UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE (ECE)
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (iii): Open source and software consortia in statistics

FREE AND OPEN SOURCE SOFTWARE AT THE US CENSUS BUREAU: BUILDING SHARED STATISTICAL DISSEMINATION SYSTEMS

Contributed Paper

Submitted by the U.S. Census Bureau, United States[1]

I. INTRODUCTION

1.   The Census Bureau has long used open source tools in its IT infrastructure and for application development. A recent survey of Census Bureau divisions showed significant usage of free and open source software (FOSS) in such areas as communication, security, data collection, office and data management, and systems. Census Bureau IT and programming staffs identified 125 FOSS packages in use and essential to operations. The most widely used FOSS includes Apache, sendmail, Perl, Linux, samba, and many GNU utilities. The survey identified another nine packages not currently in use that staffs plan to evaluate and use.

2.   Respondents expressed functionality, security, usability, and stability as being the most important reasons for using FOSS. However, the survey also found that there is a perception that migrating to FOSS in some instances could negatively impact ongoing agency legacy operations. Availability of corporate support and ease of procurement ranked lowest in reasons for using FOSS.

3.   This paper discusses how The Census Bureau and FedStats use FOSS to create a portable, interoperable application and application development environment. FedStats is the Web portal to US Federal statistical agencies, developed by the Interagency Council for Statistical Programs. This project is a case study in low-cost, rapid, and collaborative application development resulting in award-winning E-government solutions for data dissemination.

4.   The QuickFacts/MapStats cooperative software project demonstrates some of the advantages of development with free and open source software. The project is a joint activity of the Census Bureau and FedStats.

---

[1] Prepared by Lisa Wolfisch and Rachael LaPorte Taylor (lisa.wolfisch; rachael.laporte.taylor@census.gov).

II. THE NEED FOR BETTER ACCESS TO STATISTICS

5.   The Census Bureau's web site is large, rich, and diverse. This content grew independently in a number of different program areas. A user interested in local area data had to visit many different program sites, organized in different ways, to acquire the variety of data the Census Bureau published for that area. In 1999, the State Data Center (SDC) Steering Committee recommended that the Census Bureau provide a consolidated presentation of data from the full range of Census Bureau programs for any state or county. The SDCs provide easy and efficient access to U.S. Census Bureau data and information through a wide network of coordinating and affiliate agencies in each state. They work extensively with Census data users of all kinds. The QuickFacts Strike Force was assembled in December 1999 to address this key recommendation.

6.   In a parallel effort, FedStats recognized a similar need for centralized profiles that included statistics from various federal agencies. Could it leverage the work on QuickFacts to improve its MapStats pages, as well? At the time, the MapStats section was a geographic interface to links to remote datasets and applications residing at the statistical agencies' Web and FTP sites. The data was not integrated into area profiles at the FedStats site itself. FedStats achieved 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, but by enhancing its offerings with statistical profiles, FedStats had an opportunity to provide a more direct data access tool for its users. FedStats could not only be a portal to the vast, decentralized federal statistical system, but could begin to integrate data sets to provide a seamless user experience. The QuickFacts Strike Force agreed that this was an excellent opportunity to work together with FedStats.

7.   From the beginning, QuickFacts recognized the importance of satisfying user requirements, particularly those of novice users, the primary target audience.   In spring 2000, working without a dedicated staff or budget, the Strike Force returned to the original SDC group with mock-ups and lists of features to rank. In the next few months, members of the Strike Force made similar presentations to committees or meetings of the National Association of State Development Agencies, the American Chamber of Commerce Research Association, the National Association for County Community Economic Development, and the Association for University Business and Economic Research, to gather feedback on the project's progress.

A QuickFacts State Profile

8.   Using rapid prototyping techniques and open source software, the first production version of the application took six months to develop. This was a stable, production-quality version, and not a prototype. It incorporated the functional requirements of both the Census Bureau and FedStats. The joint project initially launched as Census QuickFacts in September of 2000, then shortly thereafter as MapStats on the FedStats site. Due to our past experiences with FOSS, a requirement to work with a heterogeneous computing infrastructure, and lack of funding, we chose to develop the project with FOSS tools.   An examination of some advantages of FOSS and how they contributed to the success of the QuickFacts/MapStats project follows.

