

Distr.
GENERAL

ECE/CES/SEM.54/17
12 May 2006

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANIZATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD Seminar on the Management of Statistical Information Systems (MSIS)
Sofia, Bulgaria, 21-23 June 2006

Topic (ii): Dissemination and client relations

STATISTICS IN CYBERSPACE: HOW HAS THE INTERNET CHANGED THE U.S. BUREAU OF LABOR STATISTICS?

**Invited Paper prepared by Michael D. Levi,
U.S. Bureau of Labor Statistics**

I. Introduction

1. The history of new media is a history of transformation. The very first movies, for example, were straightforward recordings of live theater. As the industry matured, however, filmmakers discovered and exploited possibilities impossible or impractical to render on stage. Effects ranging from close-ups and point-of-view photography to slow motion, animation, and computer-generated imagery revolutionized the vocabulary of cinematic expression. At the same time the spontaneity and intimacy of live performances was lost. As the new medium allowed stories to be experienced by a tremendously expanded audience, it also changed the nature of the stories being told.

2. Today, the Internet gives statistical agencies an unprecedented capacity to reach a huge audience very quickly. The Internet is, however, more than simply an effective broadcast mechanism. It is an entirely new communications medium that differs in fundamental ways – perceptual and cognitive -- from traditional paper-based publications.¹

3. Over the past ten to fifteen years, statistical agencies have increasingly moved their data dissemination activities from being primarily paper-based to being primarily web-based. Thus it

¹ For a more thorough discussion of paper vs. computer screens, albeit in a slightly different context, see Michael D. Levi, "Literature at the Human-Computer Seam" (Modern Language Association, 2000) <http://www.bls.gov/ore/pdf/st000100.pdf>

is worth asking if, and how, this change in medium is affecting the statistical products themselves, and the agencies that produce them. This paper will examine the economic and statistical outputs created by the U.S. Bureau of Labor Statistics (BLS), looking for changes that have developed since BLS went online, in an attempt to begin answering this question.

II. Core Function of the Agency

4. Despite the hype, the adoption of Internet technologies has not changed the essential nature of BLS. The agency's charge remains the same: "To acquire and diffuse among the people of the United States useful information on subjects connected with labor, in the most general and comprehensive sense of that word..."² As it has since 1884, BLS collects establishment and household-level data from individual respondents, aggregates the microdata into higher-level estimates, and publishes the results. No surveys or programs have been added or dropped because of the web, nor has the agency's budget or organizational structure changed substantively as a result of the Internet.

5. Though not the focus of this paper, it is worth mentioning that the business processes used to collect, estimate, and publish economic data, as well as those that administer the workings of the agency, have changed significantly due to – increasingly web-based – automation. Thus Internet data collection (via e-mail, file transfer, and the web) is taking off; automated review of preliminary estimates is universal; research is assisted by electronic journals and academic working papers; and staff have ready access to agency directives, best practices, administrative procedures, and productivity aids such as on-line training, conference room reservations, and personnel directories. The efficiencies achieved by this automation translate into more working hours that can be devoted to improving survey processing, performing deeper analyses, and creating innovative new products for our customers.

III. More Information

6. Perhaps the most obvious effect of the Internet on BLS products is that the agency publishes much more information than it ever did before. As of January 2006, the BLS website contained over 55,000 web pages and 20,000 FTP files. These include statistical tables, analytic documents, textual summaries, research papers, maps, graphs, and their supporting infrastructure: navigation pages and explanatory material.

7. The explosion of products is due primarily to the relative ease and negligible marginal cost of adding one more item to the online offerings. Products such as working tables originally intended primarily for internal use, and research papers previously available to the public only on demand, are now routinely posted to the website. What's more, custom tabulations developed in response to an individual customer request now typically find their way onto the web, again because it is easy to do and costs very little.

8. Previous print publications accounted for only a small fraction of the material that is currently available to any user who can navigate through the wealth of data and information.

² United States Code, Title 29, Chapter 1, Section 1

IV. Communication with Customers

9. The content and style of interactions with outside data consumers has changed quite radically. As one analyst who has been at BLS for considerable time describes the change:

When I began working as a [Current Employment Statistics] analyst in 1986, information duty was a significant part of the job. On request from the public, we ran custom table print jobs and mailed out listings. We read data to people over the phone. And on CES release Friday, everyone would man the eight or nine information line phones that would start ringing before 8:30, with none picked up until someone said “Time!”

