



December 13, 1999

CENSUS 2000 INFORMATIONAL MEMORANDUM NO. 35

MEMORANDUM FOR      Distribution List

From:                      Susan Miskura *SM*  
                                 Chief, Decennial Management Division

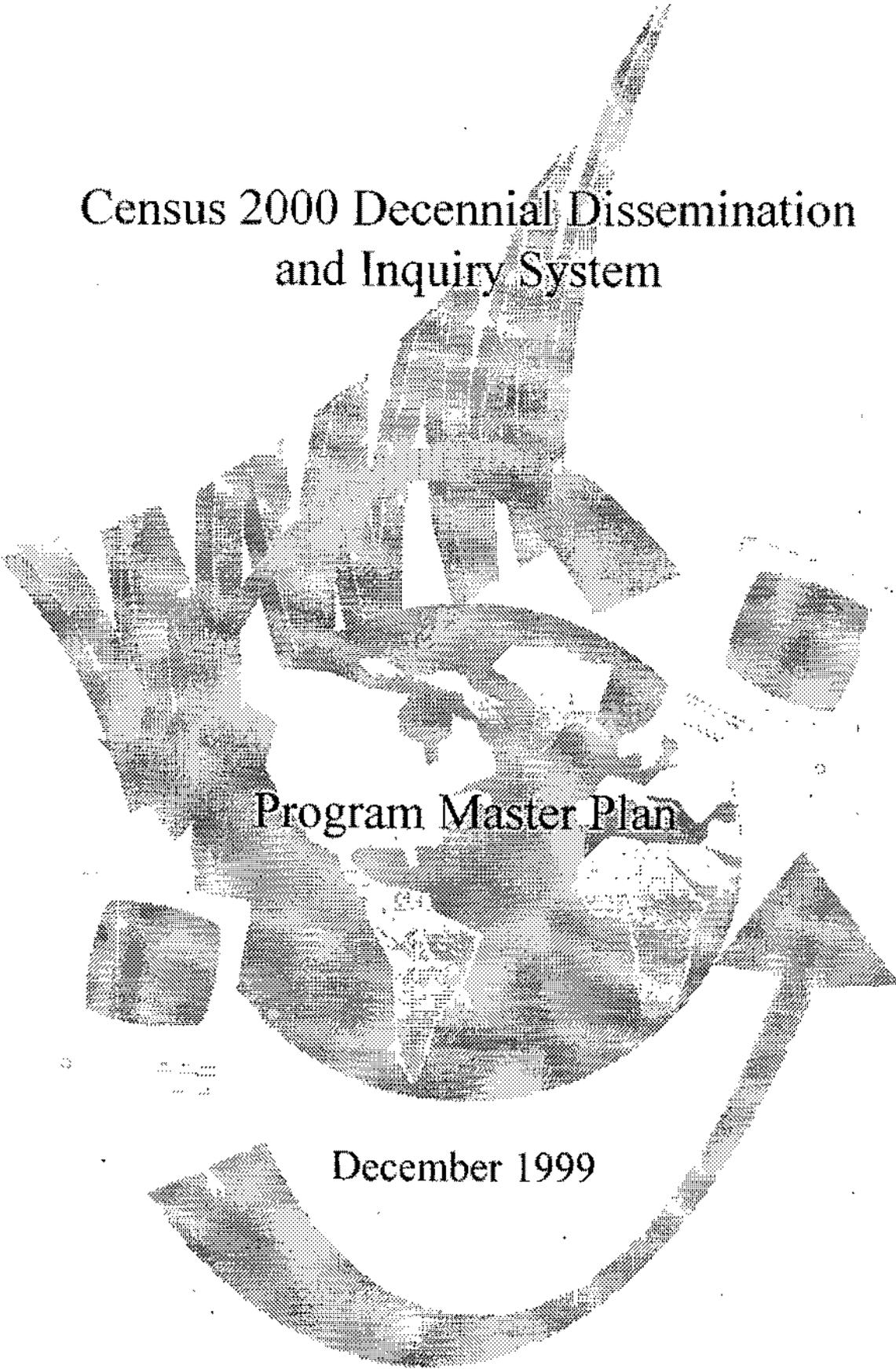
Subject:                    Program Master Plan: Census 2000 Decennial Dissemination and  
                                 Inquiry System

Contact Person:         Ramala Basu, Content and Data Products Branch, Decennial  
                                 Management Division, Room 1422-2, (301) 457-4716

The Program Master Plan (PMP) for the Census 2000 Decennial Dissemination and Inquiry System is attached.

Please forward any inquiries to the contact person for this PMP, Ramala Basu.

Attachment



Census 2000 Decennial Dissemination  
and Inquiry System

Program Master Plan

December 1999

## Table of Contents

I.	Introduction .....	1
A.	Vision of the American FactFinder System .....	3
II.	General Design and Workflow .....	3
A.	Confidentiality/Disclosure Protection .....	4
B.	Development Teams and Roundtable Discussions .....	4
C.	External and Internal Customer Input .....	6
D.	Testing and Delivery Stages of the System .....	7
1.	DADS96 Beta Testing .....	7
2.	DADS97 Beta Testing .....	8
3.	Beta Testing Survey .....	9
4.	Input from State Data Centers .....	10
	Input/feedback on data products .....	10
	Comments on the American FactFinder system .....	11
5.	DADS98 Requirements Analysis Process .....	12
6.	A key change to the DADS98 development process .....	13
7.	American FactFinder as a Production System .....	13
III.	Puerto Rico and Island Area Data on the American FactFinder .....	14
IV.	Cost Assumptions .....	14
V.	Schedule/Responsibilities .....	15
	Attachment 1: Visions for the American FactFinder System .....	16
	Attachment 2: Paper By Marian E. Brady .....	18
	Attachment 3: Bibliography for DADS PMP .....	26

# Decennial Dissemination and Inquiry System Program Master Plan

## I. Introduction

The Census Bureau collects large volumes of data that require a large and complex system of dissemination. The federal government's reinvention and reengineering policies and budget cutbacks in the 1990s led the Bureau to seek to utilize new technologies for disseminating census information faster, to more people, and electronically, using the Internet. In the past, information such as microdata, aggregated data, publications, software products, and custom software applications were made available to data users through a variety of dated mechanisms (e.g., paper copy, mail, floppy disk, tape, and download over modem technology). The most frequently cited complaint was the Bureau's inability to provide timely data products in flexible, easy-to-use, and cost-efficient format. In addition, the Census Bureau analysts had restricted or less access to the data sets required for product development, and it was difficult to provide efficient customer service without easy access to the data bases.

The Bureau designed the American FactFinder System<sup>1</sup> to be an interactive electronic system to allow efficient and cost-effective access to data generated by the various program areas of the Bureau (i.e., data sets from decennial censuses, economic censuses and surveys, demographic surveys, and the American Community Survey).<sup>2</sup> In addition, the American FactFinder complied with directives, mandates, and standards for the Federal Geographic Data Committee for the dissemination of geographic data. The overall objective is to provide one general electronic system for the dissemination, inquiry, and access of Census Bureau data for both internal and external users. This Internet system implements a process capable of efficiently disseminating census information in a consistent manner and realizes the following benefits: (1) better user interaction, service, and response time, (2) increased responses for the Bureau's collection efforts as a result of users/customers becoming more familiar with Census products and their inherent value, and (3) the ability to effectively advertise and increase public awareness about the Bureau's products and services.

---

<sup>1</sup>In the developmental stages, the American FactFinder was known as the Data Access and Dissemination System (DADS).

<sup>2</sup>See Table 1 for the stages in the development of the American FactFinder system.

**Table 1: Stages in the Development of the American FactFinder System**

<b>System Development</b>	<b>Functionality</b>	<b>Available Data</b>
DADS96 Prototype <i>March - September 1996</i>	<ul style="list-style-type: none"> <li>• Speed</li> <li>• Data warehouse structure</li> <li>• Technical architecture</li> <li>• Advanced functions</li> </ul>	<ul style="list-style-type: none"> <li>• 1990 Public Use Microdata Sample (PUMS) Files - 1% and 5%</li> <li>• 1990 Summary Tape Files (Sample only)</li> <li>• Samples of current Internet products</li> </ul>
DADS97 Prototype <i>October 1996 - September 1997</i>	<ul style="list-style-type: none"> <li>• Explore advanced software - Java</li> <li>• Improve and extend functions</li> <li>• Add thematic mapping</li> <li>• Implement metadata model</li> </ul>	<p><i>Content of DADS96 plus:</i></p> <ul style="list-style-type: none"> <li>• 1990 Summary Tape File (100%)</li> <li>• 105<sup>th</sup> Congressional District data</li> <li>• More samples of current Internet products</li> </ul>
DADS98 Production* <i>October 1997 - September 1999</i>  *At this stage, the application of this system was referred to as the <b>American FactFinder</b> .	<ul style="list-style-type: none"> <li>• Novice/expert paths</li> <li>• Improved interface</li> <li>• Confidentiality filters for custom queries</li> <li>• Improved mapping and integration of geographic components</li> <li>• Selected basic derived measures</li> <li>• Scale architecture</li> </ul>	<p><i>Content of DADS97 plus:</i></p> <ul style="list-style-type: none"> <li>• Census 2000 Dress Rehearsal products and related geographic products</li> <li>• 1997 Economic Census products</li> <li>• 1997 American Community Survey (ACS) Test data</li> <li>• 1998 ACS data</li> <li>• Economic surveys and Foreign Trade data dissemination prototypes are being developed.</li> </ul>
DADS 2000 Production <i>October 1999 - June 2000</i>	<ul style="list-style-type: none"> <li>• Scale architecture</li> <li>• Integrate Census Bureau's web site</li> <li>• Incorporate pricing mechanisms</li> <li>• Test query/results filters</li> </ul>	<p><i>Content of DADS98 plus:</i></p> <ul style="list-style-type: none"> <li>• Census 2000 data</li> <li>• 1999 and 2000 American Community Survey data</li> </ul>

*Source:* This is an updated version of the Table 1 in Marian Brady's paper in Attachment 2.

## **A. Vision of the American FactFinder System**

The key functional components of the envisioned<sup>3</sup> system are—

1. Provide users with access to census products, such as statistical briefs and abstracts, area profiles, economic indicators, press releases, summary data, and geographic files, maps, and more.
2. Provide users with access to Census Bureau data and allow creation of customized products.
3. Announce enhancements and changes made to the American FactFinder system.
4. Provide users with an on-line help system for using the system and accessing census data.
5. Provide users with an on-line feedback system for evaluations and suggestions for products and the system.
6. Provide users with links and pointers to relevant non-census federal data sets and non-federal data sources.
7. Allow users access to census data via multiple methods, such as Internet, Intranet, toll free numbers, and intermediaries such as the State Data Centers and their affiliates, Census Bureau's regional offices, libraries, etc.

## **II. General Design and Workflow**

The Census Bureau set up design and implementation teams to develop this vision into a complete proposal defining the main features of the system, including a scenario of how users would access the system, how data would be delivered, and a timeline for implementation. The system must be completed in time to serve as the vehicle for disseminating data from Census 2000 and from the American Community Surveys and will be fully tested in 1998 during the Census 2000 Dress Rehearsal operations.

The basic principles that guided the development of the system architecture were as follows: (1) disclosure protection would be built into the system and base files so that individual products from the system do not require review for disclosure; (2) all data sources with comparable levels of geographic detail would be integrated into the system; (3) geography would be the integrating principle for the data; (4) metadata, including item definitions, descriptions of edits and

---

<sup>3</sup>See Attachment 1 for a list of visions for the American FactFinder system developed by the Census Bureau in May 1995.

imputations, estimates of sampling error or the ability to calculate such estimates, and pointers to related items in other data sets in the system, would be available integrally with the data; (5) the system would provide direct access to summary data, public use microdata samples, and to a process for specifying special tabulations; (6) there would be limited standard prepackaged data and no standard printed reports except a few profile reports necessary to show appreciation to the public for cooperation; (7) development of the dissemination system would be linked with planning for decennial census and continuous measurement processing; (8) both internal and external experts would participate in defining and developing the system; and (9) the system would be accessible to the widest possible array of users through the Internet and available intermediaries, including State Data Centers and similar groups, libraries, universities, and other private organizations.

#### **A. Confidentiality/Disclosure Protection**

The Census Bureau planned to design the American FactFinder in such a way that disclosure protection, such as data swapping and/or recoding sensitive variables that could disclose confidential data, would be built into all of the data sets accessed by the system. One of the special capabilities of the system is to produce extracts and tabulations from microdata files. The dissemination of tabulations on-line from the full microdata files requires special techniques for disclosure protection.

In January 1998, the Bureau created a team of Census statisticians and systems analysts to develop recommendations for protection of disclosure through custom tabulations derived from Census 2000 data using the American FactFinder. The team recommended a "query filter" to restrict queries that are permitted to access the full microdata files and a "results filter" to restrict the details of tables generated from the full microdata files. These techniques were built into the system and tested in 1999 for use in disseminating Census 2000 Dress Rehearsal and Census 2000 data. The Bureau also hired external experts to take part in testing the disclosure protection techniques.

#### **B. Development Teams and Roundtable Discussions**

1. The Bureau's program development plan for this electronic data access and dissemination system involved three teams—technical development team, user requirements team, and the Internet support team.

The Technical Development team conducted research and procured technical support, hardware, and software for the development of the American FactFinder. It also designed and developed the American FactFinder prototypes.

The User Requirements team facilitated the identification of user and technical requirements and outreach and promotion activities, such as presentations, seminars, and the like. It also facilitated the planning and monitoring of testing the American

FactFinder prototypes as they became available.

In 1994, the Internet Support team established the Bureau's Internet site, which was a recipient of Vice President Gore's "Hammer" award for improving government operations. The primary focus of this project was to develop a cost-effective strategy for disseminating census data including microdata, aggregated data, organizational data, publications and analysis, software products, and custom software applications. The Bureau's Internet system was designed around an integrated network of UNIX workstations, personal computers, and servers connected to the "outer" Internet network. The "outer" Internet network was connected to the Census Bureau's internal local area network (LAN) by a network router (known as "firewall") that protected the Bureau's sensitive assets.

2. In May and June of 1995, the Bureau held roundtable discussions to provide initial guidance to the American FactFinder working group. There were two roundtables—a user requirements roundtable and a technical requirements roundtable—represented by staff from various directorates of the Census Bureau. The main purposes for these discussions were to—(1) lay the foundation for a communication process that would encourage Bureau-wide participation in the design and development of the American FactFinder and (2) facilitate an initial study of the vision and principles under which the system would be designed and developed.

Results from the discussions provided input to user requirements, technical requirements, and policy issues. There was broad consensus that this system could provide many benefits, such as—(1) faster access by internal and external customers to census information, (2) flexibility that would allow customization of data extraction and tabulation, and (3) cost savings by eliminating or minimizing pre-specified printed reports.

The roundtable discussants also agreed that the process to move the American FactFinder from a concept to a reality should involve soliciting direct inputs from customers in designing the system; and laying the groundwork for resolving technical and policy issues that could be barriers to the success of the system.

3. In May 1997, the Census Bureau organized the 1997 National Conference on Census 2000 Partnerships to meet with private sector organizations (for example, education, race and ethnic groups, and trade). The main objective was to discuss their data needs and how an electronic system, such as the American FactFinder, could be beneficial to them. The Bureau opened an ongoing channel of communication with these organizations to receive their recommendations or comments on the types of data to be pre-tabulated in the American FactFinder and to inform them on the progress of the system.
4. In June 1997, the Bureau consulted with Dr. Ben Schneiderman on the American

FactFinder system development activities and tested his suggestion during the DADS98 beta testing. Dr. Schneiderman heads the Human-Computer Interaction Laboratory at the University of Maryland in College Park, Maryland. He made the following proposals for improving the system:

- Developing more succinct user profiles and task classes. The Bureau's existing user profiles can be enhanced by utilizing more of the enterprise information about data users that is currently collected from the Internet usage logs, the Internet user survey mechanism, and anecdotal feedback from the Census Bureau's electronic mail post offices. According to Dr. Schneiderman, the Bureau must more clearly understand who its users are and what they do with census data *at the task level*. This information is vital to developing relevant data products and dissemination functions for Census 2000 and other censuses and surveys at the Bureau.

The Census Bureau conducted over 50 interviews with data users in December 1997 to find out what they do with census data. The interviewees represented a full cross-section of the customer typology developed by the Bureau's Marketing Services Office. The Bureau used the results of the interviews to design the gateways of census data as presented on the main screen of the American FactFinder.

- Continue Usability Studies. The Bureau needs to group information on user tasks into classes representing tasks performed by typical user communities and those that are common to all. This will build on the user requirements collection process and will benefit the design process. The Bureau can then design a system interface that most effectively accommodates the various user communities by providing different gateways to data products. Furthermore, the Bureau can test the system designs using a standard set of task and benchmark user performance as it progress through development.

During the development of the American FactFinder, the Bureau conducted three rounds of usability tests—(1) one to validate the high level architecture design and structure, (2) one to evaluate detailed screens within each path to data provided on the main page, and (3) a third one to evaluate the Advanced Query tool that provides access (with disclosure protection) to the edited detail files for custom tabulations. Each round of testing resulted in changes to the system.

- The American FactFinder system's introduction screen be redesigned. The Bureau incorporated this proposal during the beta testing of the DADS98.

### C. External and Internal Customer Input

1. In the fall of 1995, to carry out the roundtable recommendation to engage

customers in designing the American FactFinder, the Bureau facilitated 12 focus groups, including internal and external customers, that represented a full cross-section of its user community to collect information on their data requirements before designing the data access and dissemination system. A standard set of 25 questions were asked of each group. The questions covered topics such as product types, output media, geography, confidentiality, cost, and user access. Using the list of responses, the American FactFinder working group developed external and internal frequency matrices for all responses to each question to identify common and unique needs and themes.

The Bureau identified three tiers of external customers—(1) power customers, such as Congress, the Department of Commerce, the Office of Management and Budget; (2) money customers, such as other federal agencies, state and local governments, the media, and academia; and (3) service customers (most of the public data users fall into this category). The three categories of internal customers were—(1) power customers (e.g., the Office of the Director, Congressional Affairs Office, etc.) who provide services to the external power customers; (2) money customers (e.g., the Demographic Surveys division, the Field Division) providing survey or data processing services to the external money customers; and (3) service customers (e.g., Customer Service Center, Population Division's Information Staff, and the Field Division's Information Services Staff) who provide services primarily for external service customers.

2. The external and internal customer focus groups were asked questions covering topics such as product types, output media, geography, confidentiality, cost, and user access. Using the list of responses, the American FactFinder working group identified common and unique needs and themes. The results revealed that the data users would like the American FactFinder to provide—(1) predefined products and services, (2) simpler user-defined products, and (3) complex user-defined products. To meet these requirements, the American FactFinder must be supported by a successful and detailed geographic infrastructure. Users also identified several basic requirements—(1) the system should be simple to use and support print-on-demand, (2) provide on-line help and training, and (3) notify users about size, cost, and time it would take to receive requested products prior to transmission.

