

Session 5

ONE-STOP SHOPPING FOR FEDERAL STATISTICS

One-stop Shopping for Federal Statistics

Alan R. Tupek

National Science Foundation

Background

It is difficult for the general public, and even frequent data users such as social science researchers, to know about and to access the extensive amount of statistical information produced by the decentralized U.S. federal statistical system. Publications, such as the *Statistical Abstract of the U.S.*, have provided access to a limited amount of statistics from the federal statistical system. With the growing acceptance of the World Wide Web (WWW), there is an opportunity to provide access to the broad range of statistics of interest to the public. Federal agencies have, for the most part, made tremendous progress in developing easy access to their data and databases through the WWW. Data users accessing information from one Federal agency may even find out about related statistics available from other agencies. However, these efforts have been very limited to date. The technological barriers no longer exist that once made it difficult to have a coordinated effort to allow users to locate the statistics that are most appropriate for their uses.

In May 1996, the Office of Science and Technology Policy (OSTP) and the Office of Management and Budget (OMB), along with representatives of the major statistical agencies released the Economic Statistics and Social Statistics Briefing Rooms that can be accessed through the White House WWW Home Page. This initiative provides the general public easy access to key current economic and social statistics and to the approximately 15 agencies that provide these statistics.

A Task Force on One-stop Shopping for Federal Statistics was formed under the auspices of the Interagency Council on Statistical Policy (ICSP) in November 1995. The Task Force is chaired by the National Science Foundation, with representatives from the Bureau of the Census, the Bureau of Labor Statistics, the Bureau of Economic Analysis, the Economic and Statistics Administration, the Energy Information Administration, and *ex officio*, the Office of Management and Budget.

The Task Force was formed as a culmination of several activities. These included the early development stages of the Briefing Rooms, a November 1995 OMB report on Electronic Dissemination of Statistical Data, meetings of an ad hoc interagency group on dissemination of statistics on the Internet, as well as initiatives of the

Office of Statistical Policy at OMB. The Briefing Rooms would provide access to a limited amount of Federal statistics, but would also provide an avenue to access the larger array of statistical information provided by the Federal Government. The report on Electronic Dissemination of Statistical Data provided a guide to survey managers on issues to consider when expanding the modes of data dissemination. The report highlighted the broad range of dissemination modes used by Federal agencies and drew attention to the need for better coordination of dissemination practices. The ad hoc interagency group on dissemination of statistics on the Internet had been monitoring the various practices agencies were developing for Internet dissemination. This group was especially concerned with dissemination of metadata and also saw the need for interagency cooperation to provide easier access to all Federal statistics. Finally, but certainly not least, monthly meetings of the Heads of Statistical Agencies chaired by the Statistical Policy Office at OMB had just been sanctioned as the Interagency Council on Statistical Policy (ICSP). As one of its first official acts, the ICSP formed the Task Force on One-stop Shopping for Federal Statistics.

Task Force Charge

The initial charge to the Task Force was to examine the Federal Government's various systems for disseminating statistical data, and to provide recommendations to the OMB on the structure and implementation of new one-stop shopping mechanisms for users of Federal statistics. The charge to the Task Force includes: (1) identifying the major statistical series to be incorporated into the One-stop Shopping mechanism; (2) identifying and developing new concepts, structures, and methodologies to facilitate the dissemination of Federal statistics; (3) developing and testing a mechanism to provide one-stop shopping for users of Federal statistics; (4) planning and implementation of the new one-stop shopping mechanism; and (5) ensuring that there is ample opportunity for widespread public participation of stakeholders in the development process.

As one approach to its charge, the Task Force should oversee the design, development, implementation, and maintenance that would provide easy access to Federal statistics through the WWW. The Task Force will develop and monitor contracts or other mechanisms to develop, implement, and maintain the agreed upon approach.

The Task Force should adopt processes that ensure ample opportunity for public participation. These processes should involve all stakeholders, including the range of Federal data users, both government and private, as well as data collectors and data providers. The Task Force should consider forming a consultation group, composed of Federal agencies not represented on the Task Force. Such a group would meet on a flow basis, as necessary, to provide input to the work of the Task Force. Notice of the Task Force's work should be widespread and should be

published in the Federal Register for all interested public and private parties. Interested parties will be given the opportunity to be included on a mailing list.

