CONFERENCE OF EUROPEAN STATISTICIANS

ECE Work Session on Statistical Output
for Dissemination to Information Media
(Geneva, Switzerland, 28-30 October 2002)

Topic (c): Use of new audio-visual, electronic and
print technologies in disseminating data to the media

EMERGING DESIGN STANDARDS FOR WEB DATA ACCESS TOOLS

Submitted by the U.S. Bureau of Justice Statistics[1]

Many statistical agencies are moving toward the creation of browser-based data access tools and interfaces that allow users to specify the content of the output. Unless developers of such data access tools, which can be very expensive to build, pay close attention to usability, the tools fail to help users find the data they need and may instead increase user frustration.

Several U.S. statistical agencies have been working to develop a set of usability standards for web-based interfaces to statistical databases. While newly developed browser gateways to data still need independent usability testing, designers who follow these standards during design of the tool should be able to achieve a fairly satisfactory degree of usability from the start.

This paper discusses the ten design guidelines or standards that have emerged in response to problems that users encounter when they attempt to use such tools. In addition, the Bureau of Justice Statistics website application, Crime & Justice Data Online, is used to demonstrate implementation of the guidelines.

The movement to electronic dissemination via the Internet

Before the Internet, statistical data was available to expert users in electronic formats for analysis with their own resources while less sophisticated users relied on printed reports. With the advent of the Internet, agencies used the publishing paradigm placing electronic versions of documents on the Internet. However, the data was locked in formats that limited the user's ability to manipulate the data and lost the computational properties of the numbers. As technology to interface databases via the Internet became available, many statistical agencies moved to create electronic data access tools to allow users to extract the data they needed directly from agency databases.

Problems with early data access tools

Suddenly there was a proliferation of information previously challenging to get to and tools to support self-serve access to it. However, the tools were often weak and had a narrow focus. Citizens using the tools were often frustrated. A review of the tools found that they were difficult to use, depended on an expert knowledge of the data, offered too many options to do some things and no options to do others, and frequently mimicked statistical (analytic) software more than being effective access tools.

As more tools were deployed, developers recognized the need for usability testing. The U.S. Bureau of Labor Statistics staff became involved in conducting usability testing and heuristic evaluations of a number of tools from several U.S. government statistical agencies including:

---

[1] Prepared by Marianne Zawitz

FERRETT - U.S. Census Bureau access tool -(has evolved to DataFerrett)
http://dataferrett.census.gov/TheDataWeb/index.html

LABSTAT (LABorSTATistics) - Bureau of Labor Statistics
http://stats.bls.gov/home.htm

FedStats "gateway" to statistics at over 100 agencies
http://www.fedstats.gov

CDC Wonder - Centers for Disease Control and Prevention access tool (studied for research purposes only)
http://wonder.cdc.gov/#aboutWonder

As reported by Bosley and Straub[2], citizens who used these tools:
1. Did not know how to search large data repositories efficiently.
2. Did not understand technical data definitions and data types/tokens that differ only slightly.
3. Lacked familiarity with multi-variate structure of data repository, and were overwhelmed by quantity and variety of accessible data.
4. Wanted to focus on small scope, personal questions rather than surveying the big picture.
5. Were unable to cope with gaps and other irregularities in data or adopted erroneous expectations and beliefs.
6. Were unable to accurately interpret the data or observe and extract meaning from patterns in the data - the data did not speak to them.

Design guidelines

Based on testing and reviewing of many U.S. statistical sites, BLS staff, in conjunction with Human Factors International, identified 10 points that, if followed, should greatly improve the usability of interactive statistical interfaces.   Usability testing is still recommended even if you follow all of the guidelines because users are the final arbiters of what works. However, following the guidelines should result in having less to fix.

Design Guidelines for Web Data Access Tools

Orient the user to the available body of data
1. *Give an overview of the available data*
Provide general orienting information about the data that can be accessed using the interface. Highlight indicators of data scope, restrictions, deliberate omissions and other important characteristics.

2. *Support situational awareness within the available data*
Use text or graphics to propagate the data structure across levels. Tell users when they enter a disjoint partition of the data. Make it easy for users to return to the initial state (screens which link to major divergent paths?)

3. *Display and clearly define metadata*
Embed sufficient metadata with tools to get users started. Provide easy, "just-in-time" access to definitions of technical or unfamiliar descriptive terms encountered as querying proceeds. Avoid using short, cryptic labels for data sets or variables.

---

[2] Bosley, John J. and Kathleen A. Straub, "Data exploration interfaces: Meaningful web database mining by non-statisticians", *Proceedings of the IBM Make IT Easy Conference.* 2002

Design the interface for interacting with the data

4. Put adequate and clear instructions on the interface

Tell users explicitly how to work with the interactive elements on the interface. Make the association between an interactive element and user guidance clear and available. Set defaults in all data specification "widgets;" defaulting to the broadest specification is recommended.

