

## FINDING GOVERNMENTAL STATISTICAL DATA ON THE WEB: A CASE STUDY OF FEDSTATS

IRINA CEAPARU

### ABSTRACT

*Over 70 United States federal government agencies collect and store statistical data that are made available to the general public over the World Wide Web. A common portal with an easy-to-use interface was designed starting in 1997 to ensure that the general public could find appropriate data tables and reports easily (<http://www.fedstats.gov>). This study analyzes the FedStats Web site to determine its usefulness to citizens, by way of a usability test with 15 subjects who responded to three scenarios. The results show that out of 45 attempts, only 7 were successful in finding the correct answer. Also, more than 50% of the subjects reported dissatisfaction and high levels of frustration with the Web site's usefulness and ease-of-use and with the lack of results they obtained. The article recommends incorporating into the Web site accessibility and universal usability design principles that could improve the dissemination of statistical information.*

---

Irina Ceaparu is a graduate student at the University of Maryland, Department of Computer Science, Human Computer Interaction Laboratory. [irina@cs.umd.edu](mailto:irina@cs.umd.edu).

*Acknowledgements: The author appreciates the support of the National Science Foundation grant for Information Technology Research (#0086143) Understanding the Social Impact of the Internet: A Multifaceted Multidisciplinary Approach.*

Over seventy federal government agencies in the United States collect and store statistical data that eventually are made available to the general public. The World Wide Web offers the best medium for dissemination of these data. However, to facilitate access to these statistical data, a common portal with an easy-to-use interface is required. This portal is necessary to ensure that the general public, as well as researchers and statisticians, know about the existence of such data and can easily and quickly have access to the huge amount of information the federal agencies provide.

The only significant attempt to create such a portal began in 1997, when the FedStats Web site (<http://www.fedstats.gov>) became publicly available. The site is designed to complement the already existing Web sites of each federal agency, by providing a unique point of access to all collections of statistical data from almost 100 different agencies. The Web site is intended to help users find the information they need without having to visit several Web sites and without needing previous knowledge of the structure of the governmental agencies.

This article provides analysis of the FedStats Web site and determines its usefulness to citizens. In this respect, a usability test was conducted, and the article describes its results and the recommendations made to the designers of the Web site.

#### **PREVIOUS LITERATURE**

Bosley and Straub (2002) have conducted multiple usability tests of statistical Web sites, in which representative general-public end users tried to find data to answer everyday questions that have statistical answers. Their analysis led them to the following ten heuristics for interactive statistical database interfaces:

- 1) Give an overview of the available data.
- 2) Support situational awareness within the available data.
- 3) Display and clearly define metadata.
- 4) Put adequate and clear instructions on the interface.
- 5) Link users to frequently requested analysis.
- 6) Use simple interaction schemes to accomplish complex query-building.
- 7) Summarize outcome of complex data specifications for review and confirmation.
- 8) Offer choices of easy-to-interpret output formats.
- 9) Design output formats to facilitate quick and reliable query validation.
- 10) Help users avoid searching for non-existent or non-available data.

A team of researchers (Ambite et al. 2001) at the Digital Government Research Center conducted a project on how to simplify data access. They built a system for disseminating statistical data from several federal agencies. The system consists of three parts: the database manager and access planner, the overarching ontology in which terms are defined and standardized, and the interface. The interface addresses the following issues: support for adaptive, context-sensitive queries via a system of guided menus; display of tables created by the integration back-end from one or multiple individual databases, along with footnotes and links to original data sources; and browsing of the ontology that supports the entire integration model, with the capability to display concept attributes, relationships and definitions in graphics and text.

The CHI 99 Webshop, "Interacting with Statistics," addressed several issues of designing interfaces for statistical databases. At this Webshop, Donohue (1999) described two usability issues in analyzing the Bureau of Labor Statistics' Web site: cross-survey data retrieval and metadata. In cross-survey data retrieval, in which searches are done across all relevant data sources based on criteria submitted by the user, the main challenge is the design of the interface. The interface must be intuitive, assist users in creating efficient queries, and return the output in a meaningful way. Donohue identified five broad areas for cross-survey information: 1) Geography, 2) Industry, 3) Occupation, 4) Demographic Characteristics and 5) Subject/Topic. The metadata, meaning the information about the data, impacts how the data can be used. The user must know how to identify the metadata and how to use it in order to check the correctness and usefulness of the answer. Johansen (1999), from the Division for IT Development at Statistics Norway, described the Kostra project—electronic data collection from municipalities to Statistics Norway and distribution of statistics from Statistics Norway back to the municipalities. He discussed three interfaces for accessing and interacting with regional statistics over the Internet: 1) Regions (a typical interface might be maps), 2) Time (a time-line or scale would be an obvious approach for an implementation) and 3) Variables (due to the nature of these, a list or tree structure could act as an interface).

