

GAO

Testimony

Before the Subcommittee on National Economic Growth,
Natural Resources, and Regulatory Affairs
Committee on Government Reform and Oversight
House of Representatives

For Release
on Delivery
Expected at
10 a.m.
Tuesday
September 10, 1996

INFORMATION MANAGEMENT

Technical Review of the White House Data Base

Statement of Jack L. Brock, Jr.
Director, Information Resources Management/General
Government Issues
Accounting and Information Management Division



Mr. Chairman and Members of the Subcommittee:

Thank you for inviting us to participate in the Subcommittee's hearings on the White House Data Base. You requested that we do a technical review of the data base to determine who uses it, to determine the source of information populating the data base, and to evaluate the performance and operational components of the data base. Today I will provide an interim update on data base users and operational components of the data base. Additionally, I will briefly describe how the data base is configured and how it works. We do not yet have enough information to provide meaningful comments on data source and will provide that information at a later date.

In performing our review, we reviewed technical manuals and other documentation for the products supporting the data base. We reviewed documentation showing how users are assigned access rights to data base data. Further, we reviewed the data base's process models, computer programs, and directory structures;¹ toured the computer center where the main data base computer is located; and ran a small number of data base queries to initially sample the data base population. We also conducted interviews with the most active data base users as well as with the data administrator and the primary data base administrator.

White House Data Base Configuration

The White House Data Base was developed in 1994 to facilitate contacts with individuals and organizations who are important to the Presidency. It replaced a number of existing data bases with a single system which was intended to be easy to use and provide a greater level of service to a variety of users. The system has been operational since August 1995.

Among other things, the data base is used for developing invitation lists for White House events and for providing information to help prepare thank you notes, holiday cards, and other correspondence. As such, the information contained on the data base ranges from names, addresses, phone numbers, social security numbers, contributor information, and dates of birth to individual relationships to the First Family and political affiliations. According to the White House, the data base contains personal information on about 200,000 individuals.

In developing the data base, the White House used a widely accepted approach—Joint Application Development. Under this approach, users

¹A table of identifiers and references to the corresponding data items.

meet with programmers in a more intensive design session than usual—with the goals of eliminating rewrites of user interfaces and paving the way for faster application development. Development of the data base began with a series of technical interviews with potential users to determine, among other things, the sources of the data for the data base and the extent to which the data would be shared with nonfederal entities or individuals.

Once these interviews were concluded, design and development elements were pursued on several fronts. First, potential users were asked to review functional aspects of the system and provide feedback. Second, the system architecture² was developed and implemented based on detailed requirements and joint design elements provided by the customers and others.

The data base operates on and is accessible through the White House's local area network, or LAN.³ While more than 1,600 users are authorized to access the LAN, less than 150 users have been given access to the data base and even fewer actually use the data base. The products supporting the White House LAN, operating system, and data base system are widely used in the government and commercial sectors. The LAN uses version 3.12 of Novell's network operating system. The data base runs on Microsoft's Windows NT operating system using Sybase's System 10 data base management system.

Sybase's System 10 is a relational data base management system, which is a system that allows both end-users and application programmers to store data in, and retrieve data from, data bases that are perceived as a collection of relations or tables. The data base is comprised of 125 tables. Data is input to and retrieved from these tables using simple screens and drop-down menus.

Sybase's System 10 is built with published and readily available interface specifications. It is open to the extent that anyone can write a program that will connect to the server.⁴ This is unlike traditional proprietary data base management systems, which could be accessed only with

²A system architecture is the logical and physical layout of a system that includes hardware, software, communications, data management, and security.

³A local area network is a group of computers and other devices dispersed over a relatively limited area and connected by a communications link that enables the computers to interact with each other.

⁴A stand-alone computer in a local area network that holds and manages the data base.

vendor-supplied tools or programs written with vendor-specific languages and compilers.

GAO Observations on Users, Data Source, and Operating Processes and Performance

In developing the data base, the White House acquired well-established, commercially available products and created a system that users we interviewed were generally satisfied with. However, as I will discuss in more detail, the design of the data base limits system performance. Further, the system—while having in place some internal controls—needs additional controls to assure the integrity and accuracy of data.

Data Base Use and User Satisfaction

As noted earlier, data base users primarily use the data base as a tool for maintaining contact with individuals and organizations important to the Presidency. Users told us that they were generally satisfied with the system.

Less than 100 White House staff actually use the system, and only about 25 make moderate to heavy use (relative to other users) of the system—with the heaviest users representing the White House Social Office, Personal Correspondence Office, and Outreach Office, as well as system administrators. We examined user accounts and interviewed those staff making heavy use of the system in terms of amount of data both input to and read from the system. These included two staff in the Social Office, one in the Outreach Office, two on the Personal Correspondence staff, the data base data administrator, and a Sybase system administrator. We also interviewed four other business users and a system administrator who represent less heavy users of the system.

Social Office personnel use the system to assist in developing invitation lists and planning state dinners and other events. Personal Correspondence personnel use the data base to help compose letters for the President. In doing so, they retrieve information from the data base on addresses, names of family members, White House events attended, and how the correspondent knows the President. The Outreach user we interviewed entered data into the data base for use in generating lists of holiday card recipients. Many users supplement the data base with information from manually accessed address lists. All those users we interviewed who had used the prior systems believed that the new system was better, and—for some users—the system is critical to their ability to complete their tasks.

System administrators—who account for about 10 percent of all people who have accessed the data base—manage the system and maintain data base information. For example, they perform system backups, troubleshoot, and perform routine maintenance in the normal course of managing the system.