#### **D. Testing and Delivery Stages of the System<sup>4</sup>**

1. **DADS96 Beta Testing.** In September 1996, with a contractor, the American FactFinder working group and the technical team developed the first American

---

<sup>4</sup>For a summary of the schedule, general goals, and content of each iteration, see page 19 in Attachment 2.

FactFinder prototype using 1990 census data to provide a "proof of concept" for the basic design, technology, and functionality of the American FactFinder system.

2. **DADS97 Beta Testing.** In October 1997, the second prototype of the American FactFinder was developed using a scalable technical architecture for the Census 2000 production system. This prototype, with its enhanced functionality and user friendliness, supported the 1997 Economic Census and the American Community Survey.

The prototypes focused on easy access, manipulation, retrieval, and tabulation of detail census data operating in a high-performance client/server environment. The contractor provided expert technical advice in conducting research applicable to the design and development of the prototypes. Throughout the development phase of this system, the contractor worked closely with Census staff in all aspects of system development, particularly regarding interpretation of specifications and testing of the prototypes. The Bureau's technical personnel approved each phase of the project before beginning subsequent phases.

The Bureau adopted a widely recognized and proven structured approach (called CASE\*Method<sup>5</sup> and Oracle's latest enhancement of this methodology called Custom Development Method) to develop its prototypes and final system for the American FactFinder. The prototypes tested high-level functions of the system as listed below—

- Identifying and maintaining information about users accessing the American FactFinder system.
- Providing access to pre-defined census products, such as statistical briefs and abstracts, area profiles, economic indicators, press releases, summary data, geographic profiles, maps, and more.
- Providing access to Census Bureau data for creation of customized products.

---

<sup>5</sup>CASE\*Method is a structured approach to engineering systems in a data processing and warehousing environment. CASE stands for Computer-Aided Systems Engineering and consists of seven basic stages—strategy, analysis, design, build, user documentation, transition, and production. Specific activities are performed at each stage. The Bureau followed the basic stages for designing and implementing its first and subsequent prototypes. The agency also purchased Oracle's CASE tool called Designer/2000 for modeling activities and to store a repository of American FactFinder metadata and documentation. The CASE\*Method and Designer/2000 support two paths of the system—the functional side (front end of an application) and the data side (back end of an application).

- Announcing enhancements and changes made to the system.
- Providing on-line help for using the American FactFinder system and accessing data.
- Providing an on-line feedback system for evaluations and suggestions of the American FactFinder system and products.
- Identifying non-Census Bureau sources of data and providing links and pointers to relevant non-Census and non-federal data sources.
- Allowing access to census data via different means, such as Internet, Intranet, and toll-free telephone numbers, to reach the widest population of potential users.

After identifying and defining census products and services that would be made available through the American FactFinder, the working group and the technical team developed the prototypes to provide a "proof of concept" for the basic design, technology, and functionality of the system.

3. **Beta Testing Survey.** In February 1997, the Census Bureau conducted a 3-week long survey of beta testing the DADS97 prototype. Participation was strictly voluntary and by invitation to a variety of stakeholders including the State Data Centers, the DADS96 beta test group, the 1995 focus group participants, Census Advisory Committee members, and other interested volunteers. Access to the system was controlled by user name and password. Of the almost 300 invited, slightly more than half (159) established user accounts; 84 percent registered as users, 77 percent logged in, 24 percent utilized feedback mechanism in application, 21 percent completed on-line evaluation forms, and 9 percent sent electronic mail.

The top three groups that participated were—governments at all levels (45%), universities (25%), and libraries (8%). Other participants represented the media, businesses, non-profit organizations and individuals. In addition, 41 State Data Centers, two Census Information Centers, and one Business and Information Center participated in the testing. Total respondent burden for completing the evaluation was 22.2 hours.

**Survey Results.** The following comments were made by the majority of the survey participants:

- User interface was too difficult for novice users.
- System performance was unacceptably slow.

- Client-side personal computer (PC) requirements were too advanced.
- There was no cancel or stop button.
- Not enough system feedback.
- User interface design assumes knowledge of census products and data.

The results were used to improve the design process for the DADS98 system and to leverage the allocation of resources for user interface design. The major lessons learned also drove client side technology decisions for DADS98 by targeting development for commonly used browsers.

4. **Input from State Data Centers.** In the fall of 1997, two State Data Center representatives came in for six weeks to review DADS97 and the Census 2000 product proposal. They had one-on-one sessions using the system and provided feedback on what they liked or disliked, or found confusing, etc. They had a two-part mission for this internship at the Census Bureau—(1) provide input/feedback on the predefined products associated with Census 2000 and (2) work with the DADS prototype system and provide input/feedback on what they liked or did not like within the system and if there was any new requirement.

Input/feedback on data products. The State Data Center representatives made the following suggestions for Census 2000 data products:

- The Bureau develop a condensed standard profile report either in addition to or in place of the 5-page profile reports currently proposed to be produced as print-format products.
- A major barrier to the use of census data at the smaller units of geography (almost anything below county) has always been the difficulty of matching up census geography across decennial censuses. With the TIGER® technology, the Bureau should consider a project that would involve creating versions of the 1990 summary tape files for Census 2000 geographic units (e.g., block group, census tract, and place).
- ZIP Code data are one of the most frequently requested data. It will earn the Bureau big points with the public if they can enter their ZIP Code and get a quick profile for their "neighborhood."
- The Bureau has never developed a simple, standardized key for referencing the census geographic units (i.e., counties, census tracts, county subdivisions, places, etc.). The agency should develop a single-field Uniform Geographic Locator (UGL) to provide linkages required to fully utilize census data in a hyper-linked environment. Every summary record on every file created by the Bureau should have a UGL code. This will greatly simplify efforts to link

data from different data files.

Comments on the American FactFinder system. The State Data Center representatives provided the following high-level summary of general impressions and suggestions regarding the AFF system prototype 2:

- Very much liked the vision of the AFF as a universal system that can create and disseminate the Bureau's vast array of data. It is extremely ambitious and appears to be surprisingly doable.
- Saw a potential problem with the user interface of the prototype 2 which was being built with a new set of JAVA-based "widgets." While the system may be invoked via the Internet and may have a URL associated with it, it did not have the look and feel of the familiar HTML web browser-based interface.
- Scroll bars, drop down lists, and clicking on icons did not work reliably.
- It was unsettling to see that, at the end of the second prototype, no thought was given to the complex issues related to generating "on the fly" printed reports.
- The Table Viewer had no provisions for displaying FIPS or other geographic codes in the output viewer window. This was perhaps evidence of a lack of a good user-feedback loop.
- The exclusion of aggregation capability in Table Viewer was considered a serious hole in the American FactFinder system. Aggregation is an absolutely critical tool. Leaving it out is inconsistent with the overall vision of DADS as a tool that provides all the basic processing functionality.
- Did not like the name "Table Viewer," because it sounds as if it is only or primarily a tool for simply "displaying" pre-tabulated data in very pre-determined report-like formats. What it actually can and should be is a very general purpose data access tool that can manipulate and aggregate data with the same kind of flexibility that the Query Builder provides for microdata.
- The realization of the vision that the users will really be able to get, with reasonable ease, timeliness, efficiency, and cost, just the data they want for just the geography they need had a number of potential barriers such as disclosure avoidance and cost.
- There is no help button or pull-down menu that gives users descriptive texts, such as discussion of the geographic summary levels or public use microdata

area (PUMA) codes. The Details button gave descriptive texts about a data entity that were more formal than informative.

- To improve the process of gathering and implementing "user requirements," the feedback mechanism should involve developers sitting next to the users as they use the system and discuss what they like or do not like; have several users with different specialties and representing different portions of the user community on the development team; and implement a *priority* tagging of user requirements.

The Bureau made progress in all key areas identified in the comments. For example, the Quick Table have been developed as "on the fly" reports. Subsequent releases of the American FactFinder will include aggregation and manipulation capabilities for summary file data. The Bureau also renamed and redesigned the Table Viewer tool; developed a disclosure protection system built into the Advanced Query tool; and implemented a feedback mechanism to allow the public to comment on the American FactFinder. Additional requirements will be developed on a topic by topic basis as new features and functions are requested.

5. **DADS98 Requirements Analysis Process.** The DADS98 was a joint effort with IBM as a major partner. The build phase consisted of multiple short-built iterations that included both designing, building, and testing of components until the complete system was assembled and integrated. The cyclical build process allowed for continuous feedback and evaluation of the current system.<sup>6</sup> The process included gathering requirements from key stakeholders, subject matter experts, and potential users of the American FactFinder system via interviews and joint application design (JAD) sessions. The Bureau conducted 60 interviews with potential users in Denver, Dallas, and Detroit during the week of December 15, 1997. Representatives from the State Data Centers, local government, educational institutions, media, and community organizations were included in the interviews. Forty-nine of the interviews were conducted face-to-face and 11 were done by telephone. The length of each interview was approximately one hour. The interview team consisted of one Census Bureau staff, an IBM performance analyst, and an IBM technical analyst.

The interviews focused on both a task analysis and skills assessment of data users. The results were used to validate and refine existing user requirements and to fulfill the overall requirement that the system be user-centered rather than data-centered. The interviews also facilitated the transfer of business process knowledge to on-site contract staff responsible for systems integration and development. The interview results directly influenced development of a user

---

<sup>6</sup>For a diagram of the cyclical process, see Figure 1 on page 20 of Attachment 2.

interface design. They showed that most users wanted information about their community and that they like to see maps of their area. As a result of these interviews, additional features such as the ability to sum data and calculate a percentage were included in the system design.

The results of the interviews were compiled and incorporated into the Use Case documentation.

Executive interviews were conducted from November 1997 through January 1998. The Bureau prepared and distributed a summary reflecting the priorities of those interviewed and indicating the direction in which the Census Bureau should proceed.

The JAD sessions were held in December 1997 to gather requirements from subject matter experts. The two-way JAD sessions focused on defining requirements from a user perspective with respect to how the user interacts with the system. Use cases and scenarios were developed prior to the session and then reviewed and updated during the session. Subject matter experts from a cross-section of Census Bureau personnel also were asked to participate in the JAD sessions to ensure that the requirements would match the needs of a wide audience.

Documents prepared from these sessions formed the basis for sizing and scoping the project for DADS98 and led directly into the design phase of the project. Project scoping has been based on the priorities of the Census Bureau and resources available to complete the application for the Census 2000 Dress Rehearsal.

6. **A key change to the DADS98 development process.** The Bureau separated the user interface design from the development process and implemented joint design reviews with stakeholders/data providers to gain concurrence during the design and development process for DADS98. Designated points of contact facilitate this process and participate in working meetings between the larger design review sessions.
7. **American FactFinder as a Production System.** The development of the American FactFinder as a production system has been completed, and it will be made available in three releases in 1999. During the first quarter of 1999, the American FactFinder is fully supporting the Census 2000 Dress Rehearsal. It also provides access to 1990 census, 1997 Economic Census, and 1996 and 1997 American Community Survey (ACS) data. During the second quarter of 1999, the system will provide access to the 1990 Public Use Microdata Samples (PUMS) and limited access (to password holders only) to full microdata files.

The system will be made available to the public through several releases in 1999 and 2000 with progressively increasing functionality. The Census Bureau will seek user feedback about each release to improve and enhance the system throughout 1999 and 2000 in preparation for Census 2000 data dissemination.

### **III. Puerto Rico and Island Area Data on the American FactFinder**

In response to a request from the Government of Puerto Rico to the Secretary of Commerce, the Census Bureau has taken necessary steps to include Census 2000 Puerto Rico data in the national summary data products and has developed the technical requirements for presenting Puerto Rico data on the American FactFinder. In addition, the Bureau is planning to incorporate Spanish in the key points of the user interface in the American FactFinder for easy access to Puerto Rico data.

IBM has developed an extension to American FactFinder to support the dissemination of appropriate Census 2000 data for the Puerto Rican geography in Spanish. The Spanish extension will be limited to those functions that apply to the Census 2000 Puerto Rico data being disseminated. All American FactFinder users will be granted access to the Spanish extension of American FactFinder for Puerto Rico. The Spanish extension of the American FactFinder will support the following items only for Puerto Rico data and its geographies: (1) products, (2) maps, (3) community profiles, (4) quick tables, and (5) detailed tables. It also will display text in Spanish for map legends, help, data, report titles, column headings, disclaimers, and warning/error messages. The Spanish extension of the American FactFinder will not provide access to Tier 3 data or metadata.

(Note: For a list of source materials used in this plan, see the bibliography in Attachment 3. Also see the 1999 and 2000 Operational Information Technology (IT) Plans for DADS.)

### **IV. Cost Assumptions**

In order to estimate the costs to develop the American FactFinder system, Advanced Resources Technologies, Inc. and its subcontractor Gunnison Consulting Group analyzed the Census Bureau's data dissemination processes, including defining the extent of the requirement for an automated system, the organization and management structures required to support such an activity, the underlying policies and standards, and the economic implications of the entire process. The following table shows the actual and estimated costs for developing the new, electronic method for data dissemination for the period of 1998 through 2003:

Fiscal Year	Actual Cost	Estimated Cost
1998	\$10,700,000	
1999	\$13,300,000	
2000		\$19,500,000 (includes \$1.4 million for Puerto Rico concept)
2001		\$11,500,000
2002		\$11,500,000
2003		\$11,500,000

### V. Schedule/Responsibilities

The activities for the strategy, analysis, and design stages for the development of the American FactFinder system already have been completed. The following schedule depicts the key phases of the American FactFinder development and implementation:

Develop and deliver American FactFinder Prototype I	September 30, 1996
Deliver American FactFinder Prototype II	October 31, 1997
Deliver American FactFinder 98, Release 1	January 12, 1999
Deliver American FactFinder 98, Release 2	March 15, 1999
Load Census 2000 Dress Rehearsal 100 percent edited detail file (HEDF)	April 1999
Deliver American FactFinder System 2000	June 30, 2000
Load Census 2000 HEDF	June 26, 2001
Load Census 2000 SEDF	August 14, 2002
Apportionment Counts	December 31, 2000
Redistricting Data	April 1, 2001
Census 2000 products - 100%	April-December 2001
National Summary Files with urban/rural data	Spring 2002
Census 2000 products - sample	Dec. 2001-March 2003

## **Attachment 1: Visions for the American FactFinder System**

May 1995

The U. S. Census Bureau will develop and implement a plan and system for data access and dissemination focused on the 2000 Decennial Census and Continuous Measurement data sets, but with the ability to accommodate other data sets having geographic detail, like those produced from the Economic and Agriculture Censuses.

A design and implementation team or set of teams will develop this vision into a complete proposal defining the main features of the system, including a scenario of how users would access the system, how data would be delivered, and a timeline for implementation. The system must be completed in time to serve as the vehicle for disseminating data from the 2000 census and from the continuous measurement program; that is, fully tested during the 1998 census dress rehearsal operations and fully operational by early 2001.

The team(s) will enlist the assistance of internal and external users to test the system, and refine it as needed. The team will identify and, as appropriate, make use of all relevant Bureau work already underway or planned so that we have a coordinated, corporate approach to data access and dissemination. They also will specify resource requirements for the full system and recommend where the system should be maintained once developed. The team(s) will report regularly to a steering committee including senior level Bureau management and technical experts.

The principles under which the system will be designed and developed include the following:

1. The system will provide direct access to a limited number of data summaries, to public use microdata samples, and to a process for specifying special tabulations from confidential files. It will be accessible to the widest possible array of users through the Internet and all available intermediaries, including State Data Centers and similar groups, libraries, universities, private firms, and so forth. Internal access through open systems is assumed.
2. There will be limited standard, prepackaged data summaries, and no standard printed reports except those few profile reports necessary to show appreciation to the public for cooperation. All access and special product preparation will be through the dissemination system. All reports, files, etc. will be prepared on demand and rapidly, even with high demand.
3. Disclosure protection will be built into the design on the system and base files so that individual products from the system do not require review for disclosure. This implies that a confidentiality edit is performed on the basic file to minimize or eliminate additional confidentiality reviews. Work on appropriate techniques will occur outside this group.

4. All data sources with comparable levels of geographic detail will eventually be integrated into the system (e.g., economic census files, decennial census files, population estimates files.)
5. Geography is the integrating principle for the data, using both standard geographic areas and nonstandard geography based on centroids or coordinates, as appropriate.
6. Meta data, including item definitions, descriptions of edits and imputations, and pointers to related items in other data sets in the system, will be available integrally with the data. Estimates of sampling error/uncertainty or the ability to calculate such estimates will be provided within the data sets.
7. The system and its use will save money compared to the traditional publication program so that resources can be dedicated to educating users and potential users on how to access and use the system to get the products they need.
8. Most transactions will include a fee for service--for file extracts, for printed summaries, for CD-ROM summaries, for graphical summaries, and so forth.
9. Development of the dissemination system is linked with planning for decennial census and continuous measurement processing.
10. Development of ideas and plans will make use of the work already done or underway as part of CENSAS, DAPS90, the 1990 census Data Productions Specification System, the Internet efforts, the reinvention lab on post-collection processing, continuous measurement, the Current Population Survey data access system, the Governments Division work, Survey of Income and Program Participation On-Call, the Population Division data dissemination system, and so forth.
11. Both internal and external experts will participate in defining and developing this system.

## **Attachment 2: Paper By Marian E. Brady**

Marian E. Brady  
Data Access and Dissemination Staff  
U. S. Census Bureau  
FOB2, Rm. 2325  
Washington D.C. 20233  
Tuesday, April 28, 1998

### **The Census Data Access and Dissemination System: From Plans to Reality**

**Abstract:** The advent of the Internet provides an opportunity for dramatic change in the way people access Bureau of Census (BOC) information. As part of its effort to increase access to official demographic, economic, and geographic information, the BOC plans to expand electronic dissemination of data via the Internet through development and implementation of the Data Access and Dissemination System (DADS). DADS will provide timely data to the widest possible user-base through the Internet as well as through intermediaries such as State Data Centers, libraries, universities, and private firms. Many products, such as reports and files, will be prepared on demand. Disclosure protection will be built in to the system design. The primary form of output will be electronic media. Display options will include reference maps and thematic maps. DADS will be completed in a series of stages providing increasing functionality and content. The earliest system release will serve as the vehicle for disseminating data from the Census 2000 Dress Rehearsal and the 1997 Economic Census at the end of 1998.

#### **BACKGROUND**

Prior to 1994, the BOC relied on traditional methods of data dissemination, issuing information in printed publications and reports as well as in machine-readable formats, such as tape and CD-ROM. Printed publications were very popular because they reached a wide audience and were easy to use; however, they were late in the product development cycle. Tape products were released earlier than printed products but could only be used by a minority of sophisticated users, particularly secondary distributors of data. For the 1990 Census and the 1992 Economic Censuses, CD-ROMs were a popular, widely utilized format for data dissemination.

The advent and popularity of the Internet provides unprecedented ease of access to information on-line and offers a tantalizing alternative to the traditional methods of data dissemination. A 1994 "proof of concept" BOC Internet site received Vice President Gore's Hammer Award for improving government operations. Currently, the BOC Internet site receives 2.7 million hits per week serving over 100,000 unique customers in the same time period.

