

SEMINAIRE

E+; 3=! C

SEMINAR

STATISTICAL COMMISSION AND
ECONOMIC COMMISSION FOR EUROPEDistr.
GENERALCONFERENCE OF EUROPEAN
STATISTICIANSCES/SEM.43/32 (Summary)
25 January 2000

Original: ENGLISH

Seminar on integrated statistical information
systems and related matters (ISIS 2000)

(Riga, Latvia, 29-31 May 2000)

Topic IV: Improving data dissemination strategies

**FEDSTATS: CREATING THE U.S. NATIONAL STATISTICAL INFORMATION
INFRASTRUCTURE OF THE 21ST CENTURY**

Contributed paper

Submitted by the U.S. Bureau of Justice Statistics
and the U.S. Bureau of the Census¹

SUMMARY

1. The United States Interagency Council on Statistical Policy's (ICSP) One-Stop Shopping for Federal Statistical Data web-site known as FedStats (<http://www.fedstats.gov/>) opened in May 1997. FedStats has more than met its initial goal of providing the public with easy Internet access, via an initial point of entry, to the wide array of available Federal statistics. This early success is but a first step in achieving a broader vision for a National Statistical Information Infrastructure (NSII) within the United States.

2. Unlike most other countries where there is one entity that collects, analyses, and disseminates statistical information, the United States has developed a decentralized statistical system. For example, different agencies collect and disseminate statistical information on health, unemployment, demographics, and crime. With the rise of the Internet, however, an opportunity was presented in which a more uniform face of the United States statistical system could be presented to the public. FedStats was created to provide users with a common gateway that could quickly access the statistical information they were seeking.

3. To extend the usefulness of FedStats, new capabilities regarding integrated data dissemination have recently been implemented. With the advent of a new site feature called "MapStats," users can now more easily find statistical information related to the geographic units of states or counties found in the United States. To facilitate

¹ Prepared by Marshall DeBerry, Bureau of Justice Statistics; Valerie J. Gregg and Rachael LaPorte Taylor, Bureau of the Census.

the creation of this tool, Open Source software, such as MySQL and Perl, was used. The use of such robust and easily available tools greatly facilitated the rapid development of the requisite databases and scripts needed to provide users with the appropriate statistical information for a particular geographic area.

4. FedStats continues to be well received and used by both the media and the public. In the first year, FedStats logged over 800,000 user sessions, and over 1,300,000 in the second year. In July 1999, Yahoo Internet Life named FedStats as one of the fifty most incredibly useful sites for the second year in a row. The fourteen ICSP agencies annually provide resources for site design, development, maintenance, and management.

5. During 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, find common solutions and can collaboratively seek solutions for challenges that each agency cannot necessarily solve on their own.

Geographic Information

6. One of the most frequently asked questions of users is "where can I find information about my state, county, community?" The Regional Statistics Feature currently resident on FedStats is a gateway approach to finding information by geographic area on an agency-by-agency basis. However, users want to know about all the information available for a specified location, not just from one agency. That requires substantial exploration through individual agency web sites. To accommodate better those users seeking geographically based statistics, a new feature has been added, called "MapStats." Users can select a "state" or "county" from either drop-down lists or GIF images to facilitate "drilling down" to the particular geographic component. Once at the appropriate geographic level, users can then select the statistical information of interest, ranging from demographic data pertaining to population, crime, and immigration, to economic and environmental information. On selected tables, users also have the option to download the information into a format that is suitable for later analysis in either statistical software or a spreadsheet package. Future enhancements will (1) encompass additional layers of geography in which information can be made available, as well as the exploration of new technologies such as XML to facilitate the querying of heterogeneous statistical databases, and (2) integrate data sets for thematic displays.

Technological Features

7. The FedStats web site uses Open Source solutions in retrieving and disseminating information to site visitors. Some of the open source software in use by FedStats includes the Apache Group web server, MySQL relational database, Perl programming language, and Sendmail, an email program. 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 agricultural and demographic information, uses the MySQL database for queries and storage of information. The Perl language is used to parse requests for information to the MySQL database, and format the information into the appropriate HTML format. Future plans include establishing a database server using Linux, an open source operating system. The FedStats Task Force continues to explore new uses of Open Source software to ensure that the site will continue to operate in a reliable and efficient manner.

The National Statistical Information Infrastructure (NSII) for the 21st Century

8. The vision for the NSII is a national distributed statistical digital library with tools for information finding, for information extraction and reuse, information visualization, and for transforming knowledge into intelligence while maintaining the privacy and confidentiality of respondents. To achieve this vision, NSII will require

common user interfaces, data access and searching tools usable by persons with different levels of computer and statistical literacy that enables appropriate uses of the data with analysis within and between databases.

9. 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. Significant challenges in high-end computing and computation and large-scale networking exist for the NSII vision to become a reality.

10. The task force is 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 Information Infrastructure will be fully met.

FedStats' Data Integration Research Collaborations with the National Science Foundation's Digital Government Program

11. Government is a major user of information technologies, a collector and maintainer of very large data sets, and a provider of critical and unique information services to individuals, states, businesses, and other customers. The goal of the U.S. National Science Foundation's Digital Government Research Program is to fund research at the intersection of the computer and information sciences research communities and the needs of government information service communities. The Internet, which was created from a successful partnership between government agencies and the information technologies research community, is a major motivating factor and context for this program.

12. Within this context, the objective of the Digital Government Program is to support innovative projects that effectively and broadly address through research the potential improvement of agency, interagency, and intergovernmental operations and/or government/citizen interaction. Such research is expected to enable the generation and use of a continuous stream of advanced information technologies for early adoption and integration into the government information systems community.

13. One outgrowth of the May 1997 Digital Government workshop (see: <http://www.isi.edu/nsf/final.html>) was that the ICSP authorized a FedStats interagency Research and Development (R & D) working group. As a first step, the working group identified common challenges facing many statistical agencies that could potentially be overcome by applying cutting-edge information technologies. The working group sought academic and industrial researchers drawing from known social and mathematical scientists who are quite familiar with challenges faced by the statistical community. The working group encouraged these researchers to explore potential alliances with computer and information scientists within their institutions so that in partnership with Federal statistical agencies, they could develop and submit research proposals to several NSF research programs. In other cases, computer scientists approached the FedStats working group about possible information technology research collaborations. Agency staffs spend considerable time familiarizing the researchers with current architectures, systems, and methods for collecting, processing and disseminating statistical information. They also provided explanations of the challenges that serve as barriers to achieving the NSII vision. Researchers spend considerable time analysing the challenges and identified possible topics that could potentially lead to interesting scientific research. Together, the collaborators develop research proposals including objectives, time-lines, and any services-in-kind provided by government and/or industrial partners.