Since its inception, the Task Force has solicited opinions from various communities on its prototype development. Task Force members have made presentations on the One-stop development work in a variety of forums, including meetings of the Association of Public Data Users, the Council on Professional Associations on Federal Statistics, the Federal Web Consortium, and the American Statistical Association. In September 1996, the Task Force asked each of the agencies on the ICSP to appoint a liaison to the Task Force. The liaisons have been asked to provide input to the prototype development of the One-stop site.

Features of the Prototype Web Site

The purpose of the WWW site is to permit easy access via an initial point of entry to the wide array of Federal statistics of interest to the public. The site is intended to supplement the excellent WWW sites that individual agencies have developed for disseminating Federal statistics. Other currently available WWW services do not adequately meet the objectives of One-stop Shopping for Federal Statistics. For example, the Economic Statistics and Social Statistics Briefing Rooms provide access to about 100 key statistics from a dozen or so Federal agencies. STAT-USA does not cover the full array of Federal statistics and its fees may limit the potential audience. A few academic sites provide access to some of the Federal statistics that are produced by the larger statistical agencies. It is also possible to find a limited amount of Federal statistics through Government and commercial WWW search services, such as GovBot, Excite, Yahoo, and Infoseek. However, using these services to find Federal statistics often results in many more links to non-relevant information than to links to Federal statistics. It is especially hard to find Federal statistics from those agencies whose entire mission is not statistics. The primary objective of the WWW site for One-stop Shopping for Federal Statistics is to help users find the information they need without having to know and understand how the decentralized Federal statistical system works. The WWW site for One-stop Shopping is intended to help guide users to statistics from the 70 or so agencies that provide statistics of interest to the public. It is expected that most of the centralized information WWW sites mentioned above will be able to use One-stop Shopping for Federal Statistics to improve their services.

The prototype site includes the following features --

Subjects A to Z - A table of contents approach to the wide-range of statistics available from Federal agencies. The list of subjects allows users to see the various sources of statistics for each subject area. Links are provided to the statistics on agency WWW sites.

Regional Statistics - A collection of agency WWW sites that provide easy access to state, metropolitan area, and other sub-U.S. geographical statistical information. Most agencies provide a clickable map approach to access these data.

Agencies that Provide Statistics - A list of all agencies that provide statistics with links to the agencies WWW Home Page and to the statistics by subject area for each agency.

Statistical Programs of the U.S. Government - Adapted from the OMB report with the same name, this feature provides links to statistics available from federal agencies within 10 broad topical areas.

Subject Matter Contacts - A collection of agency WWW sites that provides contact names, telephone numbers, and e-mail addresses for questions about statistics.

Statistical Press Releases - A collection of agency WWW sites that includes the latest statistical news releases.

Statistical Policy - Includes links to Federal budget documents, statistical policy working papers, and selected federal register notices.

Site Maintenance

The members of the ICSP recently agreed to continue the development and maintenance of a WWW site for One-stop Shopping for Federal Statistics. The ongoing development and maintenance of the site will be funded through an annual, renewable interagency agreement between the agencies represented on the ICSP and the Bureau of the Census. Listed below are the responsibilities of the organizations involved in this agreement.

Responsibilities of the Bureau of the Census (BOC)

The BOC is responsible for the continuing development and ongoing maintenance of the WWW site for One-stop Shopping for Federal Statistics. Enhancements to the site will be made as directed by the Task Force on One-stop Shopping, operating under the aegis of the Interagency Council on Statistical Policy. The BOC will also be responsible for hardware and software maintenance and upgrades. The BOC will provide –

a separate server for the WWW site with the domain name
<http://www.fedstats.gov>;

staff to maintain the server and up-to-date links to Federal statistics and to enhance the site as directed;

usage statistics to allow the Task Force to determine traffic to and from the site and other information needed to assess the usefulness of the site;

a limited feedback mechanism to report broken links and to suggest improvements. Users will be directed to contact the appropriate agency for replies to all other questions.