When the data became available via the web, we trained users to get the data themselves. Press release Friday became a non-event. Information duty now consists much more of problem solving and redirection of users lost in the BLS information jungle.³

10. The nature of customer inquiries has shifted from “What is a particular number?” to “What does that number mean?” or “I’m trying to answer a particular question – how can you help me?” In other words, the queries received and the answers given in response have become considerably more sophisticated, drawing heavily on subject-matter expertise rather than clerical acumen.

11. Whereas in the past most communication took place either by postal mail or over the telephone, today the overwhelming majority of inquiries are handled via e-mail. E-mail is much more efficient than the postal service, can be both sent and responded to at the convenience of the writer, and allows for much shorter turn-around times than does the postal service.

12. E-mail also avoids some of the frustrations associated with the telephone, such as after-hours unavailability or being placed on hold during high-volume periods. E-mail is, however, a significantly less personal style of communication than direct conversation. In the past BLS information specialists often knew many of their regular users and developed on-going relationships that not only helped the users but also gave the agency valuable insights into the customer community. Electronic communications make it more difficult for analysts to understand needs that a user may not be able to articulate clearly, to guide users to the data that may best serve their needs, even to understand the context of questions. Thus, alongside its many advantages, e-mail also distances the agency from its customer base in unfortunate ways.

V. Educational Materials

13. In large part to address confusion among users overwhelmed by the quantity of BLS materials, and to compensate for the decrease in direct personal interaction between BLS information specialists and the public, the agency has greatly expanded its offering of explanatory and educational materials. These range from survey overviews and cross-walks between related data sets, to the BLS *Handbook of Methods*, to *The Editor’s Desk* (a daily publication which highlights specific BLS data by focusing on one or two salient points of a

³ David Hiles, Chief, Branch of Current Data Analysis, Quarterly Census of Employment and Wages. Personal communication; March 7, 2006.

report rather than presenting a full analysis) to a *Career Information for Kids* sub-site aimed at students in grades 4–8.

14. The goal of BLS' educational offerings is to provide appropriate technical guidance to a population with a wide range of statistical and economic sophistication. Thus some materials are written for a lay audience in common English, while others assume familiarity and ease with advanced mathematical concepts and are written in correspondingly specialized language.

15. In addition to the bread-and-butter economic statistics that are the agency's core product, BLS also publishes a great deal of organizational and administrative material on the web. Documents such as the official agency mission and strategic plan, listings of senior management officials and advisory bodies, guidelines on information quality and methodology, etc. were previously available only upon request, if at all. These have now all been posted to the website and, though usage of the "About BLS" sub-site is small relative to survey and program products, they do attract a steady stream of readers.

VI. Theoretically Possible but not Previously Practical⁴

16. Statistical tables are information-rich digests of data. Subject-matter experts select a subset of estimates from a larger pool and arrange them into rows and columns intended to illuminate some particular aspect of a given subject, juxtaposing series to tell a story. The advantage of agency-created tables is that they typically save the user from the labor and proficiency-intensive task of sifting through a huge collection of often disparate data items to gather the numbers they need.

17. Not even the most experienced group of analysts, however, can anticipate every question a member of the public might ask. To supplement agency-created tables, BLS now makes its database of record – the full store of approximately 200 million estimates meeting publication standards going back to 1913 – directly available to the public through a collection of interactive query tools. These applications not only retrieve the estimates in a variety of formats but also permit the user to graph time series and, in some cases, generate choropleth maps. The public-use microdata from one of the agency's most sought-after surveys, the Current Population Survey, is also available to the public from a joint BLS and Bureau of the Census website.

18. Print publications have never contained all the estimates produced by the Bureau of Labor Statistics. It would have been possible to hand-enter data from printed pages into a spreadsheet or database for further analysis, but the time and effort would have been staggering. Prior to the web, interested researchers were able to order time series data and public use microdata on computer tapes but, again, the time, effort, and expense, not to mention the requirements for access to personal computing capabilities, limited such orders to a small group of individuals and organizations. The web brings this power into the hand of every user with a browser.

⁴ Many of the ideas in this paper were sparked by a conversation with Phil Rones, Deputy Commissioner of Labor Statistics. The phrase "theoretically possible but not practical" is Mr. Rones'.