The United Nations Statistical Commission and the Economic Commission for Europe proposed a set of guidelines for statistical metadata on the Internet (2000). The guidelines refer to three types of metadata: 1) metadata assisting search and navigation (e.g., sitemap/table of contents, descriptions of statistical subject areas and institutions, a list of key words linking everyday language to statistical tables/graphs and local search engine), 2) metadata assisting interpretation (e.g., content description, measurement unit, time period, regional unit) and 3) metadata assisting post-processing (e.g., for downloads, data/metadata should allow further processing using suitable tools like spreadsheets, databases, packages for statistical analysis and the like).

The United Nations Statistical Commission and the Economic Commission for Europe (2001) proposed another set of guidelines for best

practices in designing Web sites for dissemination of statistics. The following is a summary of their guidelines:

- 1) Users have to be the center of attention. It is important to investigate who the users are, what they really need, how they use the data and what their competence levels are. At least some kind of usability test should be organized, be it even with very simple tools. Usability studies provide an essential feedback that permits improvement of the Web site and, in the long term, reduces costs necessary for redesign.
- 2) The development and maintenance of the Web site of a statistical office should not just be the task of IT experts and special dissemination staff. Strong support among the (top) management is one of the most essential success factors for a statistical Web site. It must be ensured that the necessary competence to develop and maintain a Web site is available.
- 3) The Web site architecture must provide comprehensible search and navigation across the whole Web site. It must be easy to find the desired data inside the Web site. The response time must have high priority.
- 4) A flexible and consistent metadata support should ensure that the interpretation of published statistical data is transparent and comprehensible to the users; any kind of misunderstanding and misinterpretation should be avoided. The metadata should support the comparability of data over time, i.e., historical data should be supported by metadata. The published statistical data must be consistent across the whole Web site.
- 5) Users must be able to download data into their own technical environment for post-processing. The data should be provided in well-accepted standard formats.

In 1996 Marchionini (1998; 1999) began working with the Bureau of Labor Statistics (BLS) and the FedStats group. Their goals were to gain a better understanding of the interaction of non-specialist users with statistical data and to analyze how federal statistical agencies can adopt and adapt technologies to better serve the needs of diverse constituencies. They developed a user task and type taxonomy, methodologies for assessing user behavior and user interface design guidelines—as well as a user interface tool called the Relation Browser for use in the FedStats Web site.

Haas (1999; 2000; 2001) joined the team in 1998 and focused attention on understanding the linguistic mappings between end-user vocabularies and agency vocabularies. The Relation Browser tool was revised, based on the usability tests, and implemented as an alternative site map for FedStats and field tested over a nine-month period.

In 1999, Marchionini et al. (2000; 2001) joined the team, focusing on making statistical tables from federal agencies easier to find and use. The academic team worked with several government agency partners (BLS, EIA, Census, NCHS) to study how people find and understand statistical tables. Metadata issues for statistical information were investigated and mapped to interface tools. Dynamic query tools and natural-language processing techniques were developed to facilitate finding statistical tables, and a table browser tool was developed and extensively tested for usability.

A new effort began as part of the NSF Digital Government Program (GovStat 2002) to study issues of integration requirements needed for a national, statistical-knowledge network. The goals are to work with FedStats agencies and state agencies to: 1) make government statistical data more accessible and understandable to non-specialists who need data to make better decisions; 2) address horizontal integration across federal agencies (e.g., interoperability for data and metadata vocabulary, interfaces, and the overall user experience) and 3) vertically integrate across levels of government (e.g., linking state and local data to federal sources).

In 2002, the project “Ending User Frustration” (ENUF), related to user frustrations with computers, was started at University of Maryland. The project had three phases. The first study (Ceaparu et al. 2002) measured, through 111 user-subjects, the frequency, cause and the level of severity of frustrating experiences encountered when using computers. It found that the most frequent causes of frustration were error messages, dropped network connections, long download times and hard-to-find features, while the time lost due to the frustrating experiences ranged from 31% of time spent on the computer to 46% of time spent on the computer. The second study (Bessiere et al. 2002) examined psychological and social perspectives on frustration in an attempt to clarify the relationships among variables—such as personality types, cultural factors, goal attainment, workplace anger and computer anxiety—and to develop a technology frustration model. The third study, still in progress, focuses on workers who use computers as part of their work-related responsibilities. Specific goals of the study include: determining the most frequent causes of frustration, measuring the time lost due to frustrating incidents and describing the impact that the frustrating incidents have on the users and their interactions with co-workers. This study will allow a better understanding of how user frustration impacts employees and companies.