Responsibilities of the Task Force on One-stop Shopping for Federal Statistics

The Task Force will direct the maintenance and enhancements to the site and maintain contact with agency liaisons to determine the need for maintenance and enhancements.

The Task Force will monitor the usage of the site and develop and implement a public information plan for the site in consultation with the ICSP.

The Task Force will provide periodic status reports to the ICSP.

The Task Force will provide an assessment report to the ICSP at the end of FY 1997. The assessment report will include a summary of the usage of the site, the resources expended to maintain and enhance the site, the enhancements made to the site during FY 1997, the amount of and type of feedback received, and recommendations for FY 1998, including funding options.

Responsibilities of the ICSP Agencies

Each agency will appoint an agency liaison to the Task Force on One-stop Shopping for Federal Statistics. The liaison will provide guidance on changes in an agency's WWW site that would need to be updated on the One-stop site. The liaison will also be responsible for suggesting improvements to the site.

Each agency will provide a link to One-stop Shopping for Federal Statistics on its home page.

Each agency will provide funding to support the ongoing development and maintenance of the One-stop site. The five largest Federal statistical agencies (based on FY 1997 statistical program budgets) will contribute \$20,000 each to the Bureau of the Census on an annual basis through an interagency agreement. These agencies are the Bureau of the Census, the Bureau of Labor Statistics, the National Agricultural Statistics Service, the National Center of Education Statistics, and the National Center for Health Statistics. The contribution from the Bureau of the Census will be to provide staff resources in addition to those funded through the interagency agreements. The other agencies represented on the ICSP will contribute \$10,000 each to the Bureau of the Census on an annual basis through an interagency agreement.

The funds will be used by the BOC to support one GS-13 (100%), one GS-05/07 (50%).

In addition, the Bureau of Labor Statistics has agreed to conduct usability testing of the site with library science researchers. Results of this work will be implemented, as directed by the Task Force on One-stop Shopping for Federal Statistics.

Also, the National Science Foundation has agreed to purchase the initial hardware and software needed for the site at a cost of approximately \$23,000.

FY 1997 Plans and Activities

The site is expected to be publicly available in February 1997. During FY 1997, the Task Force on One-stop Shopping for Federal Statistics plans to direct the following activities --

set up hardware, install software, procure upgraded hardware and software as needed;

through interactions with agency liaisons, identify and implement appropriate links for the one-stop site tools, "Federal Agencies That Provide Statistics," and "Subjects A to Z;"

provide access to agencies' data maps (regional statistics) and press releases;

design and develop usage statistics software and reports;

conduct routine site maintenance through interactions with agency liaisons and evaluation of reported problems;

procure, install, test, and release automated searching software which will allow users to enter their own keywords for searching for Federal statistics across agencies,

provide links to publicly available information about statistical research efforts; and

develop access tools for data maps, press releases, and statistical research that permit cross-agency searching for information.

Future Enhancements

Future enhancements will depend on an analysis of the site usage, user feedback, and the results of the usability testing of the site. Possible new features include: a data dictionary or glossary that describes differences between various terms, thematic search tools, statistical hypernews that allows users to participate in dialogues on statistical issues, and interagency coordination on database tabulation and extraction tools and user interfaces.

Acknowledgments

The members of the Task Force on One-stop Shopping for Federal Statistics are: Paul Bugg (Office of Management and Budget), Connie DiCesare (Bureau of Labor Statistics), Valerie Gregg (Bureau of the Census), Daniel Larkins (Bureau of Economic Analysis), Ken Rogers (STAT-USA), Rachael Taylor (Bureau of the Census), Al Tupek (National Science Foundation), and John Weiner (Energy Information Administration). Others who have made significant contributions to the work of the Task Force include Cherie Vaden and Marion Brady of the Bureau of the Census, Michael Levi and Deborah Klein of the Bureau of Labor Statistics, and John Wells of the Bureau of Economic Analysis. The development of the One-stop site would not have been possible without the leadership of Katherine Wallman, Office of Management and Budget and Chair of the Interagency Council on Statistical Policy.

