

PLANNING FOR THE FUTURE WITH THE SAS® SYSTEM

Bob Lohr & Bob Nunziata, U.S. Census Bureau

For more than five years, the Census Bureau's Demographic Programs Associate has been striving to build a "state-of-the-art" statistical computing environment. When we began, we had simply hoped to provide our analysts with access to demographic data, which would be available on a central computer and accessible through terminals or PCs. Once we started to build this type of facility, we began to realize that much more was possible. Rather than accessing data stored in one location and using common processors, or even using distributed processors to access centrally located data, we discovered that we could distribute the data itself as well as the access. But first, we had to modernize our computer processing philosophy.

Throughout the Census Bureau's computer history - stretching clear back to UNIVAC I - our data was processed on a few very large computers because of the size of our files. Over the last several years, however, our processing environment has begun to move from the computer room to the desktop. This does not mean that our larger computers have gone away. We are still dependent on them for production processing, and some will be used as central data sources.

Is it really possible - especially for a government agency - to build and maintain a state-of-the-art computer facility? What prevents it?

- 1) The procurement cycle guarantees that one's environment will never be the most modern available. By the time a new product is actually in-house, an improved one has been introduced.
- 2) The cost of always buying the newest and best is prohibitive. The second generation of a product is usually cheaper, and often more efficient.
- 3) Always buying the newest and best means constantly installing and training for such. The environment is never stable.
- 4) There is no guarantee that the newest and best is bug-free (it often isn't.)
- 5) There is no guarantee that the newest and best will be compatible with what's already in-house.

Historically, large computer companies have stressed proprietary hardware and software solutions. This meant two

things: a company purchasing such a system was quite likely to standardize on one vendor for all of its computer needs; and when it was time to replace all or part of the system, the company usually had such a large investment in code and training that it was not feasible to consider another vendor. When the system became outmoded, the cost of building a new system severely limited the options. A state-of-the-art computer environment, then, involved buying the newest and best proprietary hardware and software - which may not be state-of-the-art at all.

This leads to the idea of "Pareto-state-of-the-art" - a computer environment which comes as close to state-of-the-art as possible, yet places an emphasis on planning for the future and for portability between hardware platforms. It uses the second generation of the newest and the best, which usually results in cheaper, more efficient hardware and software. It gives support staff time to evaluate the prospects of a new technology, to anticipate and deal with potential bugs, and to resolve conflicts with previously installed products. There is also less emphasis placed on a fixation with "brand new," resulting in better planning for the future. Because of the current pace of technological development, hardware vendors are leapfrogging each other introducing faster, cheaper, smaller machines. The lifecycle of any machine as state-of-the-art is becoming shorter and the best machine available at any given time could come from any one of a number of different companies. And as the proprietary trend among computer vendors is giving way to a move toward open systems, the decision of which piece of hardware to buy is becoming trivial. We are also seeing a developing trend among software companies toward cooperation and standardization. Instead of relying on hardware to achieve compatibility, a Pareto-state-of-the-art environment demands standardization on networking and software.

First, a networking standard is essential. A physical standard is relatively immaterial, as most of the popular media (thick and thin Ethernet, FDDI, and Token Ring) are compatible. The protocol standard is much more important, because this is what determines the degree of contact between the different machines on the network. We have found that the de facto standard networking protocol is TCP/IP. It is available for all

popular platforms and is highly efficient. If not available from the hardware manufacturer, it can be had from a variety of third-party software companies. And because it is a DOD standard, any supported TCP/IP service such as telnet, ftp, or smtp is compatible. Support for TCP/IP is also built into a number of application software packages as well.

Standardizing on application software involves consideration of many factors. First, it must be compatible with the networking standard. This is important in order for data to be shared across machines. Either the software must have built-in support for a standard protocol, or it must be transparently supported by the operating system. Second, the software must have a common look-and-feel across all operating systems. This translates into less time and money spent on training, and a greater sense of familiarity with the product, even though it may be on a foreign machine. On many multi-user computers, a product such as X Windows provides a common graphical user interface. Third, all program code and data files must be portable across platforms. Programs copied from one machine to another should require little or no modification. In a case where data files must be moved between machines, the files should be binary-compatible, or the software should provide a facility for easily transporting them between machines. Fourth, it should be easy to import and export data and graphics into other formats. Users should be able to accomplish this by pointing and clicking or choosing a menu option to create an export file, or at the very worst by writing a simple program.