### **STUDY DESIGN**

The study was triggered by Hert’s (2002) article “Developing and Evaluating Scenarios for Use in Designing the National Statistical Knowledge Network.” As part of the NSF Digital Government Initiative, Hert’s project deals with a scenario-based approach to statistical information. In a first phase, fifteen scenarios were developed via a participatory design approach that

involved agency partners. Team members of the project searched on the Web for the answers to these scenarios, in an attempt

“...to understand the specific integration challenges represented by a given scenario, the information that was available to address the scenario, provide further information on how the scenario might be better formulated, and gain insights on how existing sites were addressing integration challenges [...] and how they were presenting data.”

As a result of this search, the team developed a database containing keywords used in the search process, URLs of the pages containing relevant data for the scenarios, types of statistics found and the like.

Three of these 15 scenarios were felt to represent three hierarchical levels of users' need of information:

- 1) Construct an understanding—the question as well as the answer is very elaborate, and the source of information to be searched is not clear from the context

*Scenario: “I’m a social activist in the Raleigh-Durham, North Carolina area and have become increasingly concerned about urban sprawl and the loss of rural areas for both farming and recreation. I need statistics to support my claim that significant differences occur when urban development occurs in rural and/or farming areas.”*

- 2) Search for specific data—the user needs only to locate information

*Scenario: “I would like to open a grocery store specializing in organic products in the greater Seattle metropolitan area. What are the trends in production and consumption of organic food products? Would the Seattle area be a good place to locate?”*

- 3) Comparative search—the user has to look for information regarding an interaction between two phenomenon

*Scenario: “I’m contemplating a move from Seattle to Bozeman, MT. How do they compare?”*

*Pilot study:* A pilot study was conducted prior to the main study to verify the correctness and usefulness of the procedure that would be followed. The pilot study helped refine the observation methodology and provided a list of most common and frequent types of frustrations the subjects might encounter during the study.

The 15 subjects (six females and nine males) came from different backgrounds: Computer Science, Library and Information Sciences, Economics, French, Sociology, Electrical Engineering, MBA and Medical Studies. The first 8 subjects were asked to navigate the FedStats Web site using only the Topic A-Z links page (<http://www.fedstats.gov/cgi-bin/A2Z.cgi>). The other 7 subjects used

the FedStats homepage (<http://www.fedstats.gov>) as a starting point, and were allowed to use all the features of the Web site.

All subjects were given the three scenarios mentioned above and asked to find the answers to the scenarios using the FedStats Web site. They were given 10 minutes for each scenario. A think-aloud protocol was used, in order to observe and register the subjects' actions and comments. After each scenario, the subjects were asked to fill out a short questionnaire intended to reflect their opinions about the scenarios and the results they obtained about the Web site (ease-of-use and usefulness) and about the level and type of frustration they experienced during the study.

*Analyzing the results:* Out of the total 45 times the queries were asked, 31 times the subjects did not find an answer. Only 7 times did the subjects find the correct answer, while the other 7 times a partial answer was found. For the first set of 8 subjects using the Topics A-Z link, in 16 of 19 cases no answer was found. For the second set of seven subjects, in 15 of 19 cases no answer was found. Thus, these results indicate that there was no significant improvement in the subjects' performance when the entire Web site, instead of just the Topics A-Z page, was used. All subjects spent an average of about seven minutes looking for information for each of the scenarios.

