectively point toward a few even more general issues. These are:

- Building usability into interfaces for applications that enable interaction with statistical data in all but the simplest cases, such as finding and retrieving pre-formatted, widely-known tabulations of data, is

extremely challenging. The FERRETT development team received, and tried to respond to, a broad range of user requirements. Fulfilling many of these called for great creativity in designing a data system architecture and other “behind-the-scenes” capacities. Representation of ways of benefiting from this computational power will require equal creativity in interface design. The limited test that formed the empirical basis of this paper showed that much more work needs to be done, and hinted at some general directions that work should take. Whatever success has been achieved in making FERRETT a usable tool should be gauged relative to where development had to begin, which was from a meager base of directly relevant interface design experience, much less systematic research.

- There are few clear mappings between existing general interface design principles and practices in non-quantitative domains, and the requirements driving interface design for quantitative work such as identifying and effectively working with statistical data. Take, for example, Ben Shneiderman’s “mantram” of “Overview first, zoom and filter, details on demand.” How can these goals be achieved when an application may give the user access to thousands of pieces of data that are qualitatively dissimilar, and yet with many of the data items carrying semantically similar labels? Data items currently are associated with extensive descriptions (metadata) to make them comprehensible. Can the components of these descriptions be prioritized so that a smaller subset of the most significant descriptors can be used as dimensional definitions in dynamic query interfaces like the starfield displays pioneered by the University of Maryland. Finally, much of the meaning attached to an individual variable inheres in complex relationships among subsets of variables that form a “concept,” with individual members of the subset having little or no meaning when looked at in isolation. How can interface designers settle on an optimal level of “granularity” be defined so that “just enough” of the network of relationships that give the data meaning be defined and displayed?

An example of an attempt to provide a broad view of a wide range of federal statistics can be found on the Interagency Task Force for Federal Statistics website called “FedStats.” (<http://www.fedstats.gov>). Both authors of this paper are on a working group to improve the usability of that site as well as of statistical sites to which it links. FedStats (and the Census Bureau’s own website) offer a “Subjects A-Z” topic list that gives the user some idea of the qualitative range of coverage accessible through FedStats. However, the links represented by a topic title on the A-to-Z list take the user to only one—and hopefully the “best”—of what may be a number of topically relevant pages scattered throughout several agencies’ websites. FERRETT hopes to offer some means of showing what data its user can reach in a more complex fashion, yet to be designed.

- FERRETT is a tool for data dissemination over the World Wide Web. As such, its interface must work for statistical data users at all levels of quantitative literacy, seeking data relevant to a virtually limitless variety of practical questions, some simple but some very complex. This usability test involved participants who were expert data users, and yet this test demonstrated that the amount of user guidance—information, procedural instructions, and feedback on the results of an action—that was present in this prototype was inadequate to support efficient interaction with the underlying database. Furthermore, statistical interface designers and usability engineers should closely coordinate their work to attempt to converge as rapidly as possible on a set of tested and proven design principles and terminology. FERRETT calls the collection of data descriptions (metadata) a “Code Book.” The test showed that this term is far from universally accepted or uniformly interpreted. The statistical interface design community needs to work together to find representations of functions that are nearly universally understood—the equivalent in this domain of “Cut” and “Paste” in office applications that are widely used.

This usability test provided the FERRETT development team with enough empirically-based information to enable them to make a limited number of important design choices with reasonable confidence. As this paper shows, however, the test opened more questions than it resolved. BLS and Census are currently working on several of these questions, so that the next version of FERRETT is easier and more natural to use for users whose skills, knowledge, needs and goals are as varied as is the growing proportion of world citizens who use the Web to get answers to significant personal questions and thereby make vital personal and professional decisions. This prototype test was a low-cost addition to the design process,

designed to identify problem areas before costly programming started. As the system becomes fully functional, additional rounds of usability tests are planned.