Operating Characteristics

The individual components supporting the data base—the network, server, and data base engine—are individually well-regarded and could be considered to be leading edge components for business applications similar to those run by the White House. However, the strength of the individual system components has been diminished by the design of the data base itself. Specifically, in developing the system, the White House attempted to meet all user requirements for a large array of potential information needs. Rather than take advantage of the relational data base capabilities of Sybase, the designers established a one-to-one relationship between the logical and physical attributes of the data base resulting in 125 tables. The data base operates more as an index sequential data base where relationships between and among data elements have to be established across many tables. This contributes to increased system overhead (requires the system to process additional steps) and thus taxes the performance capabilities of the system.

Because the data base has relatively few users and is an improvement over what users had been using, individual users have probably not been affected by the data base design. However, if demand increased, system performance could unnecessarily degrade.

In order to minimize performance impact, system administrators have made compromises which affect the data base's internal controls. First, system administrators told us that turning on the internal audit trail, which I will discuss later, would seriously slow down system performance; and that to turn on the audit trail would take several staff weeks of programming effort to minimize the impact on overall system performance. Second, system administrators have chosen not to use the referential integrity capability⁵ that Sybase offers because of performance issues. Referential integrity is critical to any data base to assure that necessary checks are in place to limit inappropriate data input and assure that output is accurate. For the White House Data Base, referential integrity is implemented through the application itself. Because of the

⁵Referential integrity is the constraint or rule that must remain true for a data base to preserve integrity. For example, that at most only two biological parents can exist for a child.

complexity of the application structure, it is difficult to assure that all edit checks are in place and work properly across the application. We found that some checks are not operational which in turn leads to a higher probability of inaccurate information being input or retrieved from the system.

Operating Processes and Procedures

Good business systems operate in a controlled environment to ensure that data within these systems is accurate, that data output is reliable, and that data integrity is assured so that only authorized users have access to the data and that such access is appropriate to their needs. To provide such assurance, an organization needs well-articulated policies and procedures, good training, and an ability to ensure compliance with established processes and procedures.

For the government, these concepts are embodied in the Office of Management and Budget's Circular A-130 which lays out the need for policies, rules of behavior governing system use, training, and the need to incorporate good controls. Circular A-130 states that accountability is normally accomplished by identifying and authenticating users and subsequently tracing actions on the system to the user who initiated them. As a system containing sensitive information on up to 200,000 individuals, and, as a system that is important to meet the work needs of several White House offices, data base users and managers need to apply the principles of A-130 to system operations.

We found that the White House has taken several positive steps to create a controlled environment. For example:

- Personalized training is available to all users.
- Users are required to sign a document stating that they will take measures to protect information including establishing and protecting passwords, logging out when leaving their computers, and reporting unauthorized access to the system.
- Password access is required to enter the system and a warning screen appears to inform the user that information within the data base is for official use only.
- The data base has an effective defense against outside intruders or "hackers" breaking into the system.
- Controls have been established within the system to limit access to certain portions of the data base to only those with a need to know. Additionally, only a limited number of users have authority to print reports.

Even with these processes in place, we found that the data base requires additional measures before data integrity and operational effectiveness can be assured. For example:

- Users do not have well-documented processes and procedures for how and when to use the data base. Written documentation, reinforced with training and operational processes, would provide a better basis for assuring system managers that the data base was being used effectively and that all users were appropriately keeping the data base current. While users were trained individually by system administrators or other users, only one user out of the nine business users that we interviewed reported having a users manual. None of these users reported having training concerning the security of the system. Such guidance can help ensure that users are familiar with the system and are entering information correctly. In talking with users we found that most everyone could navigate the system adequately; however, we also found that some duplicate information on individuals was being entered into the system and that some information was being entered into the wrong field. This causes some data base tables to contain more information than necessary and slows down the processing of information.
- Although the data base has established security policies, procedures necessary to make them effective have not been well-documented. For example, the system does not require frequent changes in passwords. Only one of the applications users we interviewed has changed their password since the system was initiated.
- Although controls exist to limit printing of reports, any user having general netware printing⁶ capability can print the screen contents. Additionally, all users have the ability to download screen content onto an electronic notebook⁷ which could then be mailed electronically to a third party. None of the users we interviewed stated that they were aware of this capability. Additionally, White House officials told us that every month they review a sample of outgoing e-mail traffic to identify inappropriate use of the electronic mail system and to comply with records management requirements.
- Most importantly, there is no audit trail. Although Sybase 10 has this capability, we were told it has not been turned on because it would inhibit system performance. The Sybase audit capability would allow system administrators to monitor and react to attempts to log on and log off the

⁶Having the capability or authority to output data over the network to a printer.

⁷A simple word processing program that allows users to copy information from one program onto the notebook.

system; execution of update, delete, and insert operations; restarts of the system; execution of system administration commands; and changes to system tables. Without this feature, data base administrators are limited in their ability to ensure that users are properly accessing and using the system.

Mr. Chairman and Members of the Subcommittee this completes my testimony. I will be happy to answer any questions you may have.

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

**U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20884-6015**

or visit:

**Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC**

**Orders may also be placed by calling (202) 512-6000
or by using fax number (301) 258-4066, or TDD (301) 413-0006.**

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

<http://www.gao.gov>

**United States
General Accounting Office
Washington, D.C. 20548-0001**

**Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100**

**Official Business
Penalty for Private Use \$300**

Address Correction Requested