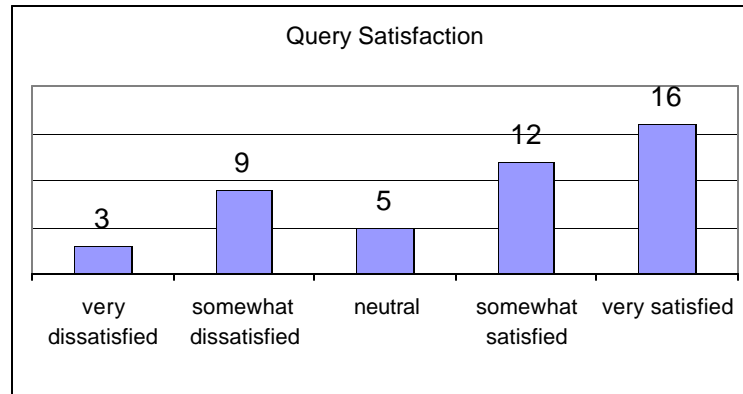
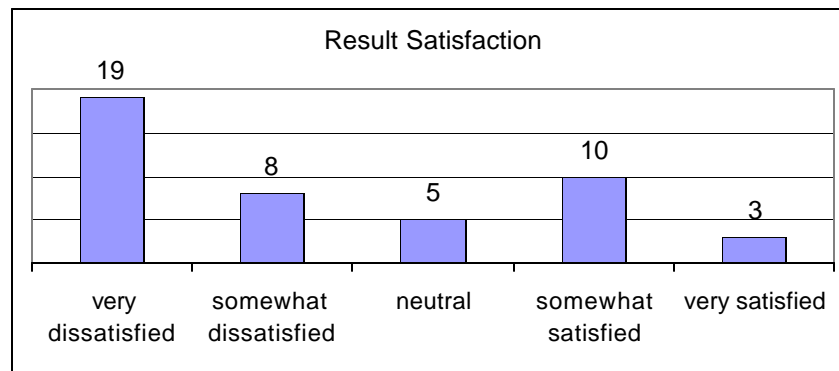
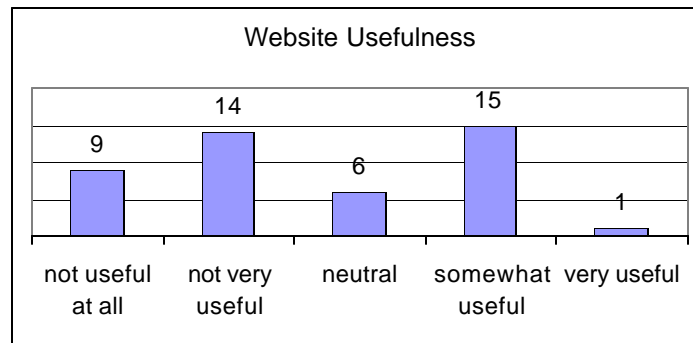
Figure 1 shows that when asked about how satisfied they were with the way the query was stated, the subjects reported being somewhat satisfied or very satisfied in 28 (62%) of the cases.

Figure 2 shows that when asked about how satisfied they were with the answer found, the subjects reported being somewhat satisfied or very satisfied in 13 (30%) of the cases.

Figure 3 shows that when asked about how useful the FedStats Web site was for them, the subjects reported that the Web site was not at all useful or not very useful in 23 (51%) of the cases.

The rest of the questionnaire assessed the level of frustration the subjects experienced while trying to accomplish each task. They were first asked about their general level of frustration after completing the task, and then about what specific frustrations they encountered and at what level. A list of seven frustrations seemed to be most common during the pilot test, and the subjects were also asked to list any other type of frustration that might have occurred during the study. The frustration scale varied from 0 to 10, with 0 being not frustrated at all and 10 being the most frustrated.

Figure 4 shows that the average general level of frustration was 5.9. Subjects reported above-mean (scale values 6 to 10) levels of frustration in 26 (57%) of the cases.

**FIGURE 1: QUERY SATISFACTION RESULTS FOR ALL 45 INSTANCES OF THE STUDY****FIGURE 2: RESULT SATISFACTION RESULTS FOR ALL 45 INSTANCES OF THE STUDY****FIGURE 3: WEB SITE USEFULNESS RESULTS FOR ALL 45 INSTANCES OF THE STUDY**

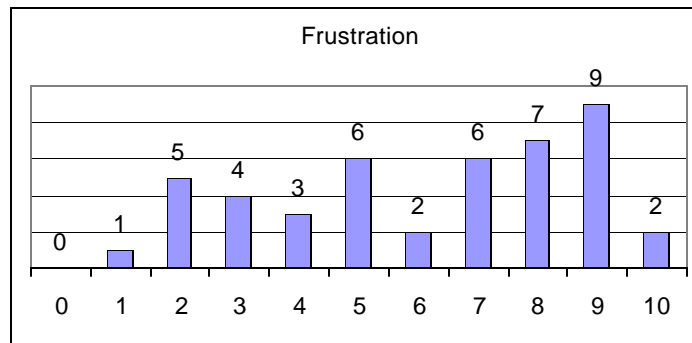
**FIGURE 4: FRUSTRATION LEVEL FOR ALL 45 INSTANCES OF THE STUDY**

Figure 5 shows that the average level of frustration in understanding what the query/scenario was asking was 1.8. Subjects reported 0-2 levels of frustration in 33 (73%) of the cases.

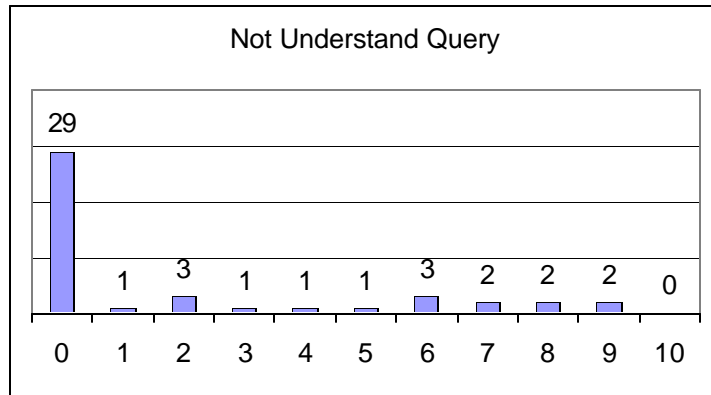
Figure 6 shows that the average level of frustration in understanding the results was 2.6. Subjects reported 0-3 levels of frustration in 31 (68%) of the cases. Their comments indicate that this is because most of them did not find any results that were relevant to the query.