By utilizing the Internet for data dissemination, information from data collection activities can be disseminated earlier than ever in the product development cycle and are available to a ever widening audience. With the available and emerging technology, tools to access traditional types of summarized data and products, as well as more flexible, advanced tools that provide the ability to customize queries can be developed. This general strategy forms the basis of the vision of DADS - the Data Access and Dissemination System.

## **SCOPE OF CONTENT DEVELOPMENT**

The DADS staff is tasked specifically with developing a corporate or enterprise-wide system that provides and expands access to both economic and demographic information via Internet. In the short term, through the year 2001, DADS development will focus on dissemination of the results of data collection activities for four major program areas:

1. Decennial, including the 1990 Census, the 1998 Dress Rehearsal and the Census 2000,
2. the 1997 Economic Censuses,
3. the American Community Survey (ACS), and
4. Geographic Reference Maps and Files.

Detailed subject matter, content, and format of the products are decided by individual program areas; they also make decisions regarding the extent of need or demand for dissemination of census and survey results on traditional media. Product development parallels the system development of the DADS functions and tools.

## **SCOPE OF FUNCTION DEVELOPMENT**

Key requirements for DADS were collected in the Fall of 1995 through a series of eight meetings with a cross section of external data users and four meetings with internal staff. The results of that process defined a set of expectations for DADS that guide the detailed development of the system.

General characteristics users envisioned for the system are that it will:

- be simple and intuitive to use,
- support a range of users from novice to expert, and
- provide fast and flexible access to all census data.

The key functional components of the envisioned system are that it will:

1. Provide users with access to census products, such as statistical briefs and abstracts, area profiles, economic indicators, press releases, summary data, and geographic files, maps, and more.
2. Provide users with access to Census Bureau data and allow creation of customized products.
3. Announce enhancements and changes made to the DADS system.

4. Provide users with an on-line help system for using the system and accessing census data.
5. Provide users with an on-line feedback system for evaluations and suggestions for products and the DADS system.
6. Provide users with links and pointers to relevant non-census federal data sets and non-federal data sources.
7. Allow users access to census data via multiple methods, such as Internet, Intranet, toll free numbers, and intermediaries such as the State Data Centers and their affiliates, BOC Regional offices, libraries, etc.

### **SCHEDULE OF DEVELOPMENT**

To implement users requirements, the DADS staff adopted an iterative development process that has the major benefits of :

- providing a proof of concept in a relatively short period of time,
- providing the ability to change the system functionality and presentation (user interface) based on feedback, and
- staying current with changing technology over the entire development period.

The DADS technical development spans a 5 year period that allows for incremental expansion of the system in three key areas of function, content and architecture. A total of four iterations of development will be completed. The first two iterations were exploratory prototypes; the final two iterations will be "production" in that new data and access tools will come on-line for dissemination to the public via the Internet. The schedule, general goals and content of each iteration are described in Table 1 below. With each iteration, functions are improved based on user feedback and testing along with a function by function assessment of resources required and the technical feasibility of implementation.

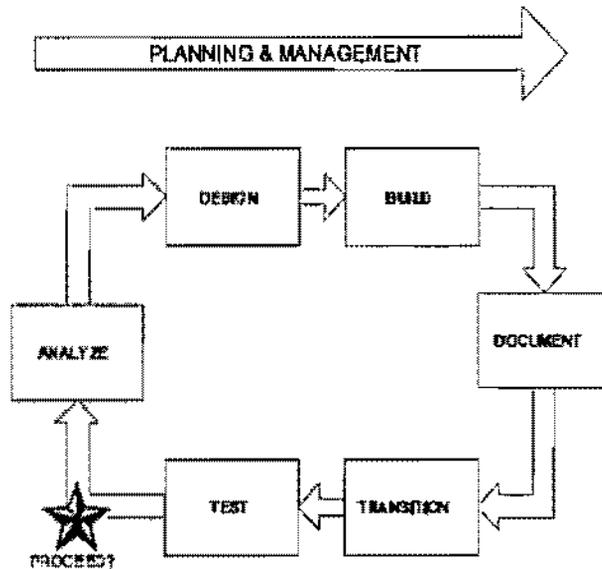
Table 1

Iteration	General Goals	Content
DADS 1996 Prototype  3/96 - 9/96	<ul style="list-style-type: none"> <li>• Speed</li> <li>• Data warehouse structure</li> <li>• Technical Architecture</li> <li>• Advanced functions</li> </ul>	<ul style="list-style-type: none"> <li>• 1990 1% &amp; 5% Public Use Microdata Sample Files (PUMS)</li> <li>• 1990 Summary Tape Files (Sample only)</li> <li>• Samples of current Internet products</li> </ul>
DADS 1997 Prototype  10/96 - 9/97	<ul style="list-style-type: none"> <li>• Explore advanced software - Java</li> <li>• Improve &amp; extend functions</li> <li>• Add thematic mapping</li> <li>• Implement metadata model</li> </ul>	<p><i>Content of DADS96 plus:</i></p> <ul style="list-style-type: none"> <li>• 1990 Summary Tape File (100%)</li> <li>• 105<sup>th</sup> Congressional District Data</li> <li>• 1992 Economic Test Files</li> <li>• 1996 American Community Survey Test Data</li> <li>• More samples of current Internet products</li> </ul>
DADS 1998 Production  10/97 - 3/99	<ul style="list-style-type: none"> <li>• Novice/expert paths</li> <li>• Improved interface</li> <li>• Confidentiality filters for custom queries</li> <li>• Improved mapping and integration of geography components</li> <li>• Selected, basic derived measures</li> <li>• Scale architecture</li> </ul>	<p><i>Content of DADS 1997 plus: *</i></p> <ul style="list-style-type: none"> <li>• 1998 Dress Rehearsal Products**</li> <li>• 1997 Economic Censuses</li> <li>• 1997 American Community Survey Test Data</li> <li>• *1992 Economic Data will not be provided.</li> <li>• ** Includes Geography Products</li> </ul>
DADS 2000 Production  10/98 - 6/00	<ul style="list-style-type: none"> <li>• Scale architecture</li> <li>• Integrate BOC web site</li> <li>• Incorporate pricing mechanisms</li> <li>• Test Query/Results Filters</li> </ul>	<p><i>Content of DADS98 plus:</i></p> <ul style="list-style-type: none"> <li>• Census 2000</li> <li>• 1998, 1999 American Community Survey</li> </ul>

## DEVELOPMENT PROCESS

Each iteration of DADS development entails the general development steps illustrated in Figure 1 below. Contrary to earlier iterations, the DADS98 cycle will be a joint effort with IBM as a major partner. Within the DADS98 cycle, the build phase will consist of multiple, short build iterations that will include both designing, building and testing of components until the complete system is assembled and integrated. At that point, complete systems load testing will be conducted prior to deployment and release of the system.

Figure 1



The cyclical build process allows for continuous feedback and evaluation of the current system as it is built allowing for modifications to the system. In turn, the results of DADS98 testing and evaluations will inform the DADS2000 iteration.

## DADS PRODUCT FRAMEWORK

As stated earlier, individual program areas make detailed decisions about the products and the content, format and media type that they choose to deliver. DADS staff must work closely with their program area data providers to understand the structure of those products. During the early iterations of development, a convenient framework emerged that describes the types of products DADS will deliver

and the associated functions that the system will deliver to access those products. The functions listed in Table 2 are not exhaustive and meant only to provide an overview.

Table 2

	Definition and Examples	Functions
Tier 1	Tier 1: "What you see is what you get". Examples: Press releases, Statistical Briefs, Publications & Reports, Census and You etc.	Key Distinction: No further manipulation possible.  Basic Functions: Browse, Search, Retrieve, Display, Download/Save and Print
Tier 2	Tier 2: Summary Files & Tables Examples: 1990 Summary Tape Files 1 & 3, 1992 All Economic-Wide File	Key Distinction: Aggregate data; limited further processing and manipulation possible.  Basic Functions: all of the above <i>plus</i> limited aggregation & manipulation such as row and column totals, calculation of percentages in DADS; when appropriate, thematic mapping of data; ability to select geographic area and geographic level of choice.
Tier 3	Tier 3: Detailed Data Files Examples: 1990 1% & 5% Public Use Microdata Summary Files, edited detail files (EDFs)	Key Distinction: Record level data files Custom tabulations must meet confidentiality standards  Basic Functions: Assuming confidentiality is met, tabulations, basic select derived measures; when appropriate, thematic mapping

## **DADS ONGOING DEVELOPMENT ISSUES**

Many issues beyond the structured DADS development affect the delivery of the system and are being addressed by the relevant program areas within the BOC. Key issues and the current strategic thinking are briefly described below.

### **Integration with the Current Web Site**

Many users see the DADS development process to be redundant given the existence of the current web site. The public site serves the ongoing needs for dissemination via Internet from many areas of the BOC while the DADS staff systematically builds new data access tools for three designated program areas of the agency. This effort is singular in that:

- 1) it entails systematic exploration and testing of emerging Internet technology via iterative prototyping,
- 2) no other staff are charged with this mission, and
- 3) these tools can be utilized by many other program areas of the BOC.

With the delivery of the DADS2000 system, newly developed tools will be integrated into the current BOC web site.

### **User Interface**

With a greater reliance on the Internet as a means of dissemination, the user interface has become a critical factor in facilitating the successful retrieval and use of information and data. IBM brings to the DADS development both a process and a methodology for designing a user interface that meets the requirements of supporting novice to expert users and providing a "user friendly and intuitive" DADS.

### **Preservation of Confidentiality**

The development of custom query capabilities will incorporate techniques to preserve confidentiality. Early work is focusing on management and filtering of both the incoming query and the out going results. Technically difficult to implement, this option of DADS will require extensive internal testing and evaluation prior to release to the public.

### **Pricing of Products and Services**

Eventually DADS will incorporate a mechanism to collect fees for products and services. Work has begun in earnest on identifying what specific products and service will require charges; once agreement is reached, the DADS staff will explore mechanisms to implement those policies.

### **CONCLUSION**

DADS development is on track to deliver the system envisioned by users in 1995. Development risks have been identified and are mitigated by implementation of the orderly, iterative development and management process. Critical success factors identified early on are being systematically addressed; key among them is the establishment of working partnerships between program areas within the agency. These working relationships are well established with the primary product program areas, the information technology infrastructure staff, and the data quality staff and will ensure the final integrity of the system.

### Attachment 3: Bibliography for DADS PMP

Flyer: Future of DADS of the Census Bureau.

Guidelines for Determining Core Data Elements for Data Dissemination.

Paper: "Design Principles for a Unified Statistical Data/Metadata System," by Daniel W. Gillman, Martin V. Appel, and William P. LaPlant, Jr.

Paper: "Proposal for a Statistical Metadata Standard," by William P. LaPlant, Jr., Gregory J. Lestina, Jr., Daniel W. Gillman, Martin V. Appel, and Technical Research Staff of SRD.

Paper: "Metadata Database Development at the Census Bureau," by Daniel W. Gillman and Martin V. Appel.

Paper: "Towards a Unified Data and Metadata System at the Census Bureau," by Bo Sundgren (Statistics Sweden) and M. Appel, D. Gillman, and W. LaPlant, Jr.

DADS Executive Summary of RoundTable Discussions.

DADS executive summary of roundtable discussions.

External Focus Group Q&As.

DADS Major Milestones.

Handout on FERRET and other technologies used for the Internet web.

DADS Requirements Analysis Documentation.

1995: Ed Spar's article on "Spreading Census Data in the 21st Century."

1995: DADS' External User Requirements Subgroup Participant List and a list of "Limited Access" Focus Group.

1995: E-mail from Sandra Nekovich containing a list of potential focus group attendees. A list of DADS User Requirements Working group.

1995: Internal Focus Group #3 (Q&As on what types of data would you like to see on DADS).

1995: Memo: From Edna Paisano on DADS Focus Groups (providing a list of organizations and a list

of tribal governments with a contact person that should be invited).

1995: DADS Working Group Memorandum from Valerie Gregg (Draft of Summary of Discussion Topics for External Focus Groups; DADS Guiding Principles).

1995: A list of the DADS Working Group members and scheduled seminars.

1995: L. Zayatz's confidentiality report draft.

1995: Pete Bounpane's comments to V. Gregg on DADS document.

1995: DADS Discussion Topics for Focus Groups.

1995: E-mail from Pat Berman for a list of names from the academic/research community for a Focus Group on user requirements for DADS.

1996: Beta Testing DADS and Status of DADS Requirements Transition Activities.

1996: DADS Communication Proposal - Executive Summary.

1996: Memo from Bryant Benton on Internet Roles and Responsibilities.

1996: Minutes from Technical Architecture Meeting with Decennial.

1996: DADS status report.

1996: DADS User Group meeting notes.

1996: DADS Communications Proposal.

1996: Memo from M. Fortier on Internal Testing of the 1996 DADS Prototype.

1996: Memo from J. Long on Internal Testing of the 1996 DADS Prototype.

1996: Alpha and Beta Test Implementation Plan for the 1996 DADS Prototype.

1996: Memo from DADS Metadata Team on Subject Classification for DADS Prototype I.

1997: DADS98 Program Management Plan.

1997: E-mail from J. Kavaliunas on case scenarios for DADS user requirements.

1997: DADS description with background, etc. from the Maize Book.

- 1997: Memo and paper: DADS Prototype 2 Requirements Analysis Report (from M. Fortier).
- 1997: User Documentation: Results, Analysis and Recommendations from User Testing of the 1996 DADS Prototype.
- 1997: JAD Session on Focus Group Participants on DADS Process and cross-functional requirements.
- 1997: Handout on DADS and the Census 2000 Data Products Proposal.
- 1997: DADS98 Comprehensive Development Plan.
- 1997: Confidentiality and DADS meeting notes.
- 1997: Confidentiality and DADS issues - e-mail from Easley Hoy.
- 1997: How DADS and Disclosure Limitation Affect Each Other.
- 1997: 1997 National Conference on Census 2000 Partnerships (Data Dissemination and User Needs).
- 1997: Census 2000 Proposed Products and Dissemination Plans.
- 1997: DADS Prototype 2 briefing - handout.
- 1997: Comments from John Blodgett and Kirin McInnis on the Census 2000 Data Product Proposal.
- 1997: Meeting notes from N. Torrieri on PUMS.
- 1997: DADS98 Requirement Process.
- 1997: Memo: DADS97 Evaluation/Testing and DADS98 Requirements from E. Gomez.
- 1997: 1990 Census P.L. 94-171 Data Official Release Dates - a listing.
- 1997: Handout on DADS Publishing System (DPS) solution.
- 1997: Alternative Publication Solution Approaches for the DADS TAB/PUB.
- 1997: System Architecture Evaluation on the Proposal for a DADS Publishing System (DPS).
- 1997: Monthly PSC report.
- 1997: A list of Targeted Associations for Census 2000 Product Input.

- 1997: Memo: Transmittal of Dr. Ben Schneiderman's Report from M. Brady.
- 1997: Baltimore Public Meeting on Census 2000 Data Plans.
- 1997: A summary of the Focus Group discussion on DADS at the Oct. 1997 APDU meeting by Celia Boertlein.
- 1997: Marian Brady's paper on "The Census Data Access and Dissemination System: From Plans to Reality."
- 1998: E-mail from Sandy Rowland on Confidentiality and DADS Meeting.
- 1998: Notes from the Advisory Committee Meeting (AMA/PAA).
- 1998: Sandy Rowland's paper on "Maintaining the Confidentiality of Census 2000 Data while Releasing Custom Tabulations of the Data Through the DADS."
- 1998: Confidentiality and DADS meeting notes.
- 1998: Memo: DADS MIT Report for March 1998 from E. Gomez.
- 1998: L. Zayatz's e-mail on confidentiality guidelines.



**Section IX.**

**Data Access and  
Dissemination System**

**(DADS)**



## IX. Data Access and Dissemination System (DADS)

### 1.0 Program Area Overview

This section of the Operational IT plan describes the Data Access and Dissemination System (DADS) 1998 Prototype system currently under development. The Plan also describes the infrastructure used to provide office automation services that support the DADS program.

The IT resources described within this plan will provide the following benefits to the Bureau:

- **Consistent data access and dissemination.** The DADS team will identify and, as appropriate, use all relevant Bureau work underway or planned so that a coordinated corporate approach to data access and dissemination is universally utilized.
- **Increased responsiveness.** By providing a means to access census data dynamically, the Bureau will be able to respond to requests for information more rapidly.
- **Reduced costs.** Adopting the most efficient and innovative processes will reduce the time and effort required to make products available.
- **Electronic commerce.** By allowing products to be ordered and paid for over the Internet, DADS will allow the Bureau to become more self-sufficient and increase the customer base being served.
- **Planned changes.** The design and implementation team will enlist the assistance of internal and external users to test the viability of each prototype and accept, reject, and/or refine with each iteration. This iterative process allows the Bureau to adjust to the rapidly changing requirements and technology.
- **User satisfaction.** Including users in the design and testing of DADS and using iterative development will ensure that DADS meets or exceeds evolving user requirements.



The major programs and their products are as follows:

### Decennial Census of Housing and Population

The U.S. Constitution provides for a census of the population every ten years, primarily to establish a basis for apportionment of members of the House of Representatives. Census counts also are required to draw Congressional or State-legislative district boundaries, for allocating Federal and State funds under various grants-in-aid, in formulating public policy at all levels and in private sector planning and decision making. The census of population is a complete count with additional detailed data collected on a sample basis to fulfill Federal legislative and programmatic needs.

A sampling of Decennial products:

- 1990 Summary Tape File 1;
- 1990 Summary Tape File 3;
- 1990 100% Edited Detail File (internal access only);
- 1990 Sample Edited Detail File (internal access only);
- 1990 1% Public Use Microdata Sample; and
- 1990 5% Public Use Microdata Sample

### Quinquennial or Economic Census

Conducted in years ending in "2" and "7", this group of censuses is commonly referred to as the Economic Census. As is shown in the following tables ("Sampling of the Following Report Series" and "Subject Areas"), the Economic Censuses compile information about business establishments, industries, products and local areas. All major sectors of the economy--manufacturers, retail and wholesale trade, service industries, finance, insurance, real estate, transportation, communication, utilities and mining--are covered by these Censuses.

<b>Sampling of the Following Report Series of the 1992 Economic Censuses</b>	
<b>Core Business Statistics Series Advance Report</b>	<b>Reference Series</b>
Comparative Statistics	Industry Series
Bridge Report	Geographic Area Series
Non-employer Report	County Business Patterns
Subject and Analytic Report Series: (a) General Summary and (b) Products Summary	



- provide users with detailed socioeconomic data throughout the decade;
- provide users with information that is more timely than currently available from the Decennial Census;
- improve the infrastructure for the federal statistical system; and
- aid state and local officials in meeting their responsibilities.

The American Community Survey was tested in 1996 and 1997; data for the eight test areas will be available through DADS. Product planning is in progress; in 1996 and 1997 a total of eight areas were tested. Products from those areas will be published by DADS and preliminary planning calls for these three types of products:

- a Public Use Microdata Sample type file;
- summary tables; and
- summary profiles.