III. USING FOSS FOR COOPERATIVE APPLICATION DEVELOPMENT

9.   Cost is often noted as an advantage for using FOSS. We've learned that successful, usable applications do not have to cost large sums. Since QuickFacts had no budget (and has none to this day), with FOSS and existing hardware, we were able to create several development and production environments with no procurement costs or delays. Procurement delays have often put projects behind schedule but were not a concern in this case.   While total cost of ownership of FOSS is the latest topic of debate among consultants and technicians, the advantage of low start up costs is clear.   However, even if funds had been available, FOSS would have been used because of its other advantages.
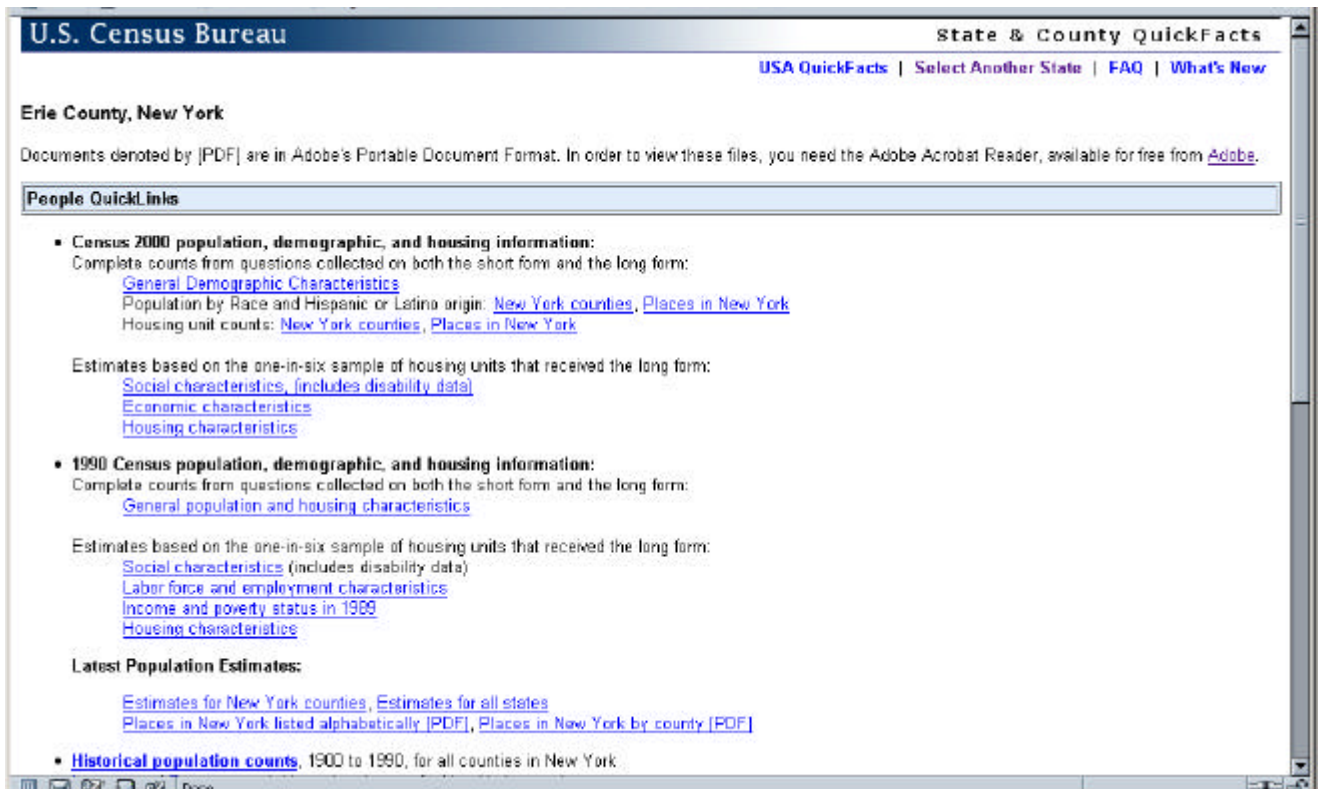
10.   Using robust and easily available FOSS tools to develop the QuickFacts/MapStats application provided much needed portability. The code base is written in the Perl programming language and data is stored in the MySQL database engine. FedStats maintains the shared database. A combination of configuration files and presentation templates creates a highly malleable system architecture.   Templates allow designers to work independently of programmers, thus allowing the core application code to remain stable while the design changes. Perl's Template Toolkit extension provides this templating functionality.

