

**The FedStats Statistical Portal:
Developing Access to City Information through MapStats**

Marshall DeBerry Program Manager FedStats mdb@fedstats.gov	Jon Sperling Manager, Geographic Information Analysis Office of Policy Development and Research, HUD jon_sperling@hud.gov
---	---

Overview

During the past six months, [FedStats](http://www.fedstats.gov), the U.S. Census Bureau Statistical Abstract Office, and the U.S. Department of Housing and Urban Development Office of Policy Development and Research (HUD/PD&R) collaborated on the development of a new city component to the FedStats MapStats website. [MapStats](http://www.fedstats.gov) effectively integrates two separate agency applications to provide users with complementary and seamless access to current and historical data from several federal statistical agencies. This case study of the recently released enhancement of MapStats provides a window into ongoing developments at FedStats and a potential model for future directions.

Background

The FedStats web site (www.fedstats.gov) serves as a portal for over 100 Federal agencies that provide statistical information to the public. An important enabling objective of FedStats is to help users find quickly the information they need, located across multiple separate agencies, without first having to know how government programs and/or data files are organized. Available to the public since May 1997, FedStats continues to meet, and improve on, its goal of providing the public with easy standardized access to the wide array of Federal statistics.

Unlike many other countries where only one entity may collect, analyze, and disseminate statistical information, the United States has developed a highly decentralized statistical system. Different government departments and agencies collect and disseminate statistical information on health, unemployment, demographics, education, crime, and a host of other categories. In the mid 1990s, with the advent of the World Wide Web and new Internet technologies, many saw an opportunity to present a more uniform face of the United States statistical system to the public. In 1995, the Interagency Council on Statistical Policy (ICSP)[1], comprised of the heads of the major statistical agencies or offices of the Federal government, began to develop plans on how to best present government statistical information to the public through the use of the Internet. The result was the FedStats web site, a portal that provides users with a common gateway and quick access to a wide variety of statistical information.

The ICSP's Interagency FedStats Task Force is responsible for designing, developing, enhancing and managing the web site. Comprised of representatives from the ICSP statistical agencies, the task force routinely makes incremental improvements to the site by enhancing existing FedStats features. These

improvements are based on user suggestions, user log analysis, and results of on-going usability testing. The task force meets monthly to review progress, hear invited speakers, share challenges, generate new ideas, and develop future directions. ICSP agencies annually provide resources to cover the costs of the site's hardware, software, and support staff activities. The agencies also provide staff time for FedStats meetings, prototype design, and assessing and testing new features for the site.

In the process of designing and developing FedStats, the task force has encountered various approaches for disseminating Federal statistics at individual agency web sites. Part of the benefit of the FedStats initiative is that agencies learn from each other, leverage common solutions and can collaboratively seek solutions for challenges each cannot necessarily solve on their own. The integration of the HUD/PD&R State of the Cities Data System and MapStats, as discussed in this paper, poses one model for future directions.

Current Features

While FedStats continues to evolve, the site offers a set of core features shown below that the task force updates and/or enhances as appropriate:

1. A-Z Topic Links -- Over 300 topics or subjects ranging from agriculture through weekly earnings.
2. MapStats -- Statistics for geographic areas such as state, county, city and congressional district.
3. Search -- FedStats search capability queries all the linked agencies and returns information related to the request.
4. Press Releases -- Current press release information is linked across agencies.
5. Statistical reference shelf -- Published collections of statistics available online (including the Statistical Abstract of the United States).
6. Agencies -- A list of Federal agencies involved directly or indirectly in statistical activities.
7. Federal statistical policy - Budget documents, working papers, and Federal Register notices.
8. Additional links -- access to other statistical sites and general government locator sites.

MapStats: A New Model?

"Where can I find key statistics and information about my state, county, or local community?" The MapStats feature on FedStats answers this question by providing one-click access to state and two-click access to county and city data. Once at the selected geographic level, users can then select the statistical information of interest, ranging from demographic data pertaining to population, crime, or home ownership status, to economic and geographic information. On selected tables, users have the option to download the information into a format suitable for desktop analysis using spreadsheet or statistical software.

A recent and valuable addition to the MapStats application, from both a data content and data integration standpoint, was the inclusion of cross-agency information for cities and places. This new initiative was quickly and successfully implemented based on close collaboration and partnership between the Department of Housing and Urban Development's Office of Policy Development and Research (HUD/PD&R), the U.S. Census Bureau's Statistical Abstract Office, and FedStats staff. Detailed demographic, economic and geographic information for cities was developed for use within the MapStats framework.