VII. Analytic Tools

19. Making all agency time series data directly accessible marks a huge step forward in customer service. It is not, however, sufficient. Many members of the public who are interested in the economic statistics that BLS produces lack the statistical software or knowledge to perform relevant calculations for themselves.

20. In response to this, BLS has developed a number of applications, including an inflation calculator, a location quotient calculator, and an illness and injuries rate calculator, which permit end-users to perform economic or statistical analyses directly from the agency website. A wage calculator also adds the ability to generate model-based estimates to supplement aggregates tabulated directly from the collected respondent data.⁵

VIII. Cross-Survey Tables

21. Internally, BLS is organized by programs, and institutional stovepipes have proved difficult to overcome. Each program or survey is responsible for developing its own publication products; with rare exceptions each program generates tables and text that relate exclusively to its own data.

22. Many end-users, however, are interested in comparing and contrasting data from multiple programs. In the past such users were forced to manually assemble content from multiple sources (even if all from within a single institution.) The creation of the BLS website highlighted this shortcoming, as users were given a single entry point into the full range of agency programs, and the entire site had a consistent structure and appearance. Thus users, legitimately, began to view the agency as a single entity rather than as a collection of disparate programs, and began to ask for better integration of the products in this repository.

23. Such customer requests, aided by the existence of a central database, gave rise to cross-survey tabulations organized first around geographic regions (states, counties, and metropolitan areas) and then around industry groups. These “Economy at a Glance” and “Regional Resources” compilations rapidly became one of the highest-volume areas of the website, and have remained extremely popular. The structure of most of these tables has remained fairly static, with new data flowing into the appropriate cells as it is released, but some cross-survey compilations have been created in response to breaking events such as the Katrina hurricane in 2005.

IX. Accommodating Perceptual Differences in Media

24. Cognitive psychologists at BLS are doing interesting work related to human factors and computer-based data dissemination. For the most part, however, this work has centered on interface design and navigation through large information spaces rather than on issues more narrowly surrounding statistical tables and analytic text.

25. One unexpected finding, however, has been implemented in practice. In the mid-1990s an early usability test asked representative end-users to find a particular estimate on the BLS site,

⁵ These calculators can all be reached from <http://www.bls.gov/data/home.htm>.

an estimate that was contained in one particular table. This test was intended to evaluate a prototype menu structure to determine whether users could understand and follow the taxonomy of categories the information architects had established. To indicate that they had found the correct cell, test subjects were asked to write the cell's value on an answer sheet.

26. A surprising number of subjects wrote down the wrong number. At first this error was attributed to a failure in navigation, but further analysis determined that most subjects had indeed found the correct table, they simply wrote down the value from an incorrect cell. Typically the erroneous value was within one or two cells horizontally or vertically from the intended target.

27. The experimenters concluded that users have great difficulty retaining a sense of position within a large two-dimensional array of numbers on a computer screen (on paper many users employ a finger or straight edge to track their location within a table, but this strategy is ergonomically much more difficult on a computer screen.) Since then, BLS tables intended for distribution over the web have alternate rows shaded in a different color to facilitate tracking, much like the old "green-bar" mainframe listings once did.

X. Conclusion

28. To date, BLS finds itself closer to the early days of film than to a mature cinematographic industry. Internet-related changes in statistical outputs are primarily ones of quantity rather than essence. We are still, metaphorically, filming live theater.

29. Nonetheless, a few qualitative differences are beginning to appear. Though in most respects quite trivial, the introduction of shaded rows in wide tables marks the first explicit recognition that a computer screen displaying a web page requires an approach that takes a different set of human factors into account.

30. Cross-program data compilations were certainly possible before the web, but were rarely created. It took a new audience and their perception of (and demand for) a unified agency to initiate this set of products.

31. Most significant is an increasing and absolutely fundamental shift in the locus of control. Whereas previously BLS, as the source agency, determined which data would be readily available, defined how its data appeared, and acted as the gatekeeper of expert knowledge, the web has moved more and more detail and analytic capacity directly into the hands of our customers.

32. Accompanying the steep rise in the number of data users is a considerable decline in direct, personal contact between BLS staff and our customers. Such a change in organizational culture carries with it the loss of a valuable source of immediate feedback.

33. BLS is no longer a solitary author telling the story inherent in the numbers. Instead BLS is working to help users find the stories for themselves.