One-Stop Shopping for Federal Statistics

Discussant: Ann Gray, Cornell University

I am here as the Association of Public Data Users (APDU) representative in place of Joe Salvo who could not attend today. At APDU's recent Annual Conference we discussed this initiative along with many other issues related to electronic delivery of federal statistics. My remarks will draw on the APDU conference. They are intended to be comments, not criticisms. This Task Force has a very ambitious charge and very little resources. They are going to need all the help they can get.

APDU members are very concerned about a group of persons that can be called the "general public" -- persons who lack knowledge of the U.S. federal statistical agencies. We believe that this group should be able to obtain the best measures they need without undue effort. Toward that end, One Stop Shopping should incorporate features that allow subject or topical access to federal data using simplified but accurate terminology. This would include features such as cross references and indexes. In fact, such products already exist both in print and in electronic formats. County and City Databook and Statistical Abstracts of the United States are proven to be useful and, in the case of Statistical Abstracts, fairly complete compendia of national level statistics.

I would urge Census to allow free access to Statistical Abstracts, even though they also market it on CD-ROM and charge for it via STAT-USA. Free, on-line use of the product may encourage the purchase of the CD-ROM as more people learn of its existence and the scope of the product.

Another APDU concern deals with the mis-use of data. This concern is not expressed in the notion that access should be limited, but rather that we need to recognize that people do make mistakes. One Stop Shopping should make sure that if mistakes are made, it is done by the end-user and not by the agency or provider. There is a standard disclaimer regarding data distribution that includes statements like "we bear no responsibility for the uses or interpretations or inferences you might make," but what is needed here is clear and ample information about the origin and history of the statistics that are presented.

APDU also encourages the inclusion of multiple levels of user support, including names, telephone numbers or electronic mail addresses of a real live person. This is useful for both novice and sophisticated data users. APDU also believes that the system should include the names and telephone numbers of local experts. Often people want local information that may not be readily available from federal agencies. There are people that are willing to be included in this list. The search of local experts could begin with those associated with the Census Bureau's State Data Center/Business and Industry Data Center program and the Depository Library Program. There may be other existing networks, such as the agricultural extension agencies, that could be brought into the system. I would also add that the media is an important intermediary in providing statistics to the public. They should be remembered in the design of the system.

Next, we need to consider the project for what it is and what it is not. It is not the democratization of information. The Depository Library Program, where everyone has equal access to government information distributed to the depository library, is democratization of information. Not everyone has equal access to the Internet, there are still problems with connectivity and we may never have universal access. This is an extension of the Depository Library Program and, in fact, not every community has a depository library.

This project is an attempt to create a single or virtual federal statistical agency. It should be possible, but there are massive intellectual problems. The success of this project depends upon solving those intellectual problems. It is not a technological problem, no matter how well technology may serve as a network or means of bringing things together. That is to say, the challenge is to organize the information in a way that makes sense.

The work of libraries is to organize information for retrieval. In order to locate information, control must be established. This control is made possible by a set of rules. We call them standards. These standards work in libraries and they work in technology also. In looking at other inter-agency or international efforts we see standards at work.

For example, the International Monetary Fund has set up a bulletin board -- it is really a web implementation -- dealing with national accounts and other economic data from various nations. There is a Special Data Dissemination Standard (copies are available from IMF: Telephone (202) 623 4415, FAX (202) 623 6165, Web: www.imf.org) that covers content, access, integrity, and quality. IMF created a voluntary standard where IMF acts as a centralized source for information about a nation's statistical products, including release dates, media, cost, and other issues dealing with availability. At this time about forty countries have subscribed to the standard. IMF enforces a consistent set of information about the statistics included in the bulletin board. Of course, it must be noted that IMF has a certain amount of influence that the Task Force on One Stop Shopping may not have and that IMF also limits the scope of the coverage for the bulletin board.