Figure 7 shows that the average level of frustration related to the usefulness of the results obtained was 4.7. Subjects reported 5-10 levels of frustration in 26 (57%) of the cases.

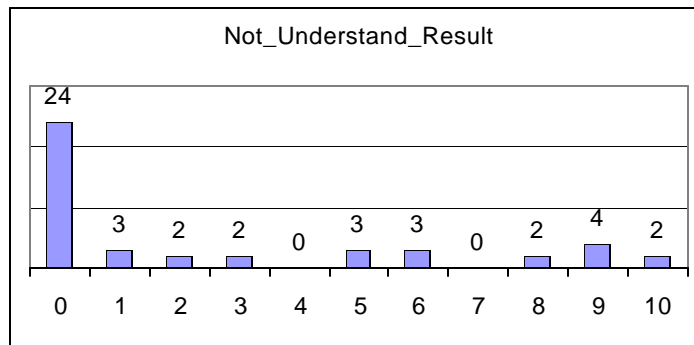
Figure 8 shows that the average level of frustration related to the granularity—in terms of geographical level and time period—of the results obtained was 4.1. Subjects reported 4-10 levels of frustration in 24 (53%) of the cases. Thus, most of the subjects who came very close to the answer were not able to get a better granularity of the data, because the refinement process was either hard (the ways to perform the refinement were not obvious) or impossible to do (there was no way to perform the refinement).

Figure 9 shows that the average level of frustration above, related to how easy it was to use the Web site, was 5. Subjects reported 5-10 levels of frustration in 26 (57%) of the cases. The subjects complained that the language used on the Web site was often not appropriate. Terms like “cities” were disguised under “metropolitan areas,” making the search process hard for a user without previous knowledge of the statisticians’ terminology. Also, the subjects expected to be able to use tools that would simplify their task. For example, when wanting to compare attributes of two different regions, subjects could not perform a comparative search or a quality-of-life analysis. Even a simple search proved sometimes to be very ineffective and misleading, since logical operators were not always implemented properly in the search function.

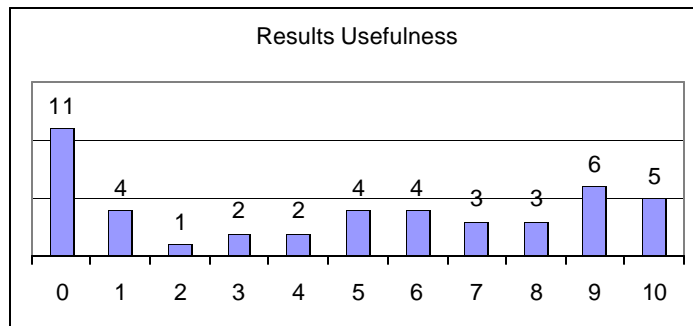
**FIGURE 5: RESULTS FROM THE STATEMENT: "I COULD NOT UNDERSTAND THE QUERY"**



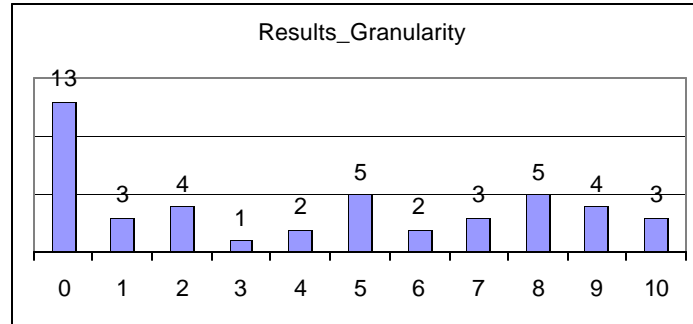
**FIGURE 6: RESULTS FROM THE STATEMENT: "I COULD NOT UNDERSTAND THE RESULTS I GOT"**



