The GovStat Statistical Interactive Glossary (SIG)

Ron T. Brown, Jesse Wilbur, Stephanie W. Haas, & Maria Cristina Pattuelli
School of Information and Library Science University of North Carolina at Chapel Hill
CB #3360, 100 Manning Hall Chapel Hill, NC 27599-3360
[ront, jdwilbur]@email.unc.edu, [haas, pattm]@ils.unc
Project URL: http://www.ils.unc.edu/~govstat

1. **Introduction**

The websites of U. S. statistical agencies provide statistical information on a wide range of American life, ranging from unemployment to crop prices to birth rates. Like other complex information resources, they are not always easy to use, especially for inexperienced or nonexpert users. One of the problems that people have in using statistical information is that the level of statistical knowledge (also known as statistical literacy) in the general population is low. People's lack of statistical knowledge is a barrier to many important tasks: searching for information, deciding which of several kinds of statistics they need, understanding the meaning of table, row, and column titles, understanding what a specific number means, interpreting the meaning of a time series, and so on.

In response to this need, a component of our NSF-funded project "Integration of Data and Interfaces to Enhance Human Understanding of Government Statistics: Toward the National Statistical Knowledge Network", also known as GovStat, (Marchionini et al., 2003) is the Statistical Interactive Glossary (SIG). The overall goal of the glossary is to help users understand important statistical terms and concepts in the context in which they are used. When users look for information on an agency website, their purpose is to answer a question or solve a problem, not to learn more about statistics. Understanding one or more statistical concepts or terms may be a necessary part of accomplishing their tasks, but should not itself become a major task. One approach that avoids pulling users away from their primary information tasks to learn about statistical terms, is to integrate glossary tools as seamlessly as possible into the statistical resources themselves.

2. Design Criteria

We have established several criteria for the design of SIG.

1. Its coverage should be limited to concepts and terms that users will encounter on the agency websites or that are crucial to understanding frequently-used statistics.
2. Its audience should be "everyday users", not statistical experts. We do not seek to turn people into experts, but rather to give them the information they need to complete their tasks successfully.
3. It should provide explanations in a variety of forms to accommodate users' learning styles, existing knowledge, interest, motivation, and persistence. Users should be able to choose the style of explanation that they find most effective.
4. The explanations must be attractive. The most sophisticated help tools are useless if people do not look at them.
5. The explanations should be coordinated with the content of the websites, to make statistical help an integral part of the users' information seeking experience.

Additional information on SIG and its supporting ontology may be found in Haas, Pattuelli, & Brown (in review).

3. Structure and Content
Each term has several presentations associated with it, which vary in three dimensions.

- Content _definitions, examples, brief tutorials, demonstrations, interactive simulations or a combination of these.
- Format _ text, text plus audio narration, still images, animation, interactive presentations
- Context specificity _ refers to ways in which the explanation is tailored to the information environment in which the user encounters the term. A context-specific explanation is associated with a single statistic or table (e.g., Consumer Price Index). An agency-specific explanation would incorporate entities used in the agency. A universally applicable explanation could be presented anywhere the term was used. Users can invoke explanations of terms directly from agency web pages.

Our demonstration will show presentations for index (including specific types of index such as Consumer Price Index), and adjustment (including specific types of adjustment such as seasonal adjustment). We will show several presentations for each term, illustrating different combinations of content, format, and context specificity. An example demonstration for index might include the universal definition of index in text, text plus audio, still images, and animations.

4. Development
In the future we will conduct user testing to determine: (1) where the explanations should be attached, (2) how to let users know that explanations are available, and (3) how to let users choose which explanation he or she would like to see.

References

Haas, Stephanie W., Pattuelli, Maria Cristina, & Brown, Ron T. (in review).Understanding Statistical Concepts and Terms in Context: The GovStat Ontology and the Statistical Interactive Glossary. Submitted to the ASIST 2003 Annual Conference.

Marchionini, G. , Haas, Stephanie W., Plaisant, Catherine, Schneiderman, Ben, & Hert, Carol (2003). Toward the Statistical Knowledge Network. Proceedings of dg.o2003.