#### Customers

DADS customers (or data users; see Figure 1) are from the public and private sectors; they are internal and external to the Bureau and they use census data to govern and make economic policy and business decisions. DADS will be built to satisfy a wide range of users from novice to expert, who will use the system in different ways for various purposes.



The Bureau's internal data customers include staff who design and create summarized data and one-time special tabulations and extracts from a given microdata set to place it within the predefined product side of DADS. These staff members play a critical role in achieving easier access to the Bureau's data and data products and comprise an important segment of the customer market for Bureau data. Bureau subject matter analysts also prepare specifications to respond to inquiries from the Congress, other Federal agencies, State and local governments and other groups. Improved access to data for these customers results in increased efficiency in designing and creating standard products and quicker response to requests for special tabulations. This group represents all major areas of the Bureau, demographic and economic, census and survey.

Additionally, the Bureau supports internal and external data disseminators who assist and provide information to data users. Internally, this group includes Bureau Regional Office staff and other individual divisions and program areas that respond to inquiries for data and data products. Externally, the group includes the State Data Center Network and other distribution networks such as the Federal Depository Library System. They essentially would use DADS as a look-up system, utilizing mostly summarized data, providing help and direction to data users.

### **1.3 DADS Business**

Because DADS consists of two fairly distinct systems that are referenced throughout this plan, and is a data delivery system, it is not within the scope of the project to determine content of any particular product or data set. The major program areas itemized in section 1.2 above will each determine product content and output format requirements based on their legal and programmatic requirements. Specific information about product file size, format, and other technical requirements are collected from the program areas through staff participation in the DADS development process.

#### **Public Inquiry (PI)**

Public Inquiry is the component of the DADS system that provides access to Bureau information to both internal and external users. The Public Inquiry system is used by all of the Bureau's external customers and 99% of the internal customer base and was the focus of the DADS97 Prototype 2 activities. The Public Inquiry system will provide an interactive interface for performing and displaying user initiated ad-hoc queries and will be accessible through the Bureau Intranet and the Internet.



The first phase of the process defines the DADS System. All work begins in parallel on requirements, architecture, planning, and prototyping (Figure 3). As the requirements are developed and accepted, they are placed under strict change control. Upon the completion of the Requirements Phase, the Integrated Product Team knows the detailed requirements, architecture, plan and deployment process.

### Fast Path for the Requirements Phase

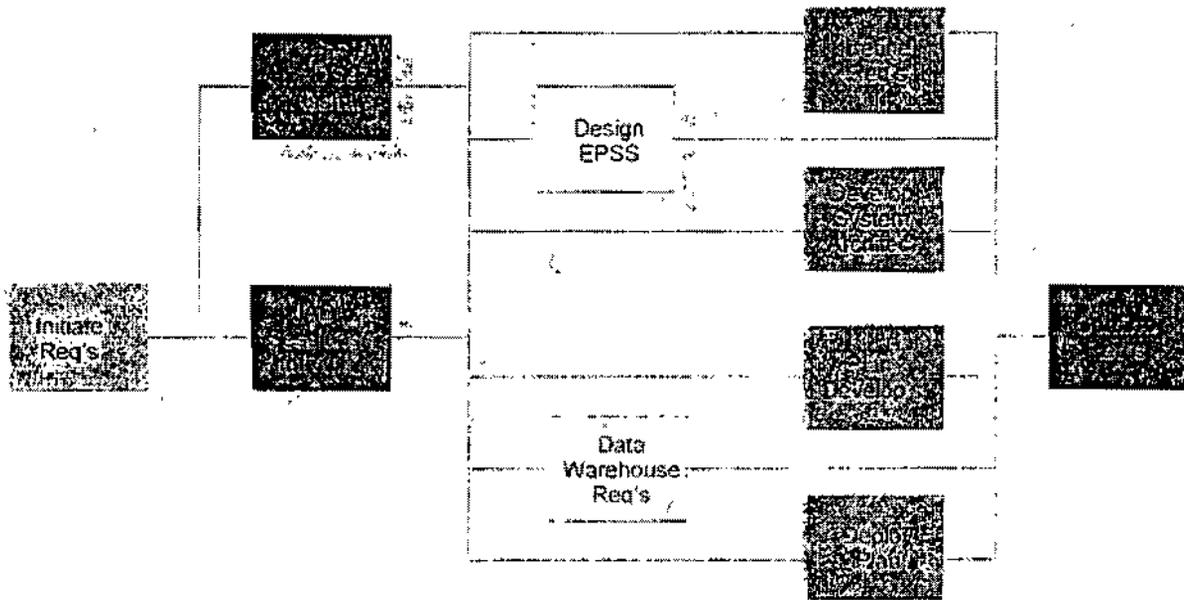


Figure 3



As each Build Cycle is completed, the result is the production of a working version of the system with the capabilities programmed into the defined process. Successive versions are designed to converge on the final system incorporating lessons learned, improvements, and culminating with the satisfaction of all requirements. Typically, each of these Build Cycles lasts ten weeks and includes a Design and Build Iteration activity intended for replication. At the end of each Build Cycle there is a working version of the planned software and a modified, detailed plan for the remainder of the project.

DADS will use the standard WSSDM Object Technology Deployment Phase (Figure 5). It will include the installation of both the software and hardware components as well as establishing and configuring the physical environment. During this period all required training will be conducted, all necessary testing and acceptance processes will occur, and in essence the job is finished.

### Fast Path Deployment Phase

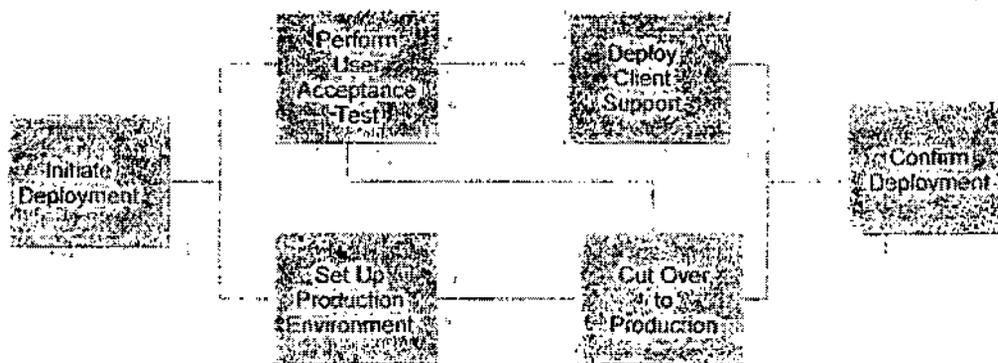


Figure 5

The DADS97 iteration was completed with the conclusion of user testing in February 1998. Initial feedback on results was provided to beta testers; final results will be posted on the DADS Internet Web site, slated to open in June 1998. The Requirements Gathering and Analysis phases for both Data Products Production and Public Inquiry have been completed as they apply to DADS98. The Build phase for Build Cycle 1/DADS98 is in its initial stages. The current system requirements are itemized below and a final Analysis Report will document any revisions to those requirements as a result of the Analysis phase of the current iteration.



## DADS Functional Requirements

The DADS98 Inquiry System is an interactive electronic system that provides access to Bureau data by internal and external users through the World Wide Web and the Bureau Intranet.

Functions That Are in Scope for DADS 98	
Browse and search for data and metadata	Map-based browsing
Retrieve data and metadata	Thematic mapping
Two levels of user interface (novice and expert) that provide assistance through help and tutorials	Extract data from summary files (internal users can also extract from detail files)
Economic profiles (geographic, industry, etc.)	Centralized maintenance of metadata by DADS
Predefined formats for Decennial Census tables that can be generated from Decennial summary files for user-selected standard Decennial Census geography such as states, counties, and places	User-defined tabulations from Decennial summary files and hundred percent and sample edited detail files appropriate to the data set being used
Perform inquiries and display, print, or download the results	Calculations of selected basic derived measures such as means, medians, percentages appropriate to the data set being used
Aggregating categories of variables to higher levels	

Functions That Are Not in Scope for DADS 98, Under Consideration for DADS 2000	
Fee-based transactions	Comparing across data sets and time frames
More than two levels of user interface that provide assistance	Advanced mapping functions such as Dot Density, Graduated Symbol mapping
Aggregating categories of variables to lower levels of aggregation or custom levels	Advanced user-defined calculations such as regressions, ANOVA, etc.
Saving queries from session-to-session	Tools for decentralized maintenance of metadata by Bureau of the Census divisions
Batch-mode query execution	Charts and graphs

The DADS98 Data Product Production (DPP) system will provide capabilities to produce PL94-171, Hundred Percent Summary File (HSF) and Sample Summary File (SSF) data products for the Decennial Dress Rehearsal.



**1. DADS will provide access to Bureau data and allow creation of customized products.**

DADS will provide a simple intuitive on-line ad-hoc query builder for data extraction and tabulation. Different navigational options will be available to allow users to limit their searches to specific geography, subject areas, keywords, and datasets. Metadata will be provided for performing navigations and for building queries. Users will have the option to create metadata files associated with their queries for viewing and downloading.

Different display options will be available for on-line query results, such as graphical, thematic mapping, and HTML displays. Also, different output options will be available, such as SAS data set, ASCII delimited, and HTML. Other functionalities include notifying the user of the estimated time it will take to perform the query, reporting the maximum length of the output, allowing the user to limit the number of records returned, and to title the query output.

**2. DADS will provide access to Bureau pre-defined products, including statistical briefs and abstracts, area profiles, economic indicators, press releases, summary data, geographic files, maps, and more.**

Users will have the capability to view and download pre-defined products. Navigation options will include searching for pre-defined products by subject areas, geography, keywords, product type, and product name. Metadata will be included that provides general product information and the associated subject areas, geography, keywords, and synonyms.

**3. DADS will identify and maintain information about users accessing the DADS system.**

DADS will identify users who connect to DADS, identify whether they are an external or internal user and whether they have access rights to Title 13 data, and will keep an audit trail on the users. The audit trail will log information about the user's connection, length of time connected, and the activities of the user (e.g., downloading, searches, ad-hoc queries). DADS will eventually add the capability of changing a user's system priority (higher or lower).

**4. DADS will announce enhancements and changes made to the DADS system.**

Some of the things that will be announced include: user interface and functionality changes; additions, deletions, or changes made to pre-defined products; and updates made to existing data sets, their availability, and the introduction of new data sets.



### **Requirements Analysis Teams**

This stage of the development process calls for seven teams to collaborate on requirements for specific functional areas of DADS98 and DADS2000. Functional requirements from each area are then organized into the full system level requirements and fully documented. The team subject areas are Tabulation/Publication, Geography, User Interface, Search/Query/Output, Predefined Products, Common Functions, Technical Architecture Requirements and Metadata. Team membership is cross-Directorate and cross-functional to ensure full representation of managerial, programming and subject matter staff. Test results from the previous iteration provide important input to this process.

### **Design Review**

The Requirements Analysis Teams and the DADS Working Group will participate in an early design review to provide early feedback to the technical teams and allow for implementation of corrective measures. The functions of an External Advisory Group have either been folded into the DADS Working Group or are accomplished through separate invitation.

### **User Testing**

A comprehensive user testing process is implemented once each prototype is delivered. The Test Plans that addressed DADS Prototype 1 and DADS98 were previously executed and fully document the methodology. The testing process incorporated the participation of a full cross-section of the customer groupings and is scaled to the capacity of the technical architecture.

### **Independent Verification and Validation (IV & V)**

The DADS program is committed to an independent verification and validation testing process to support program deliverable acceptance. It is envisioned that an independent third party organization will be contracted to perform an evaluation of the delivered system; e.g., DADS98 and DADS2000, to corroborate IBM's developmental findings and the joint government/IBM testing process. The results of this independent review will drive the government's acceptance or non-acceptance of the delivered system.



## 2.0 DADS IT Support

The current DADS architecture is based on a three-tier computing model. Each tier performs a specialized function and is connected to the other tiers via either the Internet or Census Intranet. The tiers are:

- Client platforms;
- Application Server; and
- Database Server and Map Server.

There are three architecturally similar environments maintained. The Census Intranet environment supports internal Bureau users, referred to as the "Internal Production Environment" as well as the DADS development systems, referred to as the "Development Environment." Intranet systems are protected from external access by the Bureau's Internet firewall (Figure 7). The Internet environment supports external users outside of the Bureau and is referred to as the "External Production Environment." The external systems physically reside outside of the Census Internet firewall and do not allow access to internal Bureau systems.

### Bureau Internet Environment

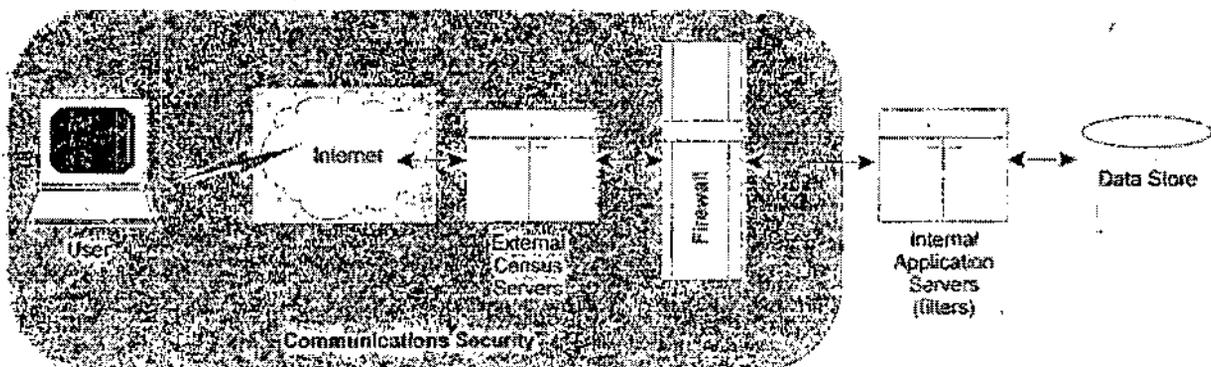


Figure 7



### Tier 3: Database Server and Map Server

Requests for access to the DADS datasets will be forwarded from the application server to the database server via high speed network (Fast Ethernet/100 Mbps). The database server will be exclusively dedicated to the generation of result sets and transmission of the results to the clients.

An additional component of the current DADS architecture is a mapping server. This system will provide processing dedicated to map-based geography selection for queries as well as thematic mapping. It is essential that we use a separate server for the mapping component so that we can tune each server to optimize performance for its specific application. The mapping platform will reside in the 3rd tier of the architecture (at the same level as the database server). Requests for mapping service will be initiated through the application server.

#### 2.1.2 DADS 1998 Prototype Architecture (Decentralized)

Figure 8 depicts a high level view of the DADS system architecture with its primary components. Like other processes in system development, the Architecture process is iterative. We go through cycles of increasing detail, narrowing in on the ultimate choices. We expect three iterations:

- coarse-grained architecture outlining the general framework without identifying specific components;
- refined high-level architecture indicating the major components (with justification) and understanding of the requirements sufficient to validate the stage of development; and
- final detailed architecture sufficient to allow software development, with component purchases underway (with justification).



## Application Development Environment

A number of software development tools will be utilized to facilitate the DADS application development process and to support IBM's WSDDM Object Technology Fast Path methodology activities during the development phase of DADS:

- Oracle Designer/2000 tool set including the Process Modeler, the Entity Relationship Diagrammer, the Function Hierarchy Diagrammer, the Designer/2000 Repository matrix tools, the Data Diagrammer and other Oracle PLSQL development tools such as Oracle Procedure Builder and Oracle SQL\*Plus;
- IBM's TeamConnection for configuration management and bug tracking;
- IBM's VisualAge for Java software development;
- IBM's VisualAge for C++ software development;
- Standard UNIX tools for C development;
- Rational Rose for Object Modeling;
- Segue SilkTest for Function/Regression Testing, Segue Surf! for Operability/Baseline Testing and Segue SilkPerformer Performance Testing; and
- PC Docs for Documentation Management.



### **Coding Standards and Review**

DADS has budgeted time for the development of project-specific coding standards. Through use of these standards, DADS will achieve high quality and easily maintained code. The standards will be few and powerful, rather than exhaustively complete. Peer code review ensures that no code goes into production without having been inspected by at least one other qualified developer.

### **Compliance with Standards**

DADS complies with a number of formally recognized and defacto standards. Access to the DADS data warehouse will be performed via Oracle's ANSI/ISO SQL-92 compliant implementation of the SQL query language. The Java programming language, used for applet development, will be submitted to an international standards body as recently announced by Sun Microsystems. The web server software will be HTTP 1.0 or 1.1 compliant. HTML pages will be developed based on the HTML 3.2 standard.

### **Public Law 94-171 Data Products**

The Bureau is required to produce and distribute the Public Law data products one year after Census Day. The same timetable applies to Dress Rehearsal dissemination in DADS98.

### **Confidentiality**

The DADS project will rely on the Bureau's Disclosure Review Board (DRB) to define and approve statistical methodology to protect data products from confidentiality risks. DADS internal security protection will be defined by the Class C2 Controlled Access Protection procedures and requirements defined in the Department of Defense Standard, DoD 5200.28-STD document.

### **General Information Technology Security**

DADS will apply general Information Technology security standards defined in the Bureau's document titled, "Handbook for Information Technology Security".



### 2.1.3 DADS 1998 Prototype Progress Against Planned Milestones

Current Fiscal Year - FY98 DADS 1998 Prototype/72-6811 Executive Milestone Overview		
Description	Planned Start	Planned Completion
Exercise contract option year for FY98	10/97	9/98
Issue Task Order 98-001; Requirements Development	10/97	completed
Issue Task Order 98-002; SP System Acquisition	10/97	completed
Requirements and High-level Architecture Planning	10/97	completed
Prepare for User Testing	10/97	completed
Complete Transition Planning	10/97	completed
Conduct Internal User Testing	11/97	completed
Conduct External User Testing	12/97	completed
SP/2 System Installation	12/97	completed
Issue Task Order 98-003; System Design and Build	1/98	9/98
Compile Results of Testing	1/98	completed
Develop IT Plan PDP	2/98	completed
Requirements Analysis and Documentation	11/97	completed
DADS98 Initial Requirements Estimate	11/97	completed
Design and Documentation	1/98	6/98
Architecture Review	2/98	completed
Usability Study 3.1	2/98	completed
Inquiry Requirements Document Complete	3/98	completed
DPP Requirements Document Complete	3/98	completed
Detailed Design Review	3/98	completed
Usability Study 3.2	3/98	completed
Build Cycle 1 (Release 1)	3/98	completed
Build Cycle 2 (Release 2)	6/98	8/98
Build Cycle 3 (Release 3)	8/98	9/98

All of the milestones included in the 1998 Operational IT Plan were accomplished on schedule to date. There is no anticipated slippage with the others.