HUD had already developed a robust and highly popular State of the Cities Data System accessible from its PD&R [HUDUSER](#) website. HUD staff saw an opportunity to leverage the separate web-based dissemination efforts by FedStats and HUD to create a one-stop portal for statistical data and information on cities.

Indeed, both applications have benefited by linking their complementary applications and development efforts. MapStats includes incorporated places (e.g., cities) of 25,000 or more population. This threshold was determined based on file availability from the Census Bureau's Statistical Abstract Office. The HUD State of the Cities Data System (SOCDS), on the other hand, includes all places located in metropolitan areas, incorporated and unincorporated, regardless of population size. As such, MapStats offers two handoff points in its application to the SOCDS. If MapStats does not include a requested place, the user is directed via links to the SOCDS for data on their place of interest. At the same time, a user is directed to the SOCDS for more information on their selected city via a link on the selected MapStats City page.

In order to create a seamless linkage or pass-off between the two systems, HUD staff designed new city pages that allowed MapStats users to "deep-link" into the HUD SOCDS from the MapStats city selection rather than having the user re-drill down from opening page of the SOCDS application. This development provides a model for other agencies to link their data at the primary geographic levels (e.g., state, county, city) to the FedStats MapStats portal.

Future enhancements to the MapStats section of FedStats will incorporate additional layers of geography in which information can be made available from multiple sources, the exploration of technologies such as XML to facilitate the querying of heterogeneous statistical databases, and creating enhanced visualization toolsets for the display of integrated statistical data.

Technological Features

The majority of the FedStats web site uses Open Source software solutions to retrieve and disseminate information to site visitors. In general, open source software can be described as software for which the original source code is freely available along with the executable program. Software can only be called "open source" if its license allows users to redistribute the program and source code at no charge. Proprietary software, on the other hand, is distributed under a very different license; it cannot be modified, copied, or redistributed without the owners' permission.

Some of the open source software in use by FedStats includes the Apache Group web server, MySQL relational database, Perl programming language, and the Linux operating system. The Apache web server is used as the site server and has operated with minimal interruptions since its installation. The retrieval of information for selected data requests, such as the MapStats feature, uses the MySQL database for queries. The Perl language is used to parse requests for information to the MySQL database, and format the information into the appropriate HTML format. FedStats continues to explore new uses of Open Source software to ensure that the site will continue to operate in a reliable and efficient manner.

Future Directions

FedStats envisions the development of a National Statistical Knowledge Network (NSKN) as a type of statistical digital library with tools for information finding, extraction and reuse, visualization, and as a means to transform knowledge into intelligence while maintaining the privacy and confidentiality of respondents. To achieve this vision, NSKN will require common user interfaces, metadata standards, and data access and searching tools usable by persons with different levels of computer and statistical literacy that enable appropriate uses of the data with analysis within and between databases.

The current decentralized, autonomous sources of statistical information have few commonalities in terms of concepts and definitions; system architectures, software, and hardware; measurement methods; interfaces; or dissemination and presentation modalities. Interoperability is a major hurdle in a variety of areas. Data integration issues abound. [2] Significant challenges in high-end computing and computation and large-scale networking exist for the NSKN vision to become a reality.

Computer and information scientists will solve some of these challenges, while others will require a more multidisciplinary, multi-sector approach. For example, efforts by mathematical statisticians with expertise in creating estimates from complex sample surveys, building small area estimation models, and estimating measures of error, including those due to sampling and non-sampling errors, will require multidisciplinary approaches to solving problem sets that cut across multiple areas.

If the metadata needed to interpret and use statistical information are to be made available and integrated with the data, the processes and procedures for collecting and compiling statistical information must also be the focus of information technologies research and development efforts.

The FedStats Task Force remains confident that these issues can be addressed within the near future, and that the challenges presented in developing and providing the public with a cohesive National Statistical Knowledge Network will be fully realized. We hope that efforts such as the MapStats enhancement presented in this paper will keep this effort moving forward.

References

[1] For more information see: *Statistical Programs of the United States Government, Fiscal Year 2004*: (<http://www.whitehouse.gov/omb/inforeg/04statprog.pdf>)

[2] Wallace, M. and Sperling, J. "User-Driven Integrated Statistical Solutions", *URISA Journal*, Vol. 12, No. 4, Fall 2000: (<http://www.urisa.org/Journal/protect/Vol12No4/JrnlContents12-4.htm>)