In another international effort, the Council of European Social Science Data Archives (CESSDA) has created an Integrated Data Catalogue that is available via the web (www.nad.uib.no/Cessda). Each data archive maintains its own catalog of data holdings and these catalog records include descriptive information about the contents of the data. Integration occurs through the use of field based WAIS indexing. WAIS incorporates a communication standard called Z39.50 that is widely used by libraries to integrate database functions.

Another effort to normalize information about data is being undertaken by a group headed by Merrill Shanks and Richard Rockwell. It is based on the use of SGML to allow automated systems to manipulate tagged text. SGML is also a standard and is, not incidentally, Z39.50 compliant.

Why are standards important? Standards are important because they allow for growth. The web itself is based on a standard and because there was a standard, the web has been able to grow into what we see today. The wide-spread availability of CD-ROM readers and CD-ROM products was made possible by a standard. The Federal Information Processing standards, especially the FIPS Codes for geographic and governmental units makes it possible to find a common identity for creating compendia of statistics about those areas from different sources. Likewise, there is a government standard for identifying government information: the Government Information Location System or GILS. If there is something wrong with GILS, fix it. Don't just ignore it.

Finally I'd like to offer some personal advice on the use of so-called "advanced" technologies and software systems based on the creation of "knowledge bases." It would be a costly waste of time to construct an enormous knowledge based system that will be obsolete by the time it is completed. What is needed to make this system successful is human intelligence and human intervention.

Council of Professional Associations on Federal Statistics
Seminar on Statistical Methodology in the Public Service

Session 5, "One-Stop Shopping for Federal Statistics"
Discussion of presentations by Alan Tupek (National Science Foundation)
and Valerie Gregg (U.S. Bureau of the Census)

Lauris Olson, Van Pelt Library, University of Pennsylvania

As I am merely a reference librarian at a large urban research university, I will restrict my remarks to the actual Internet resource (and its various parts) being presented today, leaving policy concerns to more qualified parties.

How useful is the "One-Stop Shopping" web site? Consider what's probably the federal government's single most popular and widely-used reference work, *Statistical Abstract of the United States*. The reader opens *Stat Abs* and is immediately confronted with numbers: statistical tables grouped in broad subject chapters. It never fails to surprise me that so many of our readers are satisfied with the tables they find in *Stat Abs*, or, if they're not fully satisfied, accept them with a grudging, "They'll do".

Of course, *Stat Abs* is an "abstract" in the bibliographic sense, its tables serving to indicate the contents of more expansive or detailed statistical publications through the source notes appended to each table. And *Stat Abs*'s wonderful appendix, "Guide to Sources of Statistics", serves as a subject-arranged bibliography for the whole volume by listing important U.S. statistical information resources.

An Internet counterpart to the statistical tables of *Stat Abs* has been available for some time as the "Federal Statistical Briefing Rooms" available through the White House's WWW site. As presently configured, FSBR provides current economic indicators and social statistics, sometimes with charts depicting indicators through time. Each indicator includes an hypertext source note, adding value to the presentation by allowing users to link from FSBR to the producing agency's relevant WWW page.

The "One-Stop Shopping" web site presented today seems most successful when viewed as analogous to *Stat Abs*'s "back of the book". That is, the web site is intended to provide an integrated access point to the wide range of federally-produced statistical programs and related products. Thanks to hypertext, the web site transcends *Stat Abs* by linking users immediately to data resources and documentation.

But to continue the *Stat Abs* / FSBR-"One-Stop Shopping" comparison is to show the latter's weakness. *Stat Abs* presents statistics gathered by non-federal sources; its "Guide to Sources of Statistics" includes federal and "other" statistical publications. The "One-Stop Shopping" Task Force's charge reasonably excluded these non-government data resources. But data dissemination partnerships within the federal government and between federal producers and nongovernmental distributors are proliferating. It's reasonable to link to "1990 Census Lookup" and other decennial census resources developed at Lawrence Berkeley National Laboratory, or to Cornell University's USDA statistical site. ICPSR now provides public access to the National Archive of Criminal Justice Data; CIESIN's Ulysses server pioneered decennial census microdata access; Right-To-Know Net has constructed interfaces for HMDA and other data sets. I worry that tracking these, let alone the indicators recently devolved from the Census Bureau to the Conference Board, are beyond the scope of the Task Force's charge.