DADS Final Production System for 2000 High Level Milestone Schedule				
FY 99	FY 00	FY 01	FY 02	FY 03
Issue Task Order 99-001; Build and Enhance				
Final Build/Deliver DADS98 Inq/DPP				
Prepare for User Testing of 1998 Prototype				
Conduct Internal User Testing for 1998 Prototype				
Conduct External User Testing for 1998 Prototype				
Compile Results of Testing of 1998 Prototype				
Develop Requirements Analysis and Documentation				
Develop Design and Documentation				
Conduct Usability Study 4.1				
Conduct Usability Study 4.2				
Develop Software Build 4.1	Develop Software Build 4.1			
	Conduct Usability Study 4.3			
	Develop Software Build 4.2			
	Develop Database Build - Production Database			



will report this at monthly Program Management Reviews. At the end of the Build, the DADS team will perform a compliance check, reviewing each function. Aspects to measure include:

- compliance with requirements;
- usefulness: the degree to which the product allows users to achieve their goals;
- effectiveness (ease of use): a quantitative measure determined by speed of performance and error rate;
- learnability: predefined level of competence after some predetermined amount and period of training; and
- attitude (likeability): user's perceptions, feelings and opinions of a product.

Each function will be scored (0% to 100%) by the compliance review team. A decision will be made by the DADS team as to whether functions with less than 100% compliance can be accepted. If so, these functions will be included in the compliance list for the next Build(s), and the IBM Build process will be required to indicate how full compliance can be met in the subsequent Builds.

### System Performance

The DADS team will define metrics based on projected system performance requirements. Performance projections are based on the DADS Performance Models (for Data Products Production and Inquiry). Performance areas that will be measured and tracked include:

- Capacity: Required storage and system memory capacity requirements.
  - DADS98: 1 Terabyte
  - DADS2000: 1 Terabyte
- Throughput: Required bandwidth and processing speeds necessary to meet DADS requirements.
- Response Time: Required turnaround time for the DADS subsystems to respond to a user's input.
  - DADS98: less than one minute measured from the time a user executes a query from an input device to the point of the resultant data output.
  - DADS2000: same as above.

The DADS team will record final projected measures for these three areas. Anticipated performance curves will be developed for the three measures, based on the subsystem Build cycles. Progress can then be tracked in comparison to the curves, with the DADS engineering process responding to any lags in attainment of intermediate goals.

### Business Performance Metrics

There are two areas that require measurement against business goals: performance against DADS Program Goals and Objectives, and IBM contract performance.

## Alternative 2

The second alternative considered was to "DEVELOP DATA ACCESS AND DISSEMINATION SYSTEM (DADS)." Alternative 2 involved the development and use of new, primarily electronic, methods for data access and dissemination for Census 2000. This would entail building upon the foundation laid in the successful implementation of DADS Prototype 1 and several follow-on prototypes, up through and including the production system.

It was assumed that the proposed hardware/software platform required to support DADS would be in place and functional as of the dates established for the final production system. Upon completion of the test and evaluation of Prototype 2, configurations would be re-evaluated as to type, numbers and sizes and adjusted as necessary. The LAN infrastructure would continue to be the 100Mbps Fast Ethernet system currently in place.

## Conclusion

The DADS alternative (Alternative 2) had the potential to present the Bureau with considerable cost savings during the conduct of the data access and dissemination function for the 2000 Census data. The number of Full Time Equivalents required under the DADS alternative would be much less than those required under Alternative 1 (Repeat 1990 Methodology) and should result in substantial cost savings to the Bureau. A sensitivity analysis was performed on the projected number of required Full Time Equivalents that demonstrated that the estimate done by the Bureau managers would have to be increased by 70% before Alternative 2 would be equal in cost to Alternative 1.

The risks inherent in either alternative were relatively moderate, and relatively equal to each other. Also, the Bureau is taking steps to mitigate the risks inherent in the DADS Alternative.



### **Risk: Data Products Will not be Produced on Time**

Production of data products is the primary success criterion of the DADS Program, in particular production of Public Law products. The development schedule for the Data Product Production (DPP) subsystem of DADS is extremely aggressive, which leaves little room for error. The development of DPP also relies heavily on support from other Bureau areas, for assistance with product specifications and for providing timely information on required production schedules. The DPP subsystem development, in order to succeed, requires a strong team leader on the DADS DPP development team, and cooperation from other program areas. Mitigation processes include selection of a proven, commercial off-the-shelf product as the "heart" of the DPP subsystem; selection and installation of a strong DPP leader; tight, aggressive scheduling with room to "beta test" the subsystem, and close coordination with other Bureau program offices to receive information early.

### **Risk: Data Warehousing Capability will be Late**

The DADS system relies heavily on Data Warehousing technology to access and make use of the data from many different Bureau programs. It is essential that the Data Warehousing capability be powerful, comprehensive, fast and capacious; above all else, it must be highly integrated with DADS functions. The program faces several challenges: development time is limited; the amount of data to be warehoused is huge and diverse; and it is anticipated that the time needed to "build" the requisite databases will be very large, which will limit the ability to rebuild in the event of problems. To mitigate this risk, the DADS Program is working closely with IBM on teaming and leadership of the Data Warehouse development team; in addition, the development process is being devised to provide as much time as possible for loading and testing the data warehousing capabilities.

### **Risk: DADS User Help System will be Inadequate**

The DADS user help capability, called Electronic Performance Support Solution (EPSS), is envisioned as a very sophisticated, context-sensitive user support function. For EPSS to function properly, the DADS development team requires metadata definitions and specifications from Bureau organizations, as well as assistance from Bureau user organizations in reviewing and commenting on the DADS team's proposed user interfaces. Any tardiness in either of these areas will adversely affect the development of the capability, and could render the entire system unsatisfactory. To mitigate the risk, the DADS Program is coordinating with points of contact in each of the participating Bureau organizations, both to coordinate interface reviews as well as to solicit metadata.



## 2.2 DADS Infrastructure Description

### 2.2.1 Detailed Description of System(s)/Subsystem(s)

The current DADS infrastructure consists of 60 Pentium based desktop computer systems and peripherals. Of these, approximately 45 are used for software development as well as for office automation (OA) activities. The remaining 15 are used primarily for administration and support activities using OA tools. There are four HP Laser Printers shared via the LAN. See Figure 11.

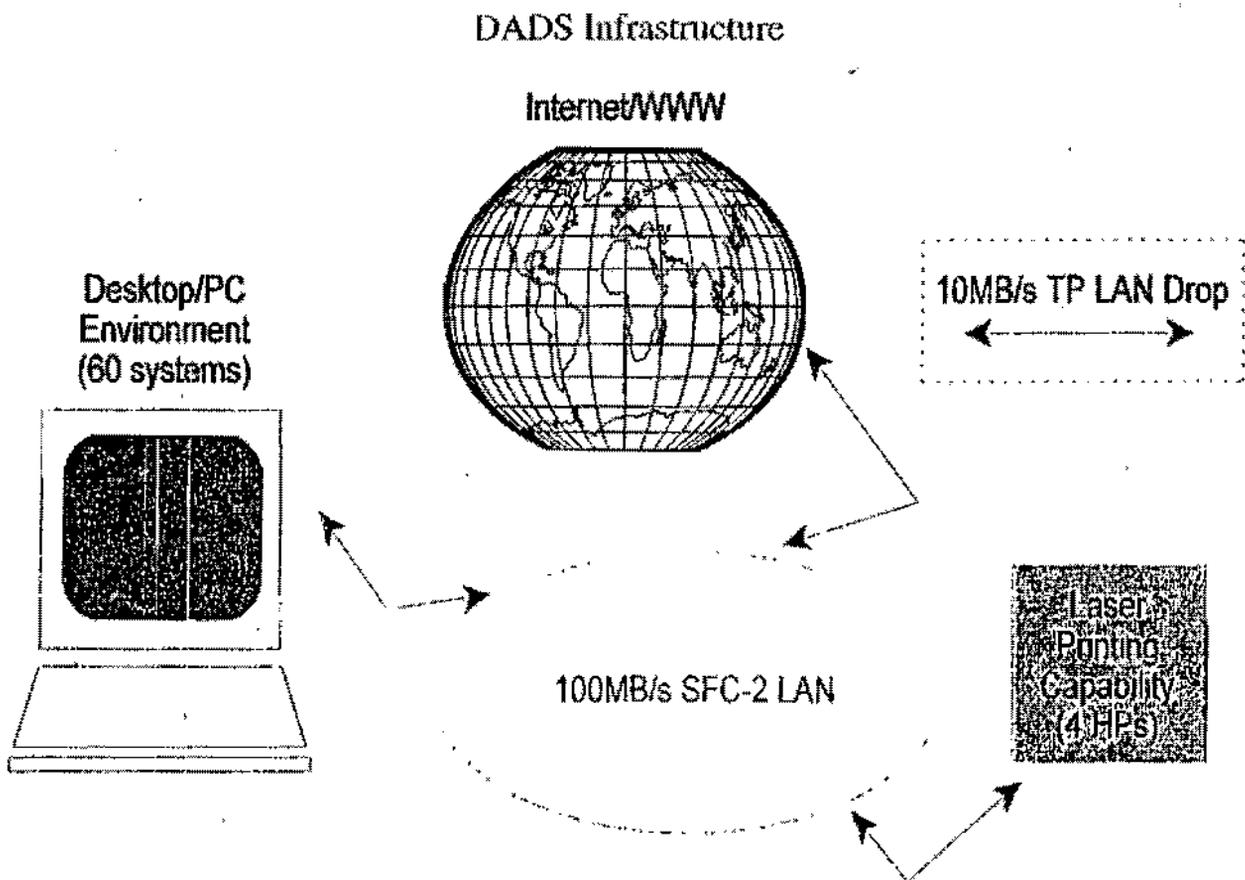


Figure 11

### 2.2.2 DADS Infrastructure Architecture (Decentralized)

All of DADS infrastructure components are supported by Bureau corporate facilities. We have no decentralized support infrastructure within DADS.



### 2.2.8 DADS Infrastructure References

The DADS Infrastructure is supported by the following planning documents:

- Strategic IT Plan, FY 1998-2002, dated December 31, 1997, pages 17, 18, 49, 60, and 80;
- Budget Submission for 1999, June 1997, page CEN-115;
- Budget Submission for 2000, June 1998;
- DADS Program Management Plan (under revision, originally dated November 1997); and
- DADS Comprehensive Development Plan (under revision, dated February 1998).

### 2.2.9 DADS Infrastructure Risks

Not applicable.

### 2.2.10 DADS Infrastructure Security

Security planning at the Bureau is described in the Introduction to this Operational IT Plan.

# Data Access and Dissemination System (DADS)

## 1.0 DADS Overview

The Data Access and Dissemination System (DADS) is comprised of two primary subsystems: American FactFinder (AFF) and Data Products Production (DPP). These subsystems jointly comprise a suite of applications designed to provide responsive, multi-tiered, near-universal access to the U.S. Census Bureau's vast storehouse of data through a state-of-the-art, Internet-based, user-interactive interface implemented via current World Wide Web technology.

We describe the IT resources for DADS as the component sub-systems and the office automation infrastructure providing program support. The IT goal of DADS is to be "the new way for the U.S. Census Bureau to expand the accessibility and delivery of its information to users, internal and external, for all skill levels, novice and expert alike."

In pursuit of that IT goal, the DADS program intends to realize the greatest gains and successes in the following areas:

- **Consistent data access and dissemination:** DADS will identify and, as appropriate, use all relevant U.S. Census Bureau work underway or planned so that a coordinated corporate approach to data access and dissemination is universally used;
  - **Increased responsiveness:** By providing dynamic access to census data, DADS will allow the U.S. Census Bureau to more rapidly respond to requests for products and information;
- (continued)

- **Reduced costs:** DADS will adopt the most efficient and innovative processes to reduce the time and effort required to make products and information available. DADS will use commercial off-the-shelf software to support the DPP;
- **Electronic commerce:** By allowing products to be ordered and paid for over the Internet, DADS will allow the U.S. Census Bureau to become more self-sufficient, expand its customer base, and maximize reimbursable income;
- **Planned changes:** The U.S. Census Bureau will enlist the assistance of internal and external users to test the viability of each DADS iteration and will accept, reject, and/or refine independently each iteration. By doing so, DADS will allow the U.S. Census Bureau to adjust to the rapidly changing spectrum of requirements and take advantage of the varied and evolving technology environments;
- **Customer relations/User satisfaction:** The U.S. Census Bureau will continue to include customers in the design and testing of DADS. By listening to our customers, we will ensure that evolving user requirements are met or exceeded; and
- **Adaptability:** DADS will respond quickly to changing customer requirements and meet them as they are identified.

The U.S. Census Bureau is in the preliminary fact-finding stages of an Integrated Information Solution (IS) project, of which DADS will be a part. The project defines the details of how the U.S. Census Bureau should proceed from DADS development to the full vision of an integrated information access and dissemination system for internal and external users. We envision that DADS will play a major part in the U.S. Census Bureau's new world of "Electronic Commerce."

The American FactFinder (AFF) Subsystem is accessible through the U.S. Census Bureau's Intranet and the World Wide Web via a standard browser. It allows users to select census products or submit queries against varied datasets to extract meaningful information and produce results, which can be in table or map form. Some queries may be generated based upon maps and most results can be reformatted, including formatting from tables-to-maps or vice versa.

The Data Product Production (DPP) Subsystem will provide an interactive interface for designing, reviewing and generating 1998 Dress Rehearsal and 2000 Decennial data products, including Public Law 94-171 (PL) listings and summary files, the Sample Summary File (SSF), and the Hundred Percent Summary File (HSF). In addition, the DPP subsystem will, based on these data products, generate a number of file output formats for printing, for stamping to CD-ROM, and for on-line Intranet/Internet access through the AFF.

### 1.1 DADS Products, Services, and Customers

As mentioned, The American FactFinder (AFF) is a corporate, enterprise-wide subsystem designed to provide access to, creation of, and delivery of census data products and data sets. AFF allows customers to select census products (existing on-line documents) or query against data sets (collections of data that can be correlated/tabulated). Executed queries generate results that can be in table or map form and can also, for the most part, be reformatted, including reformatting from tables-to-maps or vice versa. Internal government staffs may eventually use AFF in response to expressed customer needs to produce new products by assembling sets of queries and appropriately formatting the results. Once approved, these products will become standard U.S. Census Bureau

products and will then be made available to other customers through AFF.

Analysis of current trends indicate that future customers will need to be able to generate queries across data in new and unique ways, correlating information not envisioned by a single survey but as part of diverse surveys. In response to this need, AFF will incorporate the capability for customers to associate and integrate data from different datasets. Eventually, we envision that non-census datasets will become accessible through AFF, even though they are not hosted by AFF.

The table on the following page summarizes information on AFF data products.

American FactFinder (AFF) Data Products		
Class	Capability	Description
Tier 1: static products	search, browse, retrieve, view, print, and download in a what-you-see-is-what-you-get environment	<ul style="list-style-type: none"> <li>• press releases (static);</li> <li>• statistical abstracts (static);</li> <li>• census briefs; and</li> <li>• information bulletins</li> </ul>
Tier 2: summary data products	select, extract, and manipulate by geography from summary data files	<ul style="list-style-type: none"> <li>• 1990 Decennial Census Summary Files;</li> <li>• American Community Survey Summary Tables;</li> <li>• 1998 Dress Rehearsal for Census 2000 Summary Files; and</li> <li>• 1997 Economic Survey Summary Files</li> </ul>
Tier 3: microdata products	create custom tables from microdata files within strict confidentiality standards	<ul style="list-style-type: none"> <li>• 1990 Decennial Census;</li> <li>• 2000 Decennial Census;</li> <li>• American Community Survey</li> </ul>

As mentioned, the Data Products Production (DPP) subsystem will provide an interactive interface for appropriate U.S. Census Bureau staffs to design, review and generate pre-defined summary data products using the Hundred Percent Edited Detail File and the Sample Edited Detail File microdata datasets. In addition, the DPP subsystem will, based on these data products, generate a number of file out-put formats for printing, for stamping to CD-ROM, and for on-line Intranet/ Internet access through the AFF.

You can find a mid-term, futuristic vision that represents the potential end result for

DADS by considering the *Statistical Abstract of the United States*, published each year by the U.S. Census Bureau. The DPP will reach its functional maturity when the U.S. Census Bureau can produce the *Abstract* by pressing a few buttons; the AFF will reach its functional maturity when internal or external customers can access, tabulate, or download anything from the *Abstract* via the Internet. When both concepts are mature and available, DADS will have realized its full potential.

The table on the following page summarizes information on DPP data products.

Data Product Production (DPP) Data Products	
Class	Description
1998 Dress Rehearsal Public Law 94-171 (PL)	<ul style="list-style-type: none"> <li>• PL listing for printing;</li> <li>• ASCII file of metadata/data for the CD-ROM product and for FTP download from the Internet; and</li> <li>• Internal format for upload into the DADS98 Inquiry system for access to Tier 2 products</li> </ul>
Hundred Percent Summary File (HSF)	<ul style="list-style-type: none"> <li>• ASCII file of metadata/data for the CD-ROM product; and</li> <li>• Internal format for upload into the DADS98 Inquiry system for access to Tier 2 products</li> </ul>
Sample Summary File (SSF)	<ul style="list-style-type: none"> <li>• ASCII file of metadata/data for the CD-ROM product; and</li> <li>• Internal format for upload into the DADS98 Inquiry system for access to Tier 2 products</li> </ul>

The following table lists some of our clients and the data products they disseminate through DADS.

Data Access and Dissemination System Clients: U.S. Census Bureau		
Name	Census or Survey Name	Census/Survey Description
Decennial Directorate	<ul style="list-style-type: none"> <li>• 1990 Decennial Census;</li> <li>• 1998 Dress Rehearsal for Census 2000; and</li> <li>• 2000 Decennial Census</li> </ul>	Census of population and housing taken every 10 years to determine how many Congressional representatives each state will have
Economic Directorate	<ul style="list-style-type: none"> <li>• Economic Census</li> </ul>	Conducted in years ending in "2" and "7," the Economic Censuses compile information about businesses, industries, establishments, products, and local areas
Demographic Directorate	<ul style="list-style-type: none"> <li>• American Community Survey</li> <li>• Continuous Measurement</li> </ul>	A new survey being developed by the U.S. Census Bureau to provide, every year, economic, social, and housing information about communities and population groups

DADS customers (or data users) are from the public and private sectors; they are internal and external to the U.S. Census Bureau and they use census data to govern and make economic policy and business decisions. DADS will be built to satisfy a wide range of users from novice to expert, who will use the system in different ways for various purposes. Figure 1, on the following page, depicts DADS' internal and external customers.

The U.S. Census Bureau's internal data customers include staff that design and create summarized data, one-time special tabulations and extracts from a given micro-data set to place it within the pre-defined product side of DADS. These staff members play a critical role in achieving easier access to the U.S. Census Bureau's data and data products and comprise an important segment of the customer market for census data. U.S. Census Bureau subject matter analysts also prepare specifications in response to inquiries from Congress, other federal agencies, state and local governments, and other groups. Improved access to data for these customers results in increased efficiency in designing and creating standard products and quicker response to requests for special tabulations. This group comes

from all major areas of the U.S. Census Bureau.

