Interfaces to Support Customized Views and Manipulation of Statistical Data
Gary Marchionini
University of North Carolina at Chapel Hill

For the past four years, a team of information specialists has been working with the Bureau of Labor Statistics to improve outreach to constituent groups by means of the World Wide Web (WWW). In our preliminary work, we conducted investigations of user needs and tasks and completed transaction log analyses of user behavior at the BLS website, resulting in a user task-type taxonomy (http://ils.unc.edu/~march/blsreport/mainbls.html). In the second year, additional user needs analyses were done (Hert & Marchionini, 1998), transaction log analyses were replicated, an interface prototype that aimed to provide alternative entry points for different user needs and types was built and tested, and recommendations for short and long term design strategies were made (http://ils.unc.edu/~march/blsreport98/final_report.html). In the third year, the interface work extended these efforts to the Fedstats website that includes the bulk of statistics collected and disseminated by 70 federal government agencies. Fedstats is a portal service (also called a locator service or a metasite) meant to serve as a gateway to all U.S. federal government statistics (www.fedstats.gov). An interface prototype (Relation Browser) was built that aims to give people an overview of the range of federal statistics and an alternative entry to them through "look ahead" interface mechanisms. This prototype was tested with users and recommendations for revisions made (http://ils.unc.edu/~march/blsreport99/final.pdf). In the 1999-2000 phase of the project, the Relation Browser was revised based upon the previous year's usability tests, two types of assessment were made, and a final revision was prepared (http://ils.unc.edu/~march/blsreport00/final.pdf). The details of the designs and the usability tests may be found in these reports. A brief summary of the goals and usability testing results are given here.

It is clear that more and more statistical information is being made available to the public via the WWW. Many government agencies aim to improve citizen access and usage of government statistics by making these data available. Broadening access and use requires user interfaces that are easy to use and serve the needs of wide ranges of people. The desire to make statistical information more widely accessible via WWW interfaces raises several challenges to designers:

1. Universal access challenges. There is a critical need to develop alternative interface designs because human abilities and skills vary widely, we must provide access to people with a range of special physical and cognitive needs, numerical literacy skills tend to be low in the overall population, and there is a wide range of statistical needs that might be served by government statistics. In addition to the range of human capabilities and needs, technology continues to evolve and the range of platforms, software and connectivity grows. There is a limit to new requirements that government websites can put on users since the variance in citizen platforms continues to grow—there is a kind of installed base inertia that builds up as larger

portions of the population acquires WWW access. Additionally, the cultural phenomenon associated with the "attention economy" tends to demand immediate and concise (often simplistic) responses to inquiries, regardless of how complex the appropriate results may be.

2. Retrieval challenges. Online information systems offer people possibilities to search properly indexed databases in highly systematic ways or to explore opportunistically. People may use analytical search strategies via queries and selections as well as explore/browse with general queries and selections. WWW interfaces for statistical data must provide appropriate interface tools to support this entire range of search strategies.

3. Display. Once potentially relevant data is identified, it must be displayed and manipulated. People want information alchemy—answers customized to their specific needs, regardless of the form and substance in the underlying data. There are two interface issues here: what to represent (e.g., overviews of large collections; specific tables or graphs with associated annotations, etc.) and control mechanisms for reading/understanding/manipulating these representations (e.g., scroll, jump, drag, zoom, morph).

4. Metadata. Retrieval and display depend on indexes for salient data characteristics. These data about data are known as metadata. Dynamic interfaces require consistent metadata to support different ways to slice and view results. If the metadata exists, it must be organized and incorporated into the interface. This requires additional data transfer to the client side. If the metadata is not readily available, it must be created—a costly undertaking for government agencies.

5. Interface Management. Just as data must be managed from a life cycle perspective (from creation to dissemination to preservation and dispensation), so interfaces must be managed. Agencies must create iterative redesign schedules, conduct iterative usability testing, and maintain active versions of the interface for users with different capabilities. This requires changes in policies and culture as much as changes in technical operations.