The "One-Stop Shopping" web site will become one among several starting points for identifying and locating federal information. Most of these web sites are barren directory lists of agency web sites. "One-Stop Shopping" will join the handful of web sites providing multiple access tools. An instructive comparison can be made between "One-Stop Shopping" and the U.S. Government Printing Office's "GPO Pathway Services", which strikes me as possessing the best developed toolkit for finding federal information on the web.

Both "One-Stop Shopping" and "GPO Pathway Services" offer links to resources through subject arrangements. Of course, the GPO's web site attempts coverage of all federal government information Internet resources. But the two sites display significant functional differences. "One-Stop Shopping"'s "Subjects A to Z" links are selected and indexed by data-producing agencies. The collaborative nature of the project suggests that coverage will be more detailed and more comprehensive than GPO's offering. On the other hand, GPO's reliance upon the 200 or so durable subject headings of the venerable *GPO Subject Bibliographies* as applied by experienced indexers, makes its "Browse by Topic" links much easier to use, as does the skillful use of abstracts describing agency focus and web site content. I was stymied by the arcane jargon substituting for subject headings in "Subjects A to Z": "OASDI"? "Contingent workers"? "Displaced workers"? [Some alarming omissions, viz. "Immigration", probably reflect the current web site's prototypical nature, restricted to Interagency Council on Statistical Policy agencies.] Perhaps "One-Stop Shopping" should revert to the broad topics used by *Stat Abs*'s "Guide to Sources of Statistics".

Keyword searching of the content of federal information WWW pages is provided through the "GPO Pathway Indexer", whose GOVBOT-related Harvest software scans 22,000 links on more than 780 Internet servers in the "gov" and "mil" domains daily. It's safe to assume that a similar web searcher

implemented by "One-Stop Shopping" would be restricted to scanning a small number of federal Internet servers known to hold statistical information. Perhaps the collaborative nature of the "One-Stop Shopping" program will standardize the use of HTML indexing tags on relevant web pages produced by participating agencies to increase search result relevance. But these refinements won't matter: given the present interest in locking statistical information behind passwords or within WAIS servers, I suspect that keyword searching for specific pieces of federal statistical information will be useless. A similar effect has blighted the GILS initiative, albeit from a different cause, that no one server holds a copy of every GILS record makes searching that genre of fielded record impossible.

The "One-Stop Shopping" web pages on statistical programs reflect changes in public perceptions of federal statistical data and their availability that relate to the acceleration in electronic dissemination beginning in the late 1980s with the Census Bureau's CD-ROM data releases. At our reference desk, it's not just a matter of readers asking for a summary tape file or a printed report; we now meet readers asking for the Current Population Survey to reproduce analyses published in *Current Population Reports*. I hope that the web site's interest in program information moves participating agencies to make available online documentation including data dictionaries or queriable databases of program questionnaire data items. [A similar "reference shelf" approach seems to have been adopted in "One-Stop Shopping"'s "Federal Statistics Policy Documents" web pages, which link to Federal Committee on Statistical Methodology working papers and relevant *Federal Register* notices. Why not expand this to include, for instance, the NAS/NSF decennial census reports?]

The importance of collaboration in making "One-Stop Shopping" a success is most apparent in the underdeveloped web pages on regional statistics. At present, links are provided to agency web pages that serve as directories of agency regional offices. But our reference desk readers want regional, local, and small area statistical information, regardless of its bureaucratic birthplace. The Census Bureau acknowledges this universal demand with the American Community Survey and its recently advertised county data CD-ROM compendia. I hope that "One-Stop Shopping" will be able to provide more refined access to regional statistical information, perhaps by linking individual regional office web pages through geographic hierarchies.

I've examined the principal features of "One-Stop Shopping for Federal Statistics". There's much that's praiseworthy: I would certainly use it in my daily work assisting researchers, teachers, students, and the general public to identify and locate statistical information. But it doesn't matter which feature of "One-Stop Shopping" is examined: the key to the web site's success is through collaboration among its contributing agencies.

LO