**FIGURE 7: RESULTS FROM THE STATEMENT: "THE RESULTS WERE NOT VERY USEFUL"**



**FIGURE 8: RESULTS FROM THE STATEMENT:  
“THE GRANULARITY OF THE RESULTS WAS NOT SATISFACTORY”**



**FIGURE 9: RESULTS FROM THE STATEMENT: “THE WEB SITE WAS HARD TO USE”**

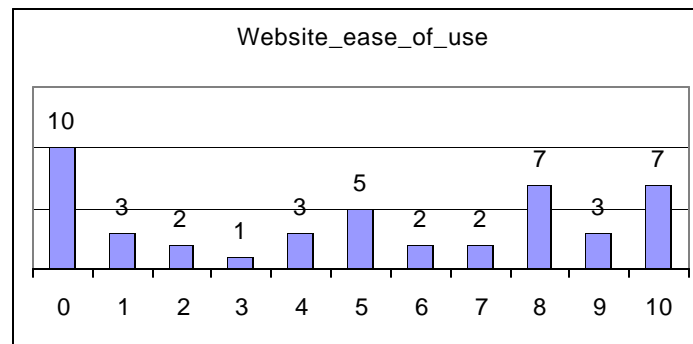
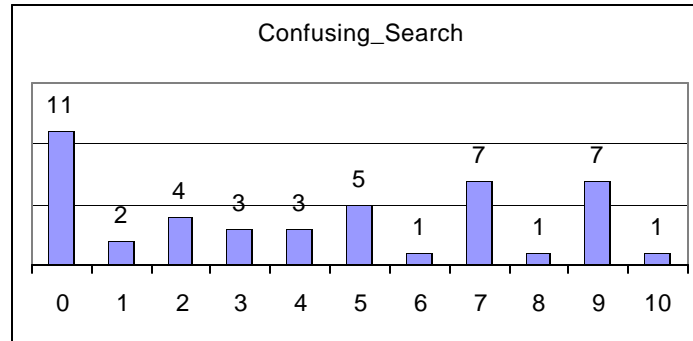
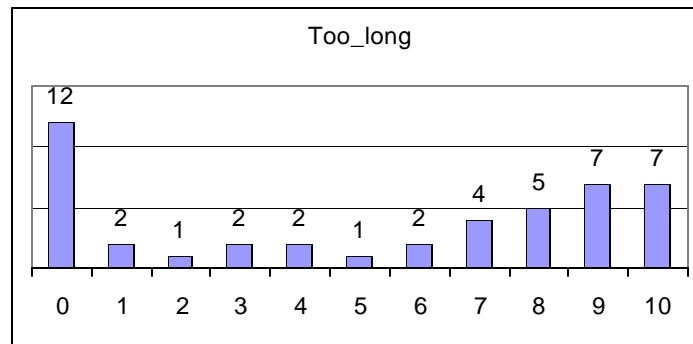


Figure 10 shows that the average level of frustration related to how confusing the searching process was while navigating the Web site was 4.2. Subjects reported 4-10 levels of frustration in 25 (55%) of the cases. The subjects had a hard time navigating through the Web site. The categories through which the data was made available did not prove too useful, and the subjects kept coming back and forth on the same paths, not knowing which one was actually relevant to their questions.

Figure 11 shows that the average level of frustration above related to how much time they spent searching for the correct answer was 5.2. Subjects reported 5-10 levels of frustration in 26 (57%) of the cases. The way the Web site was structured and presented proved to be a challenge to the subjects. All the subjects were first time users of the FedStats Web site, and most of them were novice users, as far as statistical Web sites were concerned. The time they spent on the Web site was longer than expected, because the interface was not easy to learn and the search sometimes presumed previous knowledge of statistics.

**FIGURE 10: RESULTS FROM THE STATEMENT: “THE SEARCH PROCESS WAS CONFUSING”****FIGURE 11: RESULTS FROM THE STATEMENT: “I SPENT TOO MUCH TIME”**

The following types of frustrations were also reported by the subjects:

- Need more related links of the type “See also”
- Could not find links to any keyword in the query
- Need geographic granularity by cities
- Need search keyword option on Topics A-Z page
- Obvious keywords missing
- No “cost of living” calculator
- No way to set up comparative statistics
- Use easier to understand language
- The search engine does not always correctly implement logical operators
- Topics were confusing

The subjects were asked to freely talk about the experience with the Web site. Here are some of their comments:

- Having a background or familiarity with this kind of research would probably help greatly
- FedStats was close to useless
- There is too much data
- Can I go to Google?

Exhibit 1 provides two examples of one subject's experience with the Web site in trying to find the answer to the second scenario and failing.

It is important to note that the subject knew that statistical agencies use the term "metropolitan areas" instead of "cities." Most of the subjects did not find an answer to this scenario because they could not find a link related to "cities." Also, the link to "American FactFinder" seemed familiar, and the subject went directly to it. Most of the subjects who followed the same path clicked on all the links on the US Census Bureau page except the "American FactFinder" one.