With these challenges in mind, we have taken a user-centered and iterative approach to interface design and focused on two particular issues. The design approach is to:

*   Understand human information needs and behavior by observing, interviewing, and working with users.
*   Base design on user needs as well as data characteristics
*   Create prototype user interfaces
*   Test these prototypes with users, and
*   Revise prototypes and recommend process models for implementation and maintenance

In the BLS work, we focused on two particular principles that we believe improve access and use of statistical information: a) offer users alternative entry points and data views;

and b) minimize mouse clicks to get to pertinent data. Alternative views (the Agileview framework defines overviews, previews, reviews, shared views, and peripheral views; Marchionini et. al., in press) aim to address the challenge of diversity in the user population. Minimizing mouse clicks is based on the observation that clicking on a WWW link causes users to wait for information to load and also changes the window focus—both of which lead to cognitive interrupts. The basic idea is that a click is a radical act and techniques to look ahead before committing to a click are desirable.

To this end, we created a set of interfaces that were tested and revised. The first prototype was an alternative entry point for the BLS website and aimed to minimize clicks through information abundant design and use a graphical region metaphor for layout. Figure 1 displays this interface prototype. Note the spatial metaphor that presents three views of the data (data retrieval, publications, fast info) and minimizes mouse clicks by placing large amounts of highly structured information on a single page. In the fast info "region," commonly accessed data is made immediately available (current statistics and news releases) and common scenarios are used as simple entry points.

Figure 1. BLS Prototype 1.
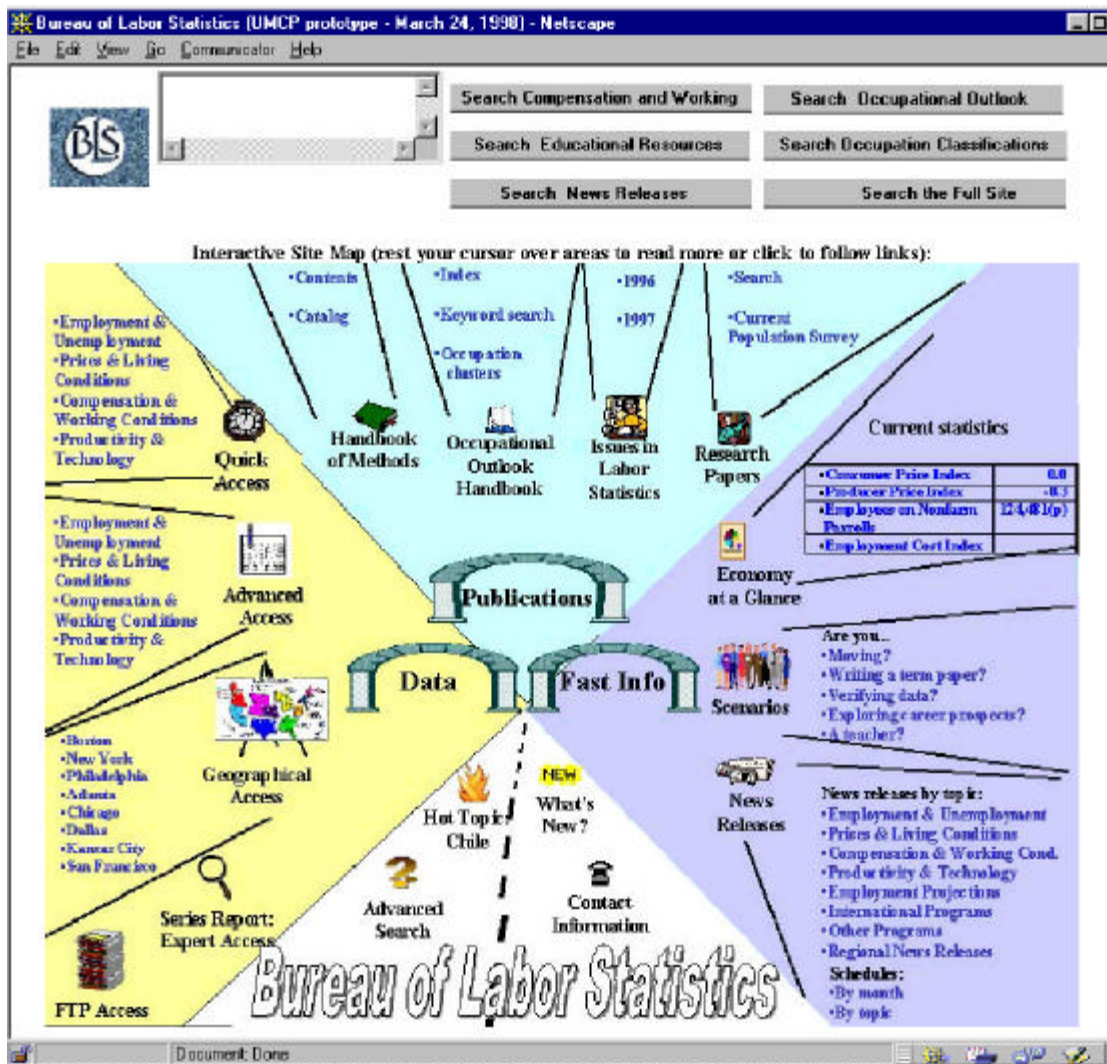
Usability tests were conducted with nine subjects in the BLS cognitive laboratory. Subjects conducted searches for assigned tasks and verbal feedback as well as video records of hands, face, and screen were collected. The results showed that although subjects liked the alternative approaches in to data access, they found the screen complex and that it tended to overload them with information choices. See http://ils.unc.edu/~march/blsreport98/final_report.html for details on this prototype and usability testing.