5. Link users to frequently requested analyses

Provide links to frequently requested numbers or datasets. Store common queries for novice users to modify to suit their needs. Build shortcuts for advanced users' quick data access.

6. Use simple interaction schemes to accomplish complex query-building

Use logical task sequences or natural language instructions to support advanced Boolean query syntax. Enable users to add or exclude data selections, and express optional inclusion ("or") by making a series of clear, discrete choices.

7. Summarize outcome of complex data specification for review and confirmation

When users apply multiple filters, especially across multiple screens, display the final specification for review and confirmation before a user submits a data request.

Help users anticipate, interpret and evaluate results

8. Offer choices of easy-to-interpret output formats

Offer users choices among well-known, understandable outputs like tables or simple graphs. Use graphics, or actual examples of output formats, as well as text to describe output options.

9. Design output formats to facilitate quick and reliable query validation

Make output labels consistent with variable selection options in query. Make labels clear and highly visible on statistical tables. Support keeping row and column labels visible as user explores table contents (large tables).

10. Help users avoid searching for non-existent or non-available data

Warn against, or actively prevent, requests for missing or unavailable data. Notify user when query will return a null result.

Source: Bosley, John J. and Kathleen A. Straub, "Data exploration interfaces: Meaningful web database mining by non-statisticians", *Proceedings of the IBM Make IT Easy Conference*. 2002

Bureau of Justice Statistics Crime & Justice Data Online Example

Bureau of Justice Statistics (BJS) Crime & Justice Data Online is an access tool for National, State and local data on –

- Crime trends
- Homicide trends and characteristics of victims and incidents
- Law enforcement agency characteristics

These data were selected for inclusion in the initial development of the tool because of user interest, the opportunity to link data among all of them, and the fact that they included both longitudinal and cross-sectional data. Usability testing was conducted on the tool during development.

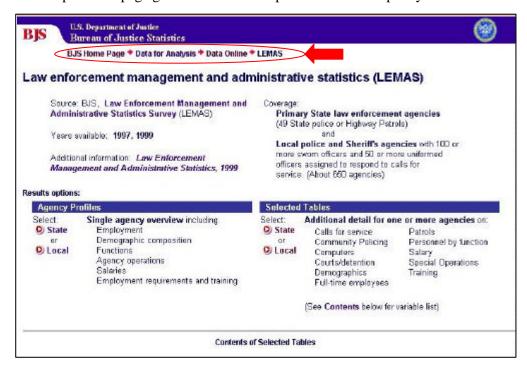**Guidelines - Orient the user to the available body of data**

1. Give an overview of the available data

BJS Crime & Justice Data Online opens with a list of data sources available and information about geography, population coverage, and time series covered. The scope of what is available is clearly presented.
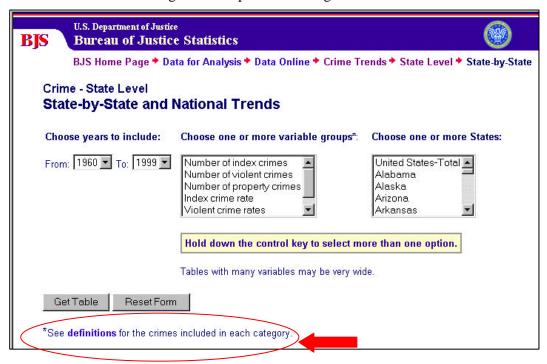

2. Support situational awareness within the available data

The crumbs at the top of each page give the user a clear picture of how deep they are in the site.
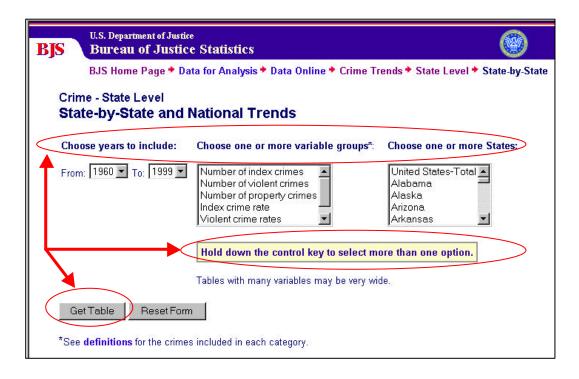
3. Display and clearly define metadata

Definitions of terms used for measuring crime are provided throughout.

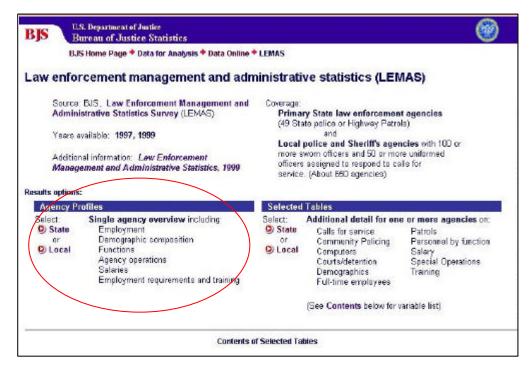### 4. Put adequate and clear instructions on the interface

The tool has few instructions. When used, they are written in clear, easy-to-understand language. User testing was required to make sure that users saw the instructions.