## CONCLUSIONS AND RECOMMENDATIONS

Accessibility and universal usability should be two of the primary concerns of the governmental agencies in providing citizen services through the agencies' Web sites. Dissemination of statistical information should be governed by the following design principles:

- 1) *Universal access*—The interface should accommodate the diversity of users—not only expert users, but also first-time and one-time users should be able to easily access and find the desired information. In the case of FedStats, most of the subjects were confused by the design of the Web site, and even after the second task, they did not "learn" the interface.
- 2) *Easy navigation*—The information available should be presented in a structured way. All subjects indicated that a home page containing categories by topics, categories by agencies and a search function would best serve their needs. An alphabetical list of topics slows down the search process, especially when it cannot ever be complete from the user's point of view. In the case of FedStats, subjects indicated frustration when searching the A-Z topics without finding keywords they were expecting to find. On the other hand, the most used feature besides the search function was "Search by agency."

### Exhibit 1: Specific Examples of Frustrating Experiences

- 1) (Scenario 2) The subject starts on the "Topic links A-Z" page (<http://www.fedstats.gov/cgi-bin/A2Z.cgi>), and looks for the following keywords suggested by the scenario: organic, food, Seattle, regions. None of them is found. The closest match was found in the link "Food consumption patterns," which the subject follows. The home page of the US Department of Agriculture Food Survey Research Group is displayed, and the subject chooses to follow the link "Foods Commonly Eaten in the United States, 1994-96" and performs a "Find in page" search for the keyword "organic."

Since no results are obtained, the subject abandons this path and goes back to the "Topic links A-Z" page, and follows the link "Food stamps." The home page of the US Department of Agriculture Food and Nutrition Service is displayed, but the subject cannot find anything relevant on the home page, so this path is also abandoned.

Finally, the subject goes back to the "Food consumption patterns" link, follows it once more, this time looking at the site map. The site map does not prove to be very helpful, and after consulting it for a while the subject chooses to follow the link "Food Consumption Surveys in the US Department of Agriculture." The document does not contain any statistical information related to organic food, so the subject gives up the path and the search process entirely. It is to be noted that the subject did not find a way to restrict the search to the Seattle region. Also, since no clear link to organic foods and markets was found, the subject got lost in the huge amount of Web sites and data that might have been relevant.

- 2) (Scenario 3) The subject starts on the "Topic links A-Z" page (<http://www.fedstats.gov/cgi-bin/A2Z.cgi>) and looks for the following keyword (even though not suggested by the scenario): metropolitan areas. The link is found, and it leads to the US Census Bureau page about Metropolitan Areas. The subject finds a link to the "American FactFinder" Web site and follows it. A number of links are available, providing information about: "General Characteristics: Population and Housing"; "Economic Characteristics: Employment, Income, Poverty"; "Social Characteristics: Education, Ancestry, Language"; "Housing Characteristics: Value, Rent, Owner Costs". All data can be selected for a specific city or town. The subject stops the searching process, since all the desired data were found.

- 3) *Common language*—The terminology used to present the information available should be easy to understand. Except for the expert users, all others most likely do not have enough knowledge to look for the “scientific” terms, and instead use common language. Also, the agencies should not expect the users to know the structure, the exact role of each agency or the interactions between agencies. All these should be transparent in the search process. In the case of FedStats, many users complained about the keywords used on the Web site (e.g., the common keyword “cities” is hidden under the not-so-common phrase “metropolitan areas”). On the other hand, the subjects found the “key statistics” link very helpful and the short description for each agency.
- 4) *Comparative search*—The Web site should allow comparative search. Since this is a very common task for users searching statistical information, the Web site should provide ways to comparatively view and analyze data. In the case of FedStats, the third scenario proved to be more difficult and more time consuming than it should have been, because the subjects had no way to perform a comparative search.
- 5) *Advanced search*—The search feature should have full functionality. It should support a comprehensive search through the huge amount of data available, support logical operators and provide relevant output. In the case of FedStats, although the search box was the most commonly used method to find the answer to the scenarios, in most cases it provided useless output, and sometimes it misled the subject by not correctly implementing the use of logical operators.
- 6) *Data tools*—Common-use ways of viewing and analyzing statistical data should be available on the Web site. Statisticians and researchers need to be able to quickly perform analysis of data by certain well-established criteria. In the case of FedStats, most of the users were unpleasantly surprised to notice the lack of a “cost-of-living calculator” feature.
- 7) *Data granularity*—Allow users to choose the granularity of the information searched in terms of geography and time. In the case of FedStats, subjects were often not able to find the information at the city level—instead they were offered data at only the state or county level. Additionally, subjects expressed the desire to be able to choose the time interval for which they wanted to search for data.

To summarize, this study was intended to discover the advantages and the shortcomings of a Web site that provides access to statistical information. The FedStats Web site was chosen because it is the biggest portal to governmental statistics on the Internet. While this usability study had a reduced scope, it provided sufficient insight and experience with this type of analysis to point the way for design of future studies with larger coverage.