A second prototype was created to provide overview information for the 200 websites from the 70 agencies accessible through the Fedstats portal to US government statistics. The aims of the Relation Browser tool were to provide an overview of federal statistics on one screen, to provide look ahead without clicking (through mouseover mechanisms), and to provide enriched data attributes to add to understanding of the data in the websites before people committed to loading them. The prototype was built as a Java applet. The first prototype crossed topics (14 topics used in the Fedstats site) with data types (the forms of data available in the statistical sites) and provided filters for geographic location and time. Figure 2 depicts this prototype.

Figure 2. Fedstats Relation Browser Prototype 1.
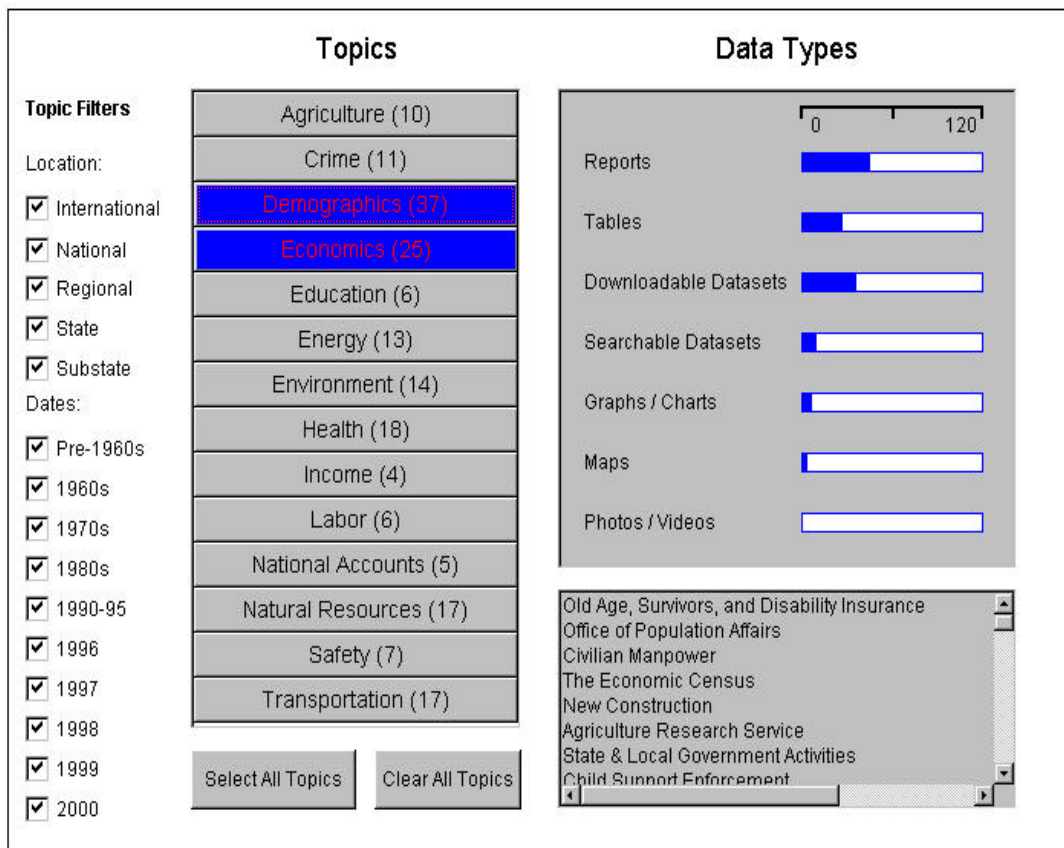
Usability tests were conducted with nine subjects in the BLS cognitive lab. Subjects conducted assigned searches using the Fedstats site map tool as well as the Relation Browser. Verbal data as well as video data were again captured and analyzed. Although subjects liked the filters, they were unable to use them effectively and sometimes got zero hits by over-filtering. A number of screen display suggestions were also made. Based upon these results, the prototype was revised and a field test conducted. Figure 3 displays the revised Relation Browser prototype.