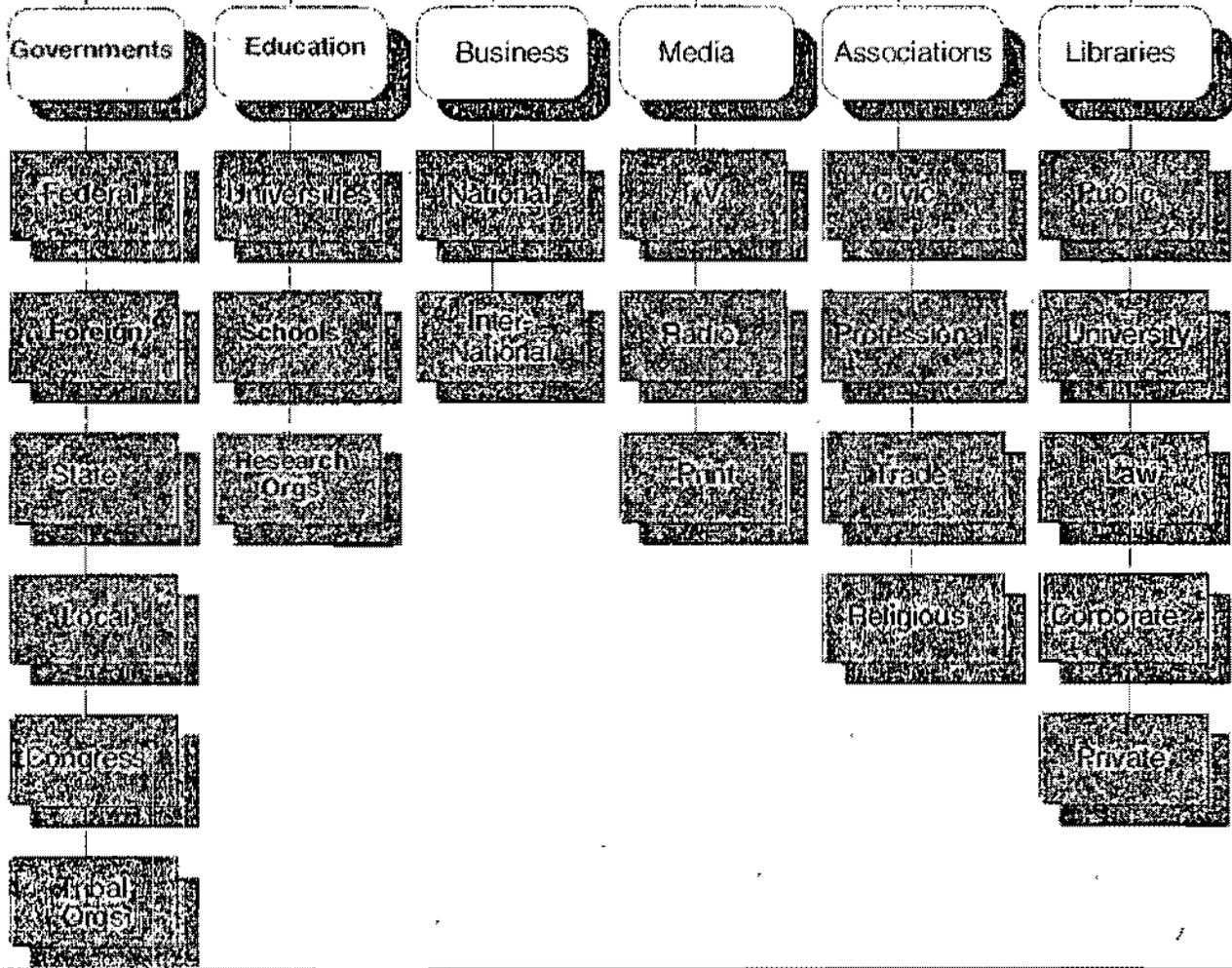
Additionally, the U.S. Census Bureau supports internal and external data disseminators who assist and provide information to data users. Internally, this group includes Regional Office staff and other individual divisions and program areas that respond to inquiries for data and data products. Externally, the group includes the State Data Center Network and other distribution networks such as the Federal Depository Library System. They essentially use DADS as a look-up system, using mostly summarized data, and providing help and direction to data users.

Because DADS is a data delivery system, it is not within the scope of the project to determine content of any particular product or data set. The clients noted in the previous table (page 4) and the customers listed in Figure 1 (page 6) will be the driving forces behind determining product content and output format requirements. Resident staff participating in the DADS development process collect specific information about product file size, format, and other technical requirements.

Internal U.S. Census Bureau Organizations and Staffs

Internal Customers  
External Customers

Internal Customers  
External Customers



Data Access and Dissemination System  
Customers

### 1.2 DADS IT Objectives

The DADS program intends to make maximum and efficient use of its IT resources to meet programmatic objectives by:

- reducing software development costs through the rigorous and efficient application of management and contract techniques;
- ensuring that DADS is hardware and operating system independent;
- ensuring that DADS is Y2K compliant;
- ensuring that there is a continuous and comprehensive transfer of technology to the government throughout the development process;
- providing a utilitarian system of enhanced data access, dissemination, and tabulation tools to U.S. Census Bureau analysts;
- ensuring that the critical components of DADS are fault-tolerant with adequate hardware and telecommunication redundancy;
- fostering innovation and judiciously applying technology to ensure that it responds to customer and programmatic requirements and is cost beneficial; and
- ensuring that existing and planned technology meshes with existing and future staff competencies.

We will achieve DADS' IT objectives by:

- maintaining a stable, secure, and responsive DADS presence on both the Internet and the U.S. Census Bureau Intranet;
- thoroughly scrubbing the most costly and resource-intensive processes to ensure maximum return on investment;
- ensuring that the products of our incremental development processes are re-useable and serve as the base components of follow-on efforts (i.e., no disposable products);
- listening and responding to our customers;
- using commercial off-the-shelf software wherever possible; and
- retaining and strengthening core staff competencies and contracting out for specialized skills.

## 2.0 DADS IT Support

DADS is designed to support many U.S. Census Bureau staffs and organizations, and as such is a resident member of the corporate network infrastructure, relying upon support from the Telecommunications Office (TCO).

Figure 2, below, depicts how the American Factfinder subsystem supports information technology.

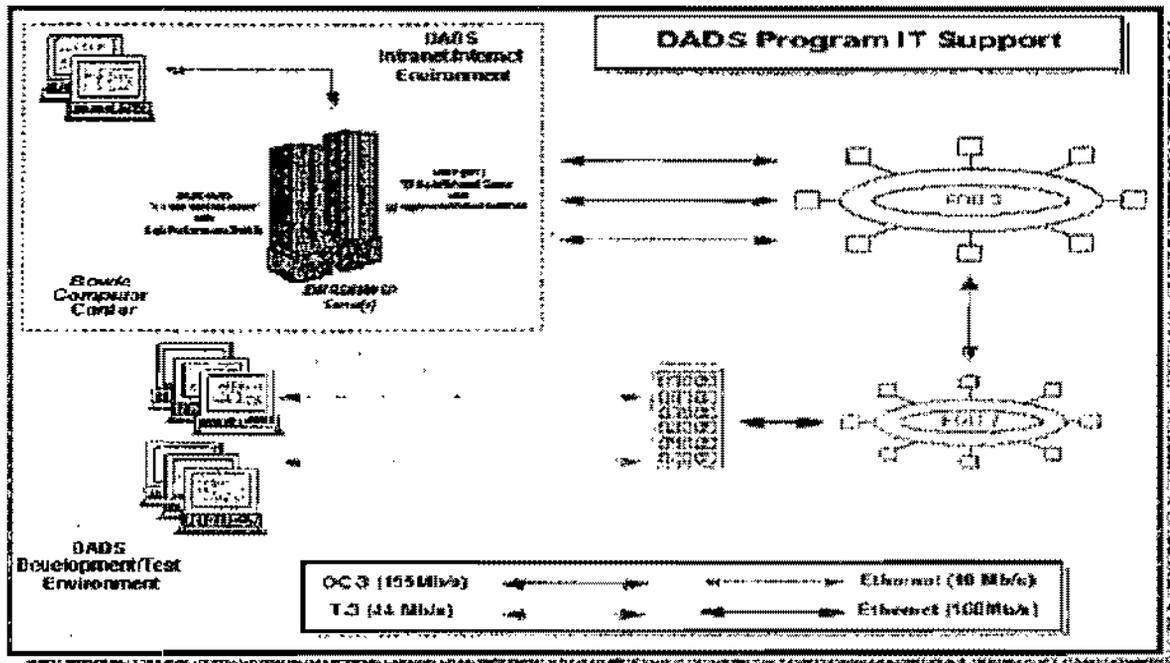


Figure 2: American FactFinder IT Support

The Intranet environment supports internal U.S. Census Bureau users as well as the DADS development systems. Intranet systems are protected from external access by the U.S. Census Bureau's Internet firewall (see Figure 3, below). The Internet environment, which physically resides outside the U.S. Census Bureau Internet firewall, supports users external to the U.S. Census Bureau and does not permit unauthorized external access to internal Bureau systems.

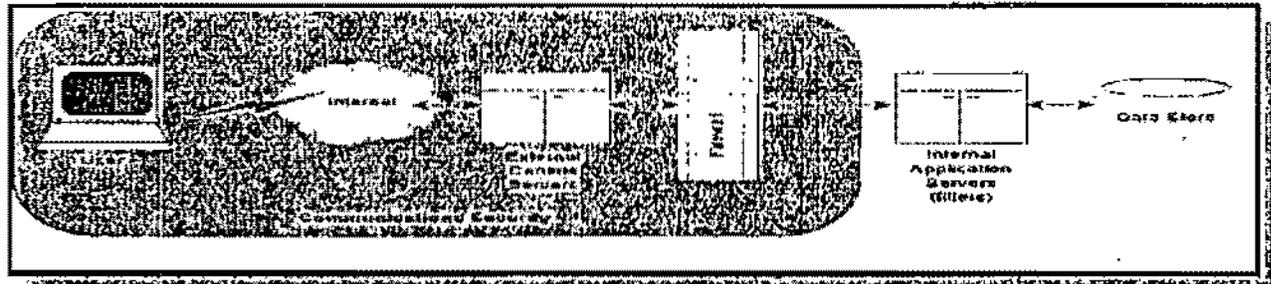


Figure 3: DADS Communications Security

The Decennial Directorate provides system management and network support for the Data Products Production (DPP) subsystem Windows NT servers. Necessary support requirements include operating system maintenance, disk backups, server monitoring, and configuration maintenance for the Windows NT Server and Novell IntraNetware operating and user environments.

In particular, the Decennial Directorate is responsible for monitoring servers to detect hardware, operating system, and network error conditions and for providing sufficient administrative coverage to ensure adequate

response to hardware, network, and security problems. Decennial systems management staff provides the ability to recover everything from the operating system to the file system to the SuperCROSS databases and source files from backup media. DPP subsystems are housed in the secure DADS terminal room in compliance with U.S. Census Bureau ADP/Title 13 security requirements.

Figure 4, on the following page, depicts how the Data Products Production subsystem supports information technology.

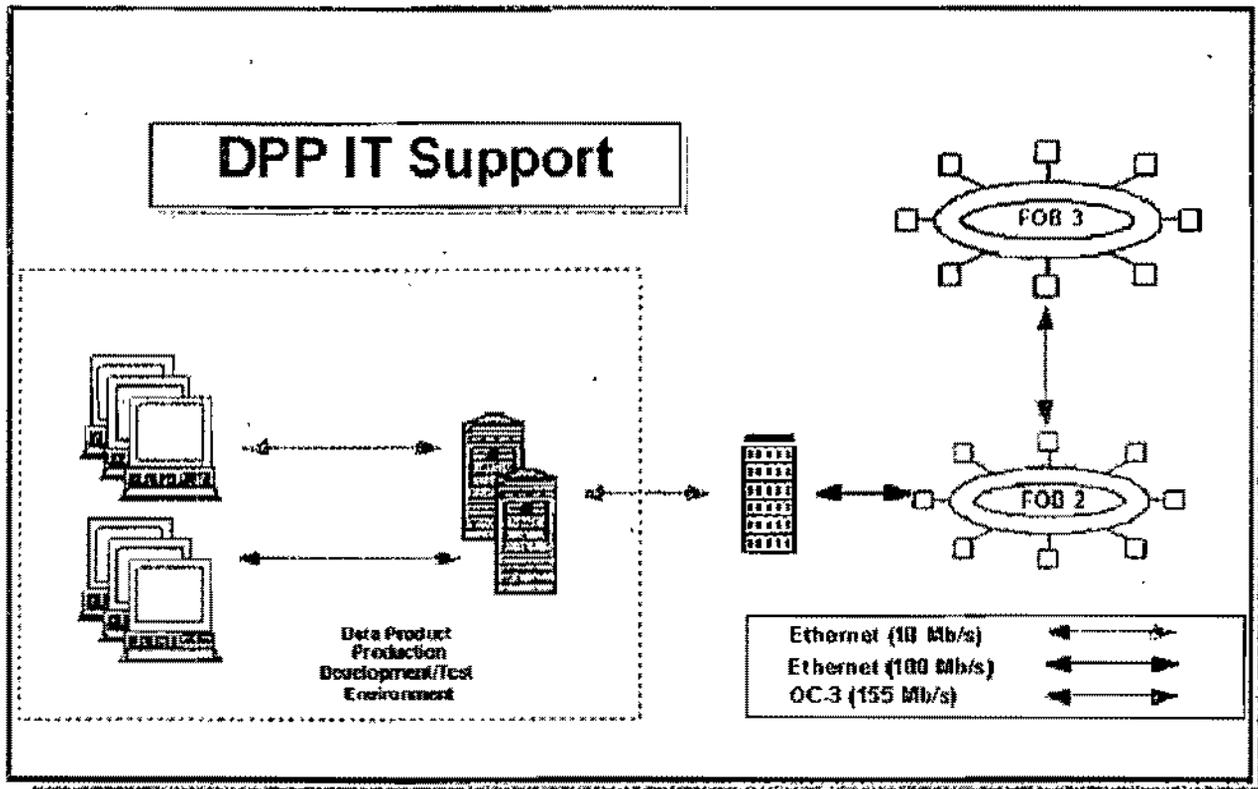


Figure 4: Data Products Production IT Support

DADS has a number of new, unfunded initiatives for the American FactFinder and Data Products Production subsystems. These are listed in the tables that follow.

American FactFinder New Initiatives	
Initiative	Description
Redundant Sites	State Data Centers, and other governmental and large-agency data organizations, require highly responsive, high-capacity support for their data operations. Redundant DADS sites would duplicate the most resource-intensive aspects of DADS and update the data at these sites on a real-time basis.
Downloading Files	Many high-level users and organizations are clamoring for the ability to access and transfer very large Census data files (very large Economic PDF files, for example) for off-line review and analysis on third-party applications. It is most efficient and cost-effective to establish a DADS FTP site, that can be managed as part of the overall DADS data management process, that maintains desired files and provides them for download.

American FactFinder New Initiatives	
Initiative	Description
Electronic Commerce	The U.S. Census Bureau has mandated that some types of data available through the external DADS AFF subsystem will have associated charges. In order to support such a requirement, the AFF subsystem will need to be able to monitor all queries originating from the user, identify those that have charges associated with them, collect all such charges, and conduct a purchase transaction.
Batch Processing	The DADS Team anticipates that users will produce very long and complex queries. These queries, if implemented in real time by the system, would consume enormous resources, reduce the overall responsiveness of DADS, and make less of the system available for other users. This requirement would provide a means for the system to process a query during non-peak conditions.
Home Page Integration	Census end users will get confused if there is a proliferation of styles, menu choices, icons, etc., when they access the U.S. Census Bureau web site. Consolidation of web designs around the current DADS user interface design would greatly reduce confusion and create an integrated, consistent look and feel for all of census data.
Enhancements for Economic Data	The Economic Directorate has extremely demanding requirements for dissemination of their data. It is essential that significant attention be paid to enhancing DADS' capabilities to completely display Economic data, otherwise vital information will be lost and studies based on that information will be compromised.
Performance Enhancements	Census data products have never before been made available via the Web in the manner envisioned for AFF. It is anticipated that, because of the innovative and user-friendly interface and the depth of information available, the American FactFinder will be very popular to in-depth users as well as casual browsers. Current estimates for usage are based on U.S. Census Bureau historical data, which are considered untrustworthy since this type of approach has never been used before; engineering estimates are that load on the AFF internal and external systems will nearly triple by 2001.
Data Integration	One of the U.S. Census Bureau's primary goals is to find means to intelligently and effectively integrate the data from the program areas. It is important to ensure consistency and accuracy, both for U.S. Census Bureau data and for the retrieval and dissemination systems for accessing the data.
Integration of Geography	This would allow DADS to provide comprehensible support to the widest user community with a consistent, integrated view of geography.
SuperCross Integration	AFF will be able to respond more rapidly, and at higher volume, by making use of SuperCross's sophisticated query capabilities. This would reduce or eliminate the requirement to produce and store pre-generated tables, significantly streamlining system resource usage and enabling AFF to accommodate more different types of data and data products.
ITS Initiatives	The ITS is focusing on long-term solutions to projected U.S. Census Bureau business needs, many of these solutions involve DADS. Early development creates a framework to support longer-range ITS initiatives, which will make it less expensive in the outyears to develop additional capabilities.
Standards Compliance	It was originally anticipated that compliance requirements for the Americans with Disabilities Act (ADA) would not become an issue until FY00 or beyond. However, several initiatives have been started on other U.S. Census Bureau programs, and it is possible that DADS will be included. Compliance with the ADA will be complex, in particular because many of the interface screens and user responses will need to be completely redesigned to accommodate special needs. An early start to address these considerations will greatly expand the potential user community, and create a positive atmosphere among our customers.
Developing DADS Requirements	In order to focus on DDP subsystem development in FY98/99, further DADS requirements analysis was postponed until FY00. These analyses must be completed as early as possible so that design/development work on DADS2000 can commence.

Data Products Production New Initiatives	
Initiative	Description
CUF/CEF Access	Selected U.S. Census Bureau internal organizations urgently require easy access to the census Unedited and Edited Files. The DPP subsystem, because of the use of SuperCross, would provide an ideal platform for that access.
Data Verification	Accuracy of the data portrayed in census data products must be unimpeachable. Without data verification capability, the U.S. Census Bureau's very small team of DPP operators will have to use labor-intensive off-line data verification techniques.
Data Product Progress/Tracking/Reporting	Data products are extremely complex documents, and the turnover associated with document creation is great. In order to prevent inadvertent compromise of data, and to ensure the most current information is provided to the user community, it is essential that data products be accurately tracked.
2000 Workloads	Because of the time press to produce a system, DADS98 DPP subsystem development very specifically focused on accommodating Dress Rehearsal capacities. We anticipate that the significantly higher workloads and capacities required for the Decennial will tax the DPP subsystem to its limits.
IIS Initiative	The IIS is focusing on long-term solutions to projected U.S. Census Bureau business needs, many of these solutions involve DADS. Early development creates a framework to support longer-range IIS initiatives, which will make it less expensive in the outyears to develop additional capabilities.