This was the philosophy that Demographic Programs elected to follow. We chose TCP/IP as our networking standard because of its hardware independence. In addition, it has allowed us to directly share file systems through the use of NFS® (Network File Systems), and to remotely access programs on various machines through X Windows. This has given us an environment in which we are free to choose machines which best fit our needs at any time. Computers currently in use consist of PCs, Digital Equipment VAXes®, Novell® LAN servers, and UNIX® workstations from Sun Microsystems®, Hewlett Packard, and IBM®. Our supported software products are available for all of these platforms and include WordPerfect® and 20/20® - and of course, the SAS system.

How well does the SAS system fit these criteria?

- 1) Portability across platforms - SAS Institute's answer to this is SAS/Connect®, which first appeared as the

Micro-to-Host link in PC version 6.03. It allowed code to be developed on a microcomputer and remotely submitted to a larger machine via a serial port. It also allowed for serial uploading and downloading of data. The product has evolved into one allowing peer-to-peer as well as client-server communications and now supports Ethernet as well.

- 2) Network compatibility - SAS/Connect allows for peer-to-peer as well as client-server communications. Combined with TCP/IP, NFS, and X Windows, it allows for software to reside on one machine, data to reside on another, and the user interface to be displayed on yet another - all transparently.
- 3) Common look-and-feel - SAS Institute has worked hard and long to make its software compatible across all platforms, resulting in the Multi Vendor Architecture (MVA®.) As most of the system is written in C language, only about 10% of the code is host-dependent. As a result, porting to new platforms is much easier. It also means that a similar look-and-feel can be maintained across all systems. In addition, all versions of SAS software since 6.06 support the X Window standard, and Windows® and Presentation Manager® versions are coming.
- 4) Import and export of data and graphics - SAS/PC® includes procedures for reading and writing dBase® and DIF file formats. Versions 6.06 and later include a variety of SAS/ACCESS® products for providing transparent access to database systems such as Oracle®, Sybase®, and RDB®. SAS/GRAPH allows the user to save graphical images in a variety of common formats.

Careful consideration of these criteria will prevent a user or an application from becoming locked into one particular machine or hardware architecture. If the networking and software standards are properly set and enforced, then the choice of a hardware platform can take a back seat to more important issues such as application functionality. The goal is to provide ease of portability when an application must be moved. For example, anyone requiring a powerful CPU may choose a UNIX workstation. Those more familiar and comfortable with PC software can use the Intel architecture or its equivalent. Larger, centralized machines in the computer room also fit easily into

the scheme. And when new machines are purchased, the standards are such that they will fit into the hierarchy. This is extremely important when a project must be moved for reasons such as unexpected growth or a catastrophic failure. A company is no longer locked into a proprietary hardware solution. Fewer companies are remaining loyal to traditionally proprietary vendors such as IBM or DEC[®], and are instead becoming computer melting pots. Thanks to portable software products such as the SAS system, open systems are becoming less a wish for the future and more a reality for the present.

SAS, SAS/Connect, SAS/Access, SAS/Graph, SAS/PC and MVA are registered trademarks of SAS Institute, Cary, NC.

DEC, VAX, and RDB are registered trademarks of Digital Equipment Corporation, Maynard, MA.

Novell is a registered trademark of Novell, Inc., Provo, UT.

UNIX is a registered trademark of AT&T.

Sun and NFS are registered trademarks of Sun Microsystems, Sunnyvale, CA.

IBM and Presentation Manager are registered trademarks of International Business Machines Corporation.

WordPerfect is a registered trademark of WordPerfect Corporation, Orem, UT.

20/20 is a registered trademark of Access Technology, Inc.

dBase is a registered trademark of Borland International, Inc., Scotts Valley, CA.

Oracle is a registered trademark of Oracle Corporation.

Sybase is a registered trademark of Sybase, Inc.

Windows is a registered trademark of Microsoft Corporation.