Figure 3. Fedstats Relation Browser Prototype 2.
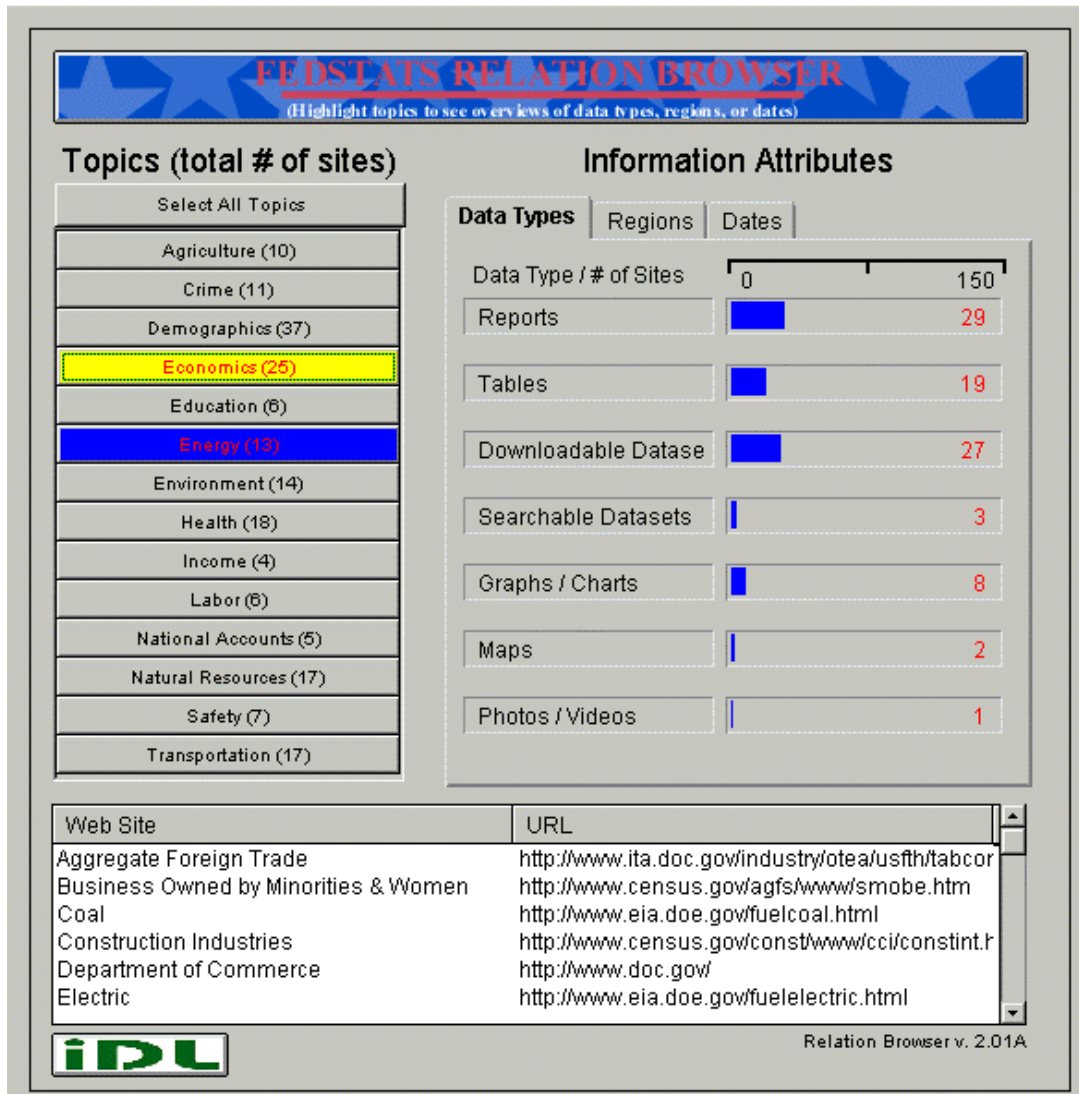
After receiving OMB clearance, the RB tool was provided on the Fedstats home page as an alternative (labeled as an experimental sitemap) for the Sept 1999-May 2000 period. Clicking on the button loaded the Java applet that included a button to provide feedback. The feedback button led to a voluntary questionnaire that contained 10 five-point Likert-scaled questions, one check list question, and two open-ended questions. Although transaction logs showed that the RB tool was used several thousands of times during this