## 2.1 DADS IT System Description

### 2.1.1 Detailed Description of American FactFinder (AFF)

The IBM RS/6000 SP family of servers serves as host to the American FactFinder (AFF) subsystem of DADS and is located at the Bowie Computer Center. This family of servers provides the widest possible range of scalability for servers of its class (Unix and NT), allowing for scalability and extendibility into the future. Because of its clustered configuration, the SP isolates most software and hardware failures, providing an overall very high level of reliability. AFF software is open in nature and allows development and deployment on other, similar hardware. Internal and external servers are separate due to U.S. Census Bureau security and confidentiality protection concerns.

Each of the SP servers consists of several Symmetric Multi Processing (SMP) shared-memory nodes. In the initial configuration, each node consists of four high-performance PowerPC processors and 3GB of memory, along with I/O interfaces. Other nodes are available which have different attributes. All nodes are connected with a high-performance switch whose aggregate bandwidth is proportional to the number of connections between nodes. Software is installed on nodes such that no single node failure can take the system down.

The American FactFinder configuration is shown in Figure 5, on the following page.

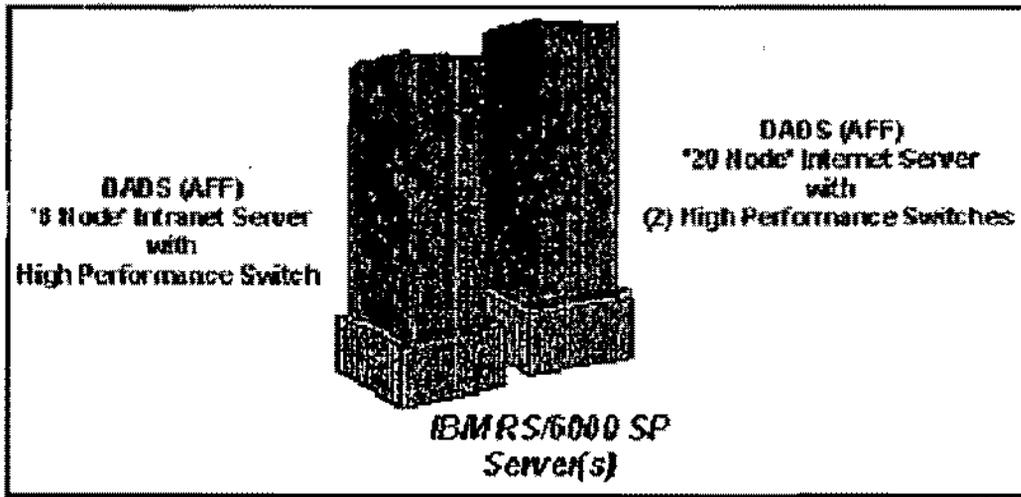


Figure 5: American FactFinder (AFF) Server Configuration

The AFF component of DADS is engineered for maximum availability and high uptime by the design implementation of failover capabilities. As a result, there are very few single points of failure. Certain components are judged to be so reliable (or not needed for normal operation) as to not need a backup. As an example, the SP switches do not have redundant backups. Possible downtime has been considered in the Reliability Model.

Disk capacity is configured such that every disk is connected to two nodes, and all data is mirrored; no single disk- or node-failure can make information inaccessible. This arrangement is depicted in Figure 6 below. Disks are also remotely accessible from other nodes, with very little latency. This process, in effect, creates one large and reliable shared disk.

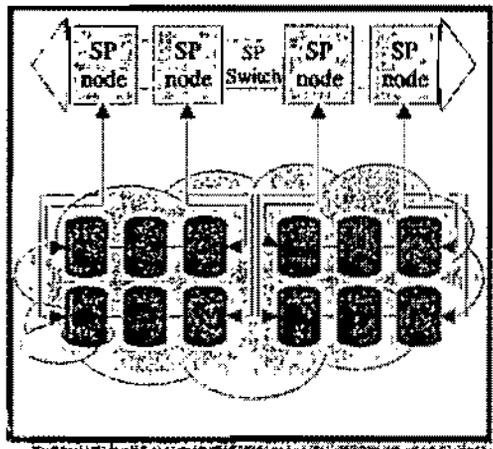


Figure 6: American FactFinder (AFF) Reliability

The following table lists the American FactFinder subsystem's major hardware components.

American FactFinder (AFF) Major Hardware Components							
Quantity	Vendor/Model	CPU	RAM	Storage	Components or Modules	Physical Location	Comments
2	IBM RS/6000 F40P Workstations	333 MHz	256 Mb	9 GB	HWACS	BCC	100MB/s NIC
2	IBM RS/6000 43P Workstations	333 MHz	128 Mb	18 GB	150 MB Switch	HQ	100MB/s NIC
1	IBM RS/6000 SP Server w/8 nodes	4 per node @333 MHz	3 GB/node	1 TB	High Speed Switch, SMP, 3570 Tape	BCC	100MB/s NIC per node
1	IBM RS/6000 SP Server w/20 nodes	4 per node @333 MHz	3 GB/node	1 TB	(2) High Speed Switches, SMP, ADSM	BCC	100MB/s NIC per node

The manner in which all software packages are installed on the RS/6000 SP ensures that scalability and reliability is maximized. Each of the significant software packages is installed on multiple nodes, ensuring operation even if a node fails, while multiple nodes also provide additional processing power. No single node failure can compromise system integrity and the RS/6000 SP nodes have been divided according to software demands (based on memory needs, CPU needs, I/O throughput, etc.). The following table lists the American FactFinder subsystem's major software components.

American FactFinder (AFF) Major Software Components	
Software	Description
IBM AIX and components	IBM's AIX is the primary DADS RS6000/SP operating system. It is a 64-bit, reliable, scalable variant of the UNIX operating system, geared for parallel processing in symmetric and distributed parallel environments. It includes components such as the High Availability Cluster Monitor Package and High Availability Control Workstation providing reliable operation, even under degraded conditions. The DADS AIX installation includes a comprehensive backup facility (ADSM) which manages a tape library and powerful monitoring and reporting tools.
ESRI: SDE and IMS	The ESRI COFS family of software products satisfies the majority of our geographic and mapping needs. Although some customization of the COFS code is necessary, we have fully utilized the COFS ESRI Internet Map Server (IMS) and Spatial Data Engine (SDE).
Oracle Application Server	Oracle's OAS is key to the DADS AFF architecture. It provides Web service and a framework for custom server-side application software in a manageable environment that spans hardware. OAS is optimized for deployment across multiple nodes (or separate computers), and handles much of the administrative burden of balancing processing load.

American FactFinder (AFF) Major Software Components	
Software	Description
	OAS allows application development in several different languages. The running applications execute in an OAS cartridge environment. OAS schedules inbound web requests to available cartridges depending on request type and load balance. Within the cartridge, application software generates HTML output (using HTML libraries provided by OAS), which are then routed back through the web server to the requesting browser. Inside the cartridge, application code has access to Oracle databases, the OAS environment, and HTML generation capability.
Oracle Parallel Database	<p>The U.S. Census Bureau selected Oracle's database software for DADS as the new standard for relational databases because it consistently performs well, runs on many platforms, and has leading-edge functionality. Oracle's version titled, "Oracle 8 Parallel Server (OPS)" is optimized for execution across multiple nodes on machines like the IBM RS/6000 SP. Based on our analysis, we expect the number of database queries to be substantial. We selected OPS for this reason and the fact that it provide better throughput in our hardware environment despite its requirement for more administrative care in setup and tuning.</p> <p>Multiple nodes can participate as part of a large "data warehouse," each contributing to overall processing capability, and each node able to support queries. Because of the RS/6000 SP's low-latency virtual disk mechanism, queries executed on any node can access database tables from any other node at very little degradation in performance. This arrangement gives the database administrators greater latitude in laying out tables.</p>
JDBC	AFF communicates with databases via the Java standard JDBC, which is supported in Java's JDK. JDBC specifies standard ways for communicating with databases that implement the SQL standard. Any database that has a JDBC driver can be connected into AFF.
Web Browsers	Web Browsers display pages of information that are received from the server, and perform limited program execution. HTML (version 3.2 or later) is the current standard for information formatting. Since users are responsible for selecting, installing, and configuring their web browsers, DADS deliberately avoided using Java applet execution in the browser for DADS, except for optional functionality, due to concerns that some users could not run Java applets. To achieve the graphical look and feel desired, some custom browser programming in JavaScript was required.
HTML, URL, HTTP	Communication between the client and the server is done strictly with web protocols: HTTP network traffic, carrying URLs from the client and HTML (version 3.2) from the server. This protocol is a widely accepted worldwide standard and is typically passed through firewalls. DADS uses this API for communicating between the internal and external servers.
Custom Java Code	Java is the standard language for DADS application development. DADS has two environments for which Java code is written: the RS/6000 SP server (inside OAS Java Cartridges) and in the browser (in optional Java applets). DADS will use JDK 1.1 despite limited browser support due to better stability, performance, and User Interface.
Custom JavaScript Code	DADS maintains session state information in browser cookies. Because cookies only store string values and are limited in number, managing cookies is difficult and has required substantial JavaScript code. It is our intention to eliminate the JavaScript-based cookie manipulation code when a decision is reached on which browsers to support.
SQL	ANSI SQL-92 is the standard for databases which are JDBC compliant.
XML	XML, the eXtended Markup Language, is a derivative of SGML (Standard Generalized Markup Language), and is a successor to HTML. Where HTML describes content-free layout, SGML describes layout-free content. XML is rapidly gaining popularity as a data exchange format. DADS uses XML 1.0 as the base for the XAP (XML Application Project) File Format, which is designed to be a project specification file for the GIS Data Server/IMS products.
CORBA	Industry standard for interconnection between object application environments; it establishes a way to send messages between objects on different systems, and conventions for locating objects. It also provides a standard collection of services available to all object environments on the network. Although AFF is not making direct use of CORBA at present, the Oracle Application Server (OAS) is built to CORBA standards (for communication between cartridges). There is planning in place to use CORBA-connected clients as an alternative to HTML-based browsers. In using Java-enabled clients talking to AFF via CORBA, we can deliver a higher-quality user interface, and off-load processing from our server.

2.1.2 American FactFinder (AFF) Progress Against Planned Milestones

Most of the milestones indicated in the 1999 Operational IT Plan were accomplished on schedule as presented. The single exception was the public release of the Internet version of AFF, originally scheduled for January 4, 1999 but released on March 15, 1999. The delay was a result of a number of issues (see the table below) but these have been overcome and the schedule re-established. The internal and external AFF subsystems have been conditionally accepted with active Defect Correction Plans.

External American FactFinder (AFF) Delay Factors	
Factor	Description
Budget	Redefining and clarifying existing requirements consumed more of the budget than originally planned. To accomplish necessary DPP subsystem development in FY98/99, further AFF requirements analysis was postponed. These analyses will be completed as early as possible so that design/development work on DADS2000 can commence.
Technology	To assure the designed level of performance, IBM accomplished a technology upgrade to the system designated for the external component of AFF. This established both external and internal systems in identical types and levels of technology and introduced the ability to load balance by moving components between individual systems as needed. This was done at no cost to the government.
User responsiveness as a result of operational experience	DADS is a tool rather than a process and therefore is sensitive to customer feedback and highly responsive to their communicated needs.  Even with the extensive design modeling and theoretical design extrapolations, preliminary user feedback was somewhat attenuated by the lack of a hands-on system. Once our customers had an AFF to test-drive, we began to gather truly effective data which, to an extent, influenced further design direction. Many of the suggestions and issues raised were deemed to be significant enough to warrant a delay in public release until we could act on these issues.

American FactFinder (AFF) Milestones, FY 99					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Initial Release of Production AFF	01/99	01/99	01/99	01/99	Completed
Security Plan approval	01/99	02/99	01/99	02/99	Completed
Conduct program/System Risk Assessment	02/99	02/99	02/99	02/99	Completed
Release 2 of Production AFF	06/99	06/99			
Security Certification and Accreditation	06/99	06/99	01/99	02/99	Temporary, valid for six months
Security Certification and Accreditation (Final)	08/99	08/99			

**American FactFinder (AFF) Milestones, FY 00**

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Y2K Compliance Testing	01/99	10/99	01/99		Ongoing
Y2K Compliance Certification	10/99	10/99			
Requirements Gathering	12/98	12/99	12/98		Ongoing
Y2K Full Regression Testing for Verification	11/99	12/99			
Requirements Analysis, Planning and Documentation	01/99	01/00	01/99		Ongoing
Security Plan Review and Revision	02/00	03/00			
Design Processes	06/99	04/00			
Post Release Development and Enhancement	01/99	07/00	01/99		Ongoing
Production Operation	01/99	09/00	01/99		Completed/Ongoing
Integration Testing	07/00	09/00			

**American FactFinder (AFF) Milestones, FY 01**

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Training	09/00	10/00			
Acceptance Testing	09/00	10/00			
Deployment for Census 2000	10/00	11/00			
Security Plan Review and Revision	02/01	03/01			
Production Operation	10/99	09/01			
Production	12/00	09/01			
Post Production enhancement	12/00	09/01			
AFF System Enhancement/ Requirements Gathering, Review and Analysis	12/01	09/01			

**American FactFinder (AFF) Milestones, FY 02**

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Security Plan 3-Year Re-Write	02/02	03/02			
Security Re-Certification and Re-Accreditation (3-year)	06/02	06/02			
AFF System Enhancement/ Requirements Gathering, Review and Analysis	10/01	09/02			
Post Production Enhancement	10/01	09/02			
Production Operation	10/00	09/02			

American FactFinder (AFF) Milestones, FY 03					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Security Plan Review and Revision	02/03	03/03			
Production Operation	10/02	09/03			
AFF System Enhancement/ Requirements Gathering, Review and Analysis	10/02	09/03			
Post Production Enhancement	10/02	09/03			

2.1.4 American FactFinder (AFF) Performance Measures

American FactFinder Performance Measures			
Performance Goals	Performance Measures	Target Performance	Current Performance
Availability of the AFF system to act as a tool in support of production and dissemination requirements	% of goals using AFF products	100%	100%
	% of performance error rate	100%	100%
	% user friendliness and product quality	100%	100%
	% of system availability	98%	98%
	% of competence after training	98%	98%
AFF system capable of producing and disseminating acceptable data products on time and in the quantities required	% of goals using AFF products	100%	100%
	% of performance error rate	100%	100%
	% user friendliness and product quality	100%	100%
	% of system availability	98%	98%
AFF will serve a larger customer base with greater speed, accuracy and more tailored customer service	% of goals using AFF products	100%	100%
	% of performance error rate	100%	100%
	% of system availability	98%	98%
Availability of Tier 1,2,3 data access and dissemination capabilities on required dates and in accordance with response time requirements	% of performance error rate	100%	100%
	% of capacity availability	100%	0%
System performance	% of planned capacity	100%	100%
	% of planned throughput	100%	100%
	% of planned response times	100%	100%

### 2.1.5 American FactFinder (AFF) Risks

The DADS Program has identified a series of strategic risks which overlap the two major subsystems (AFF and DPP) and therefore will be presented in a consolidated manner. Any segment which applies to a specific subsystem will be identified.

**1. DADS does not meet the needs of its user community due to limited empirical data on which to base planning estimates.**

The focus of DADS current design is based on a performance model with little firm empirical data and from input from internal U.S. Census Bureau representatives serving as proxies for external users and organizations. DADS current design capacity is for a maximum of 1400 concurrent users on both the internal and external systems. Current analysis projects a potential for 2,000 internal and 4.0 million external system hits monthly. If this degree of potential is realized, we conclude the current AFF system design is inadequate and would be overwhelmed.

The DPP subsystem's innovative implementation of highly sophisticated data product and tabulation tools has the potential to encourage other U.S. Census Bureau organizations to explore its use for production of their data products. However, since current DPP subsystem requirements are static in nature and focused on a finite set of tabular products, the current design is not comprehensive enough to allow expansion for new product areas. The current DPP architecture concept will not be able to accommodate a wider variety of U.S. Census Bureau data and data products without expansion to provide new product areas.

#### *MITIGATING ACTIONS*

1. Pursue avenues of funding in advance of anticipated need;
2. Target Specific functional areas of AFF for performance enhancements;
3. Develop a system limitation mechanism to control the number of concurrent users allowed access;
4. Design/redesign pricing methodologies to minimize use of resource-intensive functionality;
5. Develop and implement a Communications and Deployment Plan to educate potential AFF clients as to the latest developments in Release and Content Schedules;
6. Develop and distribute written User Guides defining functionality and content to clients when subsystems are deployed;
7. Be proactive and timely in the analysis of future requirements and ensure those requirements previously collected but not implemented are placed on the table for inclusion;
8. Follow a design process that ensures the current and future AFF architecture is (and remains) open and compatible with OSI concepts; and
9. Develop strong working relationships between AFF and other U.S. Census Bureau Planning Groups and operational staffs.

**2. DADS does not meet established deadlines due to untimely delivery or delivery of inaccurate data files from supplying organizations.**

The utility of DADS is directly tied to the availability of data and data products that are collected from diverse and disparate U.S. Census Bureau sources and are subject to internal agendas and organizational complexities. Failure of the responsible supplying organizations to deliver, or deliver on time, the necessary data files or products will lead to a serious loss of credibility for DADS and the U.S. Census Bureau. It should be noted that the release schedule of the Public Law 94-171 (PL) listings and summary files is mandated by law and failure to meet that schedule will incur significant legal repercussions.

**MITIGATING ACTIONS**

1. Establish effective working relationships with supplier organizations to foster a cooperative working environment and to establish a clear understanding that timely delivery of accurate data files and products is in the best interest of all;
2. Support corporate buy-in and acceptance of DADS as the primary tool for census data dissemination;
3. Set milestones for the development process to bilaterally monitor the progress of data and data products destined for dissemination through DADS;
4. Establish a tracking mechanism/tool to address the previous mitigator; and
5. Establish and publish 'drop dead' dates as management tools to meet the established schedules and projections.

**3. DADS does not meet established deadlines due to its inability to acquire and/or retain staff with requisite critical skills.**

DADS depends heavily upon skilled government staff with appropriate policy and practice expertise made available to this program only through the cooperation of the program areas. Additionally, hindered by the salary guidelines of the General Schedule, DADS is not competitive on the open market when searching for the requisite technical skills. Training the staff has proven ineffective since the extensive cutting-edge technical and in-depth subject matter experience needed can't be taught and those staff garnering high-demand skills often leave for better career opportunities.

DADS faces a tangible risk in [1] having difficulty finding qualified and effective replacements, and [2] losing the cumulative programmatic knowledge and experience. This has grave impacts on current development efforts and future intentions involving the transfer of technology to government staff and the eventual assumption of program responsibilities by government staff. This could make DADS contractor-dependent for an extended period.