11.   Portability was a critical requirement for the project's success since the Census Bureau and FedStats Web sites have always used different operating systems and hardware platforms. QuickFacts was originally developed under Irix 6.5 on an SGI Octane Workstation, and was in production on an SGI Challenge XL. After three weeks in production, the application migrated seamlessly to a Dell PC running Red Hat Linux. The MapStats version is in production on a Sun Workstation running Solaris 5.8. There is also a portable version of MapStats on both Windows and Linux laptops. This leads to another advantage that enables collaboration. Using FOSS, the project is not dependent on any hardware or operating system, and therefore the Census Bureau and FedStats are able to make independent choices in their platform specifications. The two need not agree on these specifications to ensure the project's success, nor agree on vendor-supplied components.

IV. APPLICATION FUNCTIONALITY

12.   In late 2002, thematic maps were added to MapStats and QuickFacts to provide another view of the data. The maps include color and grayscale maps for the US and states by county for several demographic themes. The maps are in PNG format and were created with Perl and the GD library for graphics manipulation.

13.   The QuickFacts/MapStats profiles enrich the data by presenting the data in easily digestible form, with percentages or other derived measures, and by making side by side comparisons: state profiles include national figures; county profiles include the corresponding state data. Over 50 data items are presented. Each data item has a link to definition and methodology information, and every area profile invites the user to "Browse more data sets" for the same area. MapStats uses this mechanism to link to various agencies' data sets, while QuickFacts links to the many detailed datasets ("QuickLinks") produced by the Census Bureau.

QuickLinks to Census Datasets

14.   The application interface provides multiple means of navigation. States and counties can be selected from maps or drop-down menus, and a county may even be selected by specifying the name of a city or town in it. In January, 2004, city data was added to MapStats. MapStats city profiles link directly to the Department of Housing and Urban Development's State of the Cities database, an effort that fostered more cooperation among agencies and interoperability among applications. The shared code base between QuickFacts and FedStats enables the migration of enhancements quickly and easily.   City pages, showing Census Bureau data items only, will be migrated to QuickFacts later this year. While the functionality is inherently the same in the code base, the applications can be configured to include or exclude any specific feature.

People MapStats	Los Angeles	California
Population, 2000	3,694,820	33,871,648
Population, net change, 1990 to 2000	209,321	4,080,221
Population, percent change, 1990 to 2000	6.0%	13.6%
Population under 5 years old, 2000	285,976	2,486,981
Persons under 5 years old, percent, 2000	7.7%	7.3%
Persons under 18 years old, 2000	981,311	9,249,829
Persons under 18 years old, percent, 2000	26.6%	27.3%
Persons 65 years old and over, 2000	357,129	3,595,658
Persons 65 years old and over, percent, 2000	9.7%	10.6%
Female persons, percent, 2000	50.2%	50.2%
White persons, 2000 (a)	1,734,036	20,170,059
Black or African American persons, 2000 (a)	415,195	2,263,882
American Indian and Alaska Native persons, 2000 (a)	29,412	333,346

A MapStats City Profile

15. Additionally, it is important to note that QuickFacts/MapStats pages, including the thematic maps, are accessible to persons with disabilities. In fact, blind users have provided positive feedback on their experiences with the sites.

16. As development continues on QuickFacts/MapStats, future considerations include looking at how the use of XML can enhance the data storage and delivery architecture, and looking at open source development models for better source code and work process control as development grows. Traditional open source development, in which many application programmers may contribute to the code base, can promote wider collaboration efforts among statistical agencies.

V. CONCLUSION

17. The return on investment of MapStats and QuickFacts includes the intangible assets beyond financial benefits, among them are collaboration, a culture of innovation, and reputation that comes from improved citizen access to data. The QuickFacts/MapStats project received the 2002 Federal CIO Council Award, and the 2001 Census Bureau Innovation Award. QuickFacts/MapStats, a product of a collaboration process across government, persuasively demonstrates the benefits to be anticipated from e-government initiatives that focus on the fundamentals: improving the quality of information service for citizens and businesses, maximizing the information capabilities and productivity of all people in a collaborative information management process, and eliminating redundant costs for system components.