**REFERENCES**

- Ambite, J. L. et al. 2001. Simplifying Data Access: The Energy Data Collection (EDC) Project, *IEEE Computer*, February 2001, p. 47-54.
- Bosley, J. and Straub K. 2002. Data exploration interfaces: Meaningful Web database mining by non-statisticians, *Proceedings of the IBM Make IT Easy Conference 2002*.
- Bessiere, K., Ceaparu, I., Lazar, J., Robinson, J. and Shneiderman, B. 2002. Social and Psychological Influences on Computer User Frustration (to be published).
- Bessiere, K., Ceaparu, I., Lazar, J., Robinson, J. and Shneiderman, B. 2002. Understanding Computer User Frustration: Measuring and Modeling the Disruption from Poor Designs, *The Information Society*, 2003 (to be published).
- Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J. and Shneiderman, B. 2002. Determining Causes and Severity of End-User Frustration, *International Journal of Human Computer Interaction*, 2002 (to be published).
- Donohue, K. 1999. Designing Interfaces to Statistical Databases, In: CHI Workshop Interacting with Statistics 1999
- GovStat Project. 2002. Integration of Data and Interfaces to Enhance Human Understanding of Government Statistics: Toward the National Statistical Knowledge Network, 2002, Available at: <http://ils.unc.edu/govstat/>
- Haas, S. 1999. Knowledge Representation, Concepts, and Terminology: Toward a Metadata Registry for the Bureau of Labor Statistics. In Final Report to the United States Bureau of Labor Statistics, (July, 1999).
- Haas, S. 2000. A Terminology Crosswalk for LABSTAT: Mapping General Language Words and Phrases to BLS Terms. In Final Report to the United States Bureau of Labor Statistics, (September 29, 2000).
- Haas, S. 2001. From Words to Concepts to Queries: Helping users Find Series and Variables to Satisfy Their Information Needs. In Final Report to the Bureau of Labor Statistics, (November 12, 2001).
- Hert, C. A. and Marchionini, G. 1997. Seeking Statistical Information in Federal Websites: Users, Tasks, Strategies, and Design Recommendations: Final Report to the Bureau of Labor Statistics. (July 18, 1997).
- Hert, C. A. 1998. FedStats Users and Their Tasks: Providing Support and Learning Tools: Final Report to the United States Bureau of Labor Statistics. (August, 1998).
- Hert, C. A. 1999. Federal Statistical Website Users and Their Tasks: Investigations of Avenues to Facilitate Access: Investigations of Avenues to Facilitate Access: Final Report to the United States Bureau of Labor Statistics. (1999).

- Hert, C., Marchionini, G., Liddy, E. and Shneiderman, B. 2000. Interacting with Tabular Data through the World Wide Web. White paper presented to the FCSM Statistical Policy Seminar, November 2000.
- Hert, C., Marchionini, G., Liddy, E. and Shneiderman, B. 2001. Integrating Electronic Systems for Disseminating Statistics. Federal Committee on Statistical Methodology Statistical Policy Working Paper 32: Seminar on Integrating Federal Statistical Information and Processes. (April 2001). Washington, DC: Office of Management and Budget. p. 219-226.
- Hert, C. 2002. Developing and Evaluating Scenarios for Use in Designing the National Statistical Knowledge Network, 2002
- Johansen, P. 1999. Exploring Regional Statistics, In: CHI Workshop Interacting with Statistics 1999.
- Liddy, E.D. and Liddy, J.H. 2001. An NLP Approach for Improving Access to Statistical Information for the Masses. Paper presented to the FCSM 2001 Research Conference, November 2001 (Abstract)
- Marchionini, G. 1998. Advanced Interface Designs for the BLS Website: Final Report to the Bureau of Labor Statistics. (June, 1998).
- Marchionini, G. 1999. An Alternative Site Map Tool for the FedStats Website. In Final Report to the Bureau of Labor Statistics, (June 30, 1999).
- Marchionini, G., Hert, C., Liddy, E. and Shneiderman, B. [2000] Extending Understanding of Federal Statistics in Tables. *ACM Conference on Universal Usability*. (Washington, DC). p. 132-138.
- Marchionini, G., Hert, C., Shneiderman, B. and Liddy, E. 2001. E-tables: Non-Specialist Use and Understanding of Statistical Data. In *Proceedings of dg.o2001: National Conference for Digital Government*. (Los Angeles, May 21-23, 2001). p. 114-119.
- United Nations Statistical Commission and Economic Commission For Europe – Guidelines for Statistical Metadata on the Internet, Conference of European Statisticians Statistical Standards and Studies No. 52, Geneva, 2000.
- United Nations Statistical Commission and Economic Commission For Europe. 2001. Best Practices in Designing Websites for Dissemination of Statistics. In: Conference of European Statisticians Methodological Material, Geneva 2001.