period, only 74 individuals completed the questionnaire. The results showed strong support for adding the tool as an option (50% strongly agreed it should be added). One-third of the respondents reported using the tool from home, thus suggesting that the goal of serving the general public was at least partially met.

Based on the field test, a third iteration of the Relation Browser was implemented and a usability test conducted. Figure 4 displays the final version of the Relation Browser. As in previous usability tests, nine subjects conducted assigned searches with the Fedstats site map tool and the Relation Browser at the BLS cognitive lab. Subjects expressed strong satisfaction with the way that data types, time data, and regions were linked to categories and claimed that the tool would save them time by helping them avoid going to websites that did not have data relevant to their needs. They also noted that some basic level of sophistication and/or training would be needed to take full advantage of the tool and that the font sizes were difficult to see.   Overall, there is good support for incorporating the Relation Browser as an optional overview tool for federal statistics. See (http://ils.unc.edu/~march/blsreport00/final.pdf) for details on the design evolution and usability testing.

Figure 4. Fedstats Relation Browser Prototype 3.

This work has demonstrated that dynamic interfaces can help people gain an overview of statistical data, better understand the range and type of data offered at websites, and save people time as they seek statistical data. Moving from prototype to operational system requires changes in several areas of management. First, interface techniques and options that serve diverse user communities and capabilities must be adopted and integrated into the agency design culture. Second, user-centered, iterative design processes must be adopted that systematize updates and new releases and integrate usability testing throughout the design and production process. Third, the process by which data is captured and released must include procedures for assigning and maintaining metadata that supports dynamic interfaces. This includes possibly classifying data according to attributes across different agencies. For example, each of the 200 people responsible for the different statistical websites linked from Fedstats must specify (index) how their data fits into the various data type, region, and time attributes. In some cases, this is a simple notation, in others, original classification decisions are required, which may in turn require policy decisions. These policies and classifications must be standardized across all the agencies and therein lays an organizational challenge. This final challenge is perhaps the most difficult one, as institutional inertia and turf protection must be overcome.

Although the challenges are great, the possible rewards of improving citizen access are considerable. These may include: equitable access to valuable decision making information; satisfying work experiences; improved government service (and citizen perception and satisfaction); augmentation of the intellect for individuals empowered by access; augmentation of the social intellect as the general population makes better personal decisions; and generally improved performance and quality of life. David Moore's comment in the *International Statistical Review* summarizes the reason we should endeavor to overcome the challenges of universal access and use of government statistics: "Democratization is driven in part by the quantization of society."

References

See the WWW sites referenced in the text for full reports.

Hert, C. & Marchionini, G. (1998). Information seeking behavior on statistical websites: Theoretical and design implications. Proceedings of the 61st Annual Meeting of the American Society for Information Science. (Pittsburgh, PA, Oct. 25-29, 1998). P 303-314.

Marchionini, G., Geisler, G., & Brunk, B. (in press). Agileviews: A Human-Centered Framework for Interfaces to Information Spaces. Proceedings of the Annual Meeting of the American Society for Information Science (Chicago).

Marchionini, G., Brunk, B., Komlodi, A., Conrad, F., & Bosley, J. (in press). Look Before You Click: A Relation Browser for Federal Statistics Websites. Proceedings of the Annual Meeting of the American Society for Information Science (Chicago)