18. Today, QuickFacts serves over 170,000 pages per day, accounting for over 10% of traffic to Census.gov Web sites. Yet, QuickFacts still has no dedicated funds. All resources for the project remain voluntary, and there is no line-item budget. MapStats is a very popular part of the FedStats Web site, and includes rich profiles beyond those developed in the joint project. For example, some data is available by Federal Judiciary district.
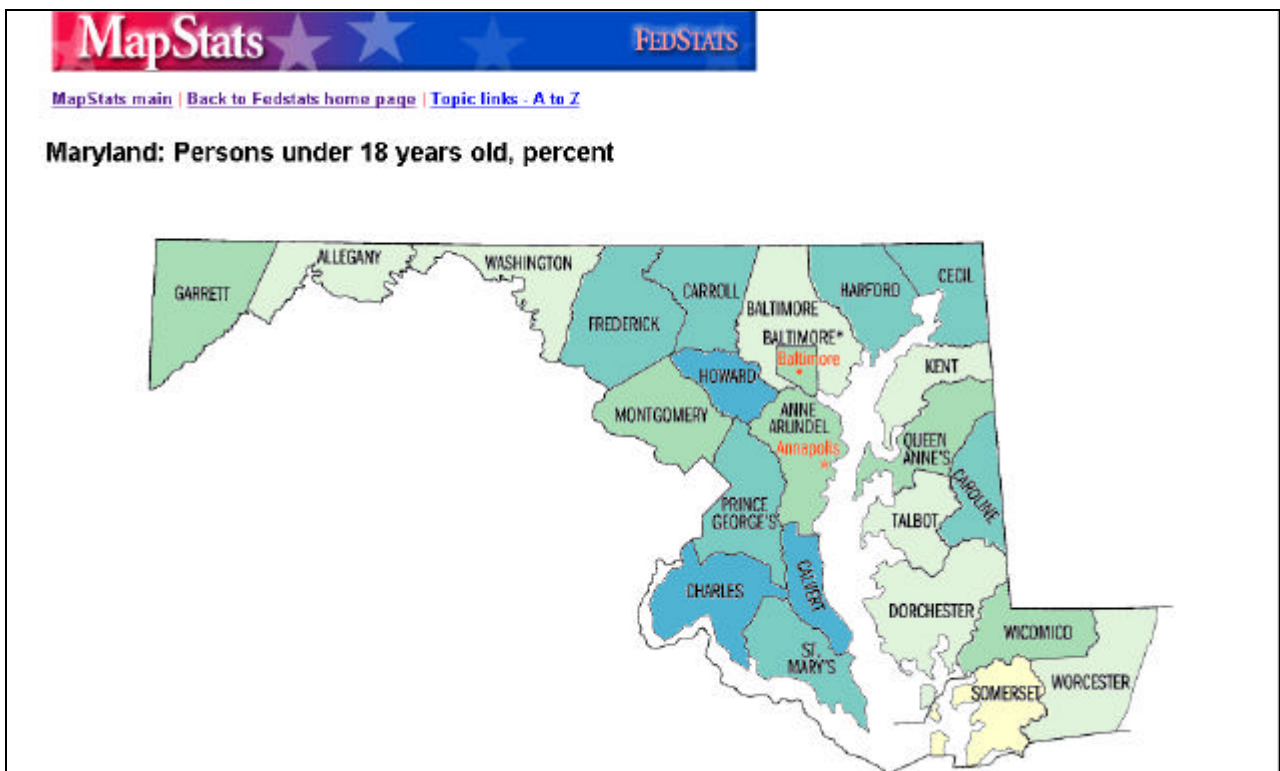
19.   Because they are an easy and inviting starting point, both MapStats and QuickFacts are bringing thousands of new users to their respective websites, introducing many to the usefulness of Federal statistics for the first time. The following is typical of user feedback for the applications:

"On many occasions, students/teachers do not have the time to engage in this type of research. To be able to access, or "call up" information about states in a matter of minutes is a plus for the teacher."

"Thanks for such a thorough and easy to access set of data, county by county. I am writing a couple of grants, and found exactly what I needed in one easy stop."

"This was one quick, and consumer-easy source for the most relevant information. This page allowed me immediate access for all information I needed. I was able to do this from one page, not having to jump from web site to web site. I wish all of our tax payers' money could be spent in such an easy to understand, user-friendly, and common sense manner."

"It is nice to have this available to all citizens who have access to the net. Makes me feel some of my taxes are going to the right place."

A Thematic Map