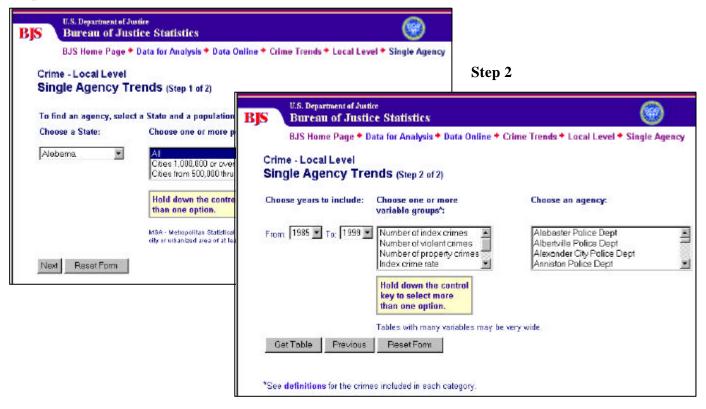**Guidelines - Design the interface for interacting with the data**

5. Link users to frequently requested analyses

BJS Crime & Justice Data Online provides the option to get an agency profile that has a summary of the most frequently requested statistics.

6. Use simple interaction schemes to accomplish complex query-building

Step 1

Step 2

7. Summarize the outcome of complex data specification for review and confirmation

BJS Crime & Justice Data Online presents three-dimensional data, which is hard for users to understand. Originally, the option descriptions were complete sentences. Through user testing, we saw users having a hard time choosing an output format. We determined that most users could not envision an output format from the descriptions. We highlighted the differences between the formats, used two word dimension descriptions in a list, and moved and simplified the examples to be with the descriptions. Testing of the revised design presented here resulted in greatly improved user performance.

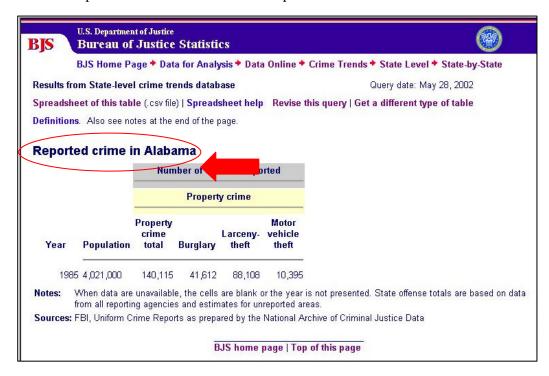**Guidelines - Help users anticipate, interpret and evaluate results**

8. Offer choices of easy-to-interpret output formats

BJS Data Online makes extensive use of examples of the output to assist the user in deciding what to select. Originally, the examples were further down the page and filled the horizontal space with data. We found that the large table examples intimidated most users. Using smaller extracts of the tables that could result from a query combined with the dimension descriptions aided users in getting what they wanted.
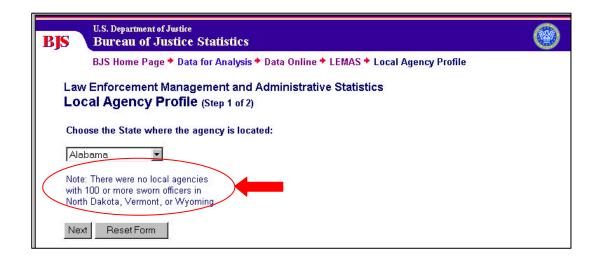
9. Design output formats to facilitate quick and reliable query validation

In user testing, we saw users scan the results to see if they got what they selected. Our original design did not clearly differentiate the selections like state or years from the other material on the page. We revised the results to emphasize those elements that were part of the selection and are not in the tables.

10. Help users avoid searching for non-existent or non-available data

BJS Data Online explicitly states when data are not available. By positioning the notes about data that are not available right below the appropriate drop down menu, users saw the information in context.

Even though the design guidelines were not yet drafted when BJS Crime & Justice Data Online was developed, it followed many of the guidelines. Specifically, it included:

- Simple two or three step interface.
- Descriptions of the data at beginning, links to definitions throughout, and additional metadata on results pages.
- Drill down to get user to smaller set of relevant data.
- "You are here" clues through "crumbs".
- Simple instructions where relevant.
- Proper use of "widgets" (Drop down menus, radio buttons, etc.).

In user testing, the application scored very well, but we found a few problems as outlined above.   We redesigned and retested several problematic features, which are now incorporated in the live site. Bosley and Straub applied the guidelines to the current site and found that it met most of the guidelines as demonstrated above. Clearly, by following the guidelines, we had much less to change as a result of user testing and could get real performance improvements quickly and easily because we needed to focus on only a few issues.

 The site was launched in November 2001 and has had over 80,000 visitors between November 2001 and the end of June 2002. BJS has received very positive responses about the application. The greatest evidence about the success of the site is that no one has asked a question about how to use the site.