*MITIGATING ACTIONS*

1. Define clear career paths for government staff and provide necessary training. Place conditions ensuring continuation of government employment for staff receiving census-sponsored training;
2. Effectively use the IT Services Contracts for support to find the highly specialized technical skills necessary to the success of DADS. Project staffing requirements in a timely manner;
3. Effectively use contractor support to achieve the necessary flexibility in staffing to maintain the appropriate number and types of personnel to meet workload levels and technical requirements;
4. Maximize the salary pool to effectively compensate qualified staff, e.g., increase program funds;
5. An effective transfer of technology must take place through the machinations of the Integrated Product Team (IPT) process. Government staff must be fully involved and cognizant of contractor efforts; and
6. Employ a strategy of COTS software and hardware standardization.

**The confidentiality and security requirements of the U.S. Census Bureau are compromised due to DADS' use of Internet technology to disseminate data.**

For any Web-based application, malicious or frivolous access, i.e., hacking, is a major concern. The connection to the World Wide Web, Internet technology concepts such as firewall security and the implementation of confidentiality filters all have inherent risks and promulgate a degree of serious concern.

*MITIGATING ACTIONS*

1. Work with appropriate U.S. Census Bureau organizations to develop and employ a rigorous set of edits and programmed rules (confidentiality filters) that will insure that no disclosure of Title 13 data can occur;
2. Identify and become intimately familiar with all required government standards;
3. Implement a rigorous discovery and testing process to identify any security weaknesses;
4. Attain proper security accreditation and certification through established U.S. Census Bureau Security processes;
5. Implement, at earliest opportunity, the software interface between IBM AIX and Novell NDS;
6. Utilize the corporate Novell infrastructure to achieve independent and centralized control of access and utilization; and
7. Employ IBM's corporate knowledge regarding Web-page security design and implementation.
8. Conduct 'ethical hacking' exercises to identify weak points in system security.

**5. New or redefined requirements cause the scope of contract or baseline requirements to creep.**

With the significant number of 'unknowns' inherent to developmental programs such as DADS, requirements are often poorly defined or misinterpreted due to a lack of understanding or poor communication and achieve clarity only after much investigation and discussion.

Additionally, the focus of system development is often subject to changes in organizational priorities, corporate politics, and resistance to change. Capabilities not envisioned at the outset become critical due to new guidance, newly perceived need, or the introduction of new technology. As the system becomes functional and the user community becomes aware of the system's utility, demand increases and the capacity/capability to do more is required.

**MITIGATING ACTIONS**

1. DADS has an efficient and effective Change Control Process in place to manage the introduction of baseline variances (new, deleted, or modified/clarified requirements etc.); Apply the DADS Change Control Process rigorously;
2. Involve user groups and client organizations early in the process to define and clarify the requirements gathered. Establish concise lines of communication and foster open discussion;
3. Apply programmatic knowledge, during each iteration or release of DADS, of the lessons learned in a proactive manner during collection and distillation of follow-on requirements; and
4. Ensure that the contract vehicle is designed to best address the current and potential needs of the program. Maintain adequate contract expertise to accomplish necessary modifications, negotiations, and maintenance of contract vehicles.

**6. DADS Service fails due to single points of failure.**

DADS is designed around two IBM RS/6000 SP computing platforms, one for each the Internet and Intranet components. Any major failure that overwhelms the system's 'fail-over' redundancy design would incapacitate the applicable component system and result in a loss of that specific Internet or Intranet capability. The current program paradigm does not take into consideration the concept of a contingency site or of a redundant, mirrored, or 'hot' system that would replicate system capability.

With DADS permanently housed at the Bowie Computer Center, we must rely upon the resident organization to address contingency operational issues, and we have serious concerns with regard to the lack of a current and regularly tested Site Contingency Plan that addresses the particularities and requirements of the IBM SP environment.

*MITIGATING ACTIONS*

1. Work with U.S. Census Bureau and Bowie Computer Center staffs to develop a comprehensive U.S. Census Bureau Contingency Plan to activate in the event of anomalous conditions at the Bowie Computer Center;
2. Begin contingency planning focused on addressing the need for a redundant, mirrored, or 'hot' site;
3. Establish a reciprocal agreement(s) with existing sites to serve as a backup to DADS in the event of a system failure, catastrophic event or act(s) of vandalism; AND
4. Secure funding to implement a duplicate system or capability at a second geographically separate site.

**2.1.6 American FactFinder (AFF) References**

The American FactFinder subsystem is supported by the following planning documents:

- Budget Submission for FY 2001, dated June 1, 1999;
- Census 2000 Systems Architecture Report, dated March 1, 1999;
- Census Information Technology Review Guide, dated February 2, 1999;
- DADS98 AFF Release 1 Test Results, dated January 4, 1999;
- 1999 Strategic IT Plan, dated December 18, 1998, pages 28, 61, 63, 93;
- DADS Y2K Compliance Review, dated October 10, 1998;
- DADS Program Management Plan, continuous updates from June 1997 to February 1999;
- DADS Comprehensive Development Plan, continuous updates from June 1997 to February 1999;
- Requirements Analysis Documentation for DADS Inquiry System, continuous updates from June 1997 to February 1999;
- DADS Architecture Plan (version 2.0), continuous updates from June 1997 to January 1999;
- DADS Business Case Analysis Report, May 1997;
- DADS 1996 Test Results, dated March 17, 1997;
- 1997 DADS Analysis Report, dated March 11, 1997; and
- Census Security Plan CEN230, continuous updates from February 1996 to February 1999.

2.2.1 Detailed Description of Data Products Production (DPP)

DPP creates and delivers 1998 Dress Rehearsal and 2000 Decennial data products, including Public Law 94-171 (PL) listings and summary files, the Sample Summary File (SSF), and the Hundred Percent Summary File (HSF).

The DADS DPP implementation will start with installing the latest version of SuperCROSS client and server software, configuring the operating environment to meet production requirements, and creating data preparation and loading scripts. The SuperCROSS tabulation products do not presently run on SP/AIX systems, but the development of a compatible version is in progress and will be delivered in the last quarter of FY1999. In the meantime, DADS is developing DPP on a pair of Compaq NT servers and is planning to accomplish the initial

processing of DPP data in this environment. When the SP/AIX version of SuperCROSS is delivered, we will move all DPP development and production process to the RS/6000 SP (internal) system.

While the NT servers are not as scalable nor as fault-tolerant as the SP server, they meet our anticipated short-term needs. SuperCROSS tabulation makes extensive use of memory, and can handle datasets of such size as to be limited only by the amount of available memory. Thus any server hosting SuperCROSS software must be configured with a significant amount of memory.

The following table lists the Data Products Production subsystem's major hardware components.

Data Products Production (DPP): Major Hardware Components							
Quantity	Vendor/ Model	CPU	RAM	Storage	Components or Modules	Physical Location	Comments
2	Compaq MS NT-based	333 MHz	128 Mb	18 GB	File servers	HQ	100 MB/s NIC

The table on the following page lists the Data Products Production subsystem's major software components.

Data Products Production (DPP) Major Software Components	
Software	Description
Microsoft Windows NT	Microsoft's Windows NT is the operating system running on the Compaq servers that support the DPP Component of DADS. Windows NT is a 32-bit, reliable operating system geared for single- and small-parallel processing in symmetric environments. It includes monitoring and reporting tools, and is integrated with Novell software to allow access to remote file systems.
SuperCROSS	SuperCROSS is a COTS integrated data aggregation and dissemination system software technology that performs tabulations across large datasets with minimal consumption of computer system resources. We are using it specifically for tabulating Decennial large-scale statistical data. One of the main components is SuperSTAR, which supports client/server architecture and currently runs on Windows 95/NT systems. STR is porting the SuperSTAR suite to AIX and DADS will move all existing SuperCROSS processing to the IBM RS6000/SP by late 2000.
SAS	We use SAS for transforming files from one format to another as part of DPP process. We use SAS programs to process incoming data (from data providers) into a SuperCROSS compatible format, and extract information from SuperCROSS into other necessary forms.
HTML, URL, HTTP	Communication between the client and the server is done strictly with web protocols: HTTP network traffic, carrying URLs from the client and HTML (version 3.2) from the server. This protocol is a widely accepted worldwide standard and is typically passed through firewalls. DADS uses this API for communicating between the internal and external servers.

SuperCROSS is purpose-built for DPP-type tabulations and is used at other national statistical agencies. It offers:

- a graphical interface well suited to component and tabular specification;
- native understanding of important concepts (such as dimensions, universes, recodes, and geographical hierarchies);
- the ability to handle hierarchically organized input data; and
- a production module capable of volume production of statistical tables.

SuperCROSS operates with Windows 95/NT clients running against a local database or against a Windows NT server.

2.2.2 Data Products Production (DPP) Progress Against Planned Milestones

All of the milestones indicated in the submitted 1998 Operational IT Plan were accomplished on schedule as presented. The DPP subsystem has been conditionally accepted with an active Defect Correction Plan in place. We expect formal acceptance to take place by the end of April 1999.

Data Products Production (DPP) Milestones, FY 99					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Conduct DPP System Risk Assessment	02/99	02/99	02/99	02/99	Completed
Conditional Acceptance of NT based DPP system	02/99	02/99	02/99	02/99	Completed
Security Certification and Accreditation	06/99	06/99	01/99	02/99	Temporary, valid for six months
Formal Acceptance of NT based DPP system	04/99	04/99			In process
Initial release of PL 94-171 data	04/99	04/99	04/99	04/99	Completed
Security Certification and Accreditation (Final)	08/99	08/99			
Port to AIX/SP Platform	09/99	09/99			

Data Products Production (DPP) Milestones, FY 00					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Y2K Compliance Testing	04/99	10/99	04/99		Ongoing
Y2K Compliance Certification	10/99	10/99			
Requirements Gathering	12/98	12/99	12/98		Ongoing
Y2K Full Regression Testing for Verification	11/99	12/99			
Requirements Analysis, Planning and Documentation	01/99	01/00	01/99		
Security Plan Review and Revision (DPP Component)	02/00	03/00			
AIX-based Development	09/99	07/00			
Production Operation	04/99	09/00	01/99		Completed/ongoing
Integration Testing	07/00	09/00			

Data Products Production (DPP) Milestones, FY 01					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Training	09/00	10/00			
Acceptance Testing	09/00	10/00			
Deployment for Census 2000	10/00	11/00			
Security Plan Review and Revision	02/01	03/01			
Production	12/00	09/01			
DPP System Enhancement/ Requirements Gathering, Review and Analysis	12/00	09/01			
Post Production Enhancement	12/00	09/01			

Data Products Production (DPP) Milestones, FY 02					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Security Plan Review and Revision	02/02	03/02			
Security Re-Certification and Re-Accreditation (3-year)	06/02	06/02			
DPP System Enhancement/ Requirements Gathering, Review and Analysis	10/01	09/02			
Post Production Enhancement	10/01	09/02			
Production Operation	10/01	09/02			

Data Products Production (DPP) Milestones, FY 03					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Security Plan Review and Revision	02/03	03/03			
Production Operation	10/02	09/03			
DPP System Enhancement/ Requirements Gathering, Review and Analysis	10/02	09/03			
Post Production Enhancement	10/02	09/03			

2.2.3 Data Products Production (DPP) Performance Measures

Data Products Production Performance Measures			
Performance Goals	Performance Measures	Target Performance	Current Performance
Availability of the DPP system to act as a tool in support of production and dissemination requirements	% of goals using DPP products	100%	100%
	% of performance error rate	100%	100%
	% user friendliness and product quality	100%	100%
	% of system availability	98%	98%
	% of competence after training	98%	98%
DPP system capable of producing and disseminating acceptable data products on time and in the quantities required	% of goals using DPP products	100%	100%
	% of performance error rate	100%	100%
	% user friendliness and product quality	100%	100%
	% of system availability	98%	98%
DPP will serve a larger customer base with greater speed, accuracy and more tailored customer service	% of goals using DPP products	100%	100%
	% of performance error rate	100%	100%
	% of system availability	98%	98%
DPP System performance	% of planned capacity	100%	100%
	% of planned throughput	100%	100%
	% of planned response times	100%	100%

2.2.4 Data Products Production (DPP) Risks

Please see section 2.1.4, American FactFinder Risks.

2.2.5 Data Products Production (DPP) References

The Data Products Production subsystem is supported by the following planning documents:

- Budget Submission for FY 2001, dated June 1, 1999;
- Census 2000 Systems Architecture Report, dated March 1, 1999;
- Census Information Technology Review Guide, dated February 2, 1999;
- DADS98 AFF Release 1 Test Results, dated January 4, 1999;
- 1999 Strategic IT Plan, dated December 18, 1998, pages 28, 61, 63, 93;
- DADS.Y2K Compliance Review, dated October 10, 1998;
- DADS Program Management Plan, continuous updates from June 1997 to February 1999;
- DADS Comprehensive Development Plan, continuous updates from June 1997 to February 1999;
- Requirements Analysis Documentation for Data Products Production (DPP) System, continuous updates from June 1997 to February 1999;
- DADS Architecture Plan (version 2.0), continuous updates from June 1997 to January 1999;
- DADS Business Case Analysis Report, May 1997;
- DADS 1996 Test Results, dated March 17, 1997;
- 1997 DADS Analysis Report, dated March 11, 1997;
- 1997 DADS DPP Analysis Report; and
- Census Security Plan CEN230, continuous updates from February 1996 to February 1999.

### 3.0 DADS Infrastructure Description

#### 3.1 Detailed Description of DADS Infrastructure

DADS infrastructure is supported by the Decennial program area.

#### 3.2 DADS Infrastructure Progress Against Planned Milestones

Data Access and Dissemination Systems (DADS) Infrastructure Milestones, FY 99					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Technology Refreshment	10/98	09/99	10/98		Ongoing

Data Access and Dissemination Systems (DADS) Infrastructure Milestones, FY 00-05					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Technology Refreshment	10/99	09/05			

#### 3.3 DADS Infrastructure Performance Measures

Not applicable.

#### 3.4 DADS Infrastructure Risks

Not applicable.

#### 3.5 DADS Infrastructure References

Not applicable.

June 7, 2000

CENSUS 2000 INFORMATIONAL MEMORANDUM NO. 9

MEMORANDUM FOR      Distribution List

From:                    Susan Miskura (Signed)  
                              Chief, Decennial Management Division

Subject:                Census 2000 Language Program Master Plan

Contact Person:        Carol Briggs, Content and Products Branch, Decennial  
                              Management Division, Room 1422, Bldg.2, (301-457-8228)

The Census 2000 Language Program Master Plan is attached for your information.

This Program Master Plan was reviewed by Census Managers. Every attempt has been made to document the Language Program Operations and Procedures which will occur during Census 2000. Readers are advised that any major changes to policies, procedures, or operations will be submitted as an addendum to this plan.

Attachments

# Census 2000 Language Program Master Plan

<u>Table of Contents</u>	<u>Page</u>
<b>I. Summary Description and Purpose of the Operation.....</b>	<b>1</b>
<b>II. General Design and Work Flow.....</b>	<b>1</b>
<b>A. Questionnaire Special Requests.....</b>	<b>1</b>
1. Overview .....	1
2. Translation, Printing and Distribution .....	2
3. Processing .....	2
<b>B. Questionnaire Assistance Centers .....</b>	<b>3</b>
1. Overview .....	3
2. Questionnaire availability .....	4
3. Detailed Information .....	4
<b>C. "Be Counted" .....</b>	<b>4</b>
1. Overview .....	4
2. Detailed Information .....	5
<b>D. Language Guides .....</b>	<b>5</b>
1. Overview .....	5
2. Languages .....	5
3. Distribution .....	6
4. Translation .....	6
5. Printing .....	6
<b>E. Telephone Questionnaire Assistance (TQA).....</b>	<b>6</b>
1. Overview .....	6
2. Responsibilities .....	6
3. Questionnaire availability .....	7
<b>F. Education and Outreach .....</b>	<b>7</b>
1. Overview .....	7
2. Educational Materials .....	8
3. Printing and Distribution .....	9
4. Implementation Strategy .....	10
<b>G. Advertising .....</b>	<b>10</b>
1. Overview .....	10
2. Young and Rubicam Coordination .....	11

<b>III. Cost Assumptions .....</b>	<b>12</b>
A. Overview .....	12
B. Components .....	12
<b>IV. Schedule/Responsibilities .....</b>	<b>12</b>
A. Overview .....	12
B. Schedule .....	12
<b>V. Evaluation Requirements .....</b>	<b>13</b>
A. Overview .....	13
B. Specifications .....	13
<b>VI. Major Differences from Dress Rehearsal .....</b>	<b>13</b>
A. Dual Language Mailout .....	13
B. Questionnaire Special Requests .....	13
C. Language Assistance Guides .....	14
D. QAC Paid Temporary Staff .....	14
E. Large Household Follow-up .....	14
F. Processing .....	14
G. TQA Special Mail Requests .....	14
<b>VII. Final Translation Contractor Report .....</b>	<b>15</b>
<b>VIII. Census 2000 Language Memorandum.....</b>	<b>16</b>
A. Census 2000 Decision Memorandum No. 77: Changes Planned for the Census 2000 Language Program (March 25, 1999).....	16
B. Census 2000 Informational Memorandum No. 8: Procedures for Obtaining a Foreign-Language Questionnaire by Return of the Advance Letter (April 9, 1999).....	17
C. Census 2000 Informational Memorandum No. 18: Process Used to Select Languages for Census 2000 Language Assistance Guides (June 21, 1999).....	18

## **Attachments**

- Attachment A. Census 2000 Advance Letter  
Census 2000 Questionnaires (D-1 and D-2): English, Chinese, Korean, Spanish, Tagalog, and Vietnamese
- Attachment B. Census 2000 Contracts Matrix - Language Related - DSCMO
- Attachment C. "Be Counted" Questionnaires (D-10) and Envelopes: English, Chinese, Korean, Spanish, Tagalog, and Vietnamese
- Attachment D. Language Assistance Guide - Census 2000 Languages (49)
- Attachment E. Language Assistance Guide - Questionnaire Format/Layout
- Attachment F. External Communications Materials for Census 2000 (Appendix A) - C2PO
- Attachment G. RFP: Advertising Services for Census 2000 (Section C) - C2PO
- Attachment H. Cost Model - Questionnaire Printing
- Attachment I. Master Activity Schedule - Language Related
- Attachment J. Census 2000 Language Assistance Guides (D-60A/short, D-60B/long) and Language Identification Flashcard (D-3309)

# Census 2000 Language Program Master Plan

## I. Summary Description and Purpose of the Operation

The objective of all Census 2000 programs is to raise the response rates, improve the data accuracy, and conduct a cost effective decennial census. The purpose of the Census 2000 language program is to support that goal by providing census information and assistance in languages other than English (languages other than Spanish in Puerto Rico). Census Bureau research indicates that the inability to speak or read English (Spanish in Puerto Rico) well is a barrier to a successful enumeration. The availability of multiple languages helps create a Census 2000 climate that promotes goodwill and cooperation between the Census Bureau, our census partners<sup>1</sup>, and respondents throughout the nation. This language program master plan applies to stateside and Puerto Rico (PR) operations.

## II. General Design and Work Flow

In Census 2000, as in every census since 1940, a sample of households will be asked to respond to more questions than other households. The majority of households will receive a "short" form which has 7 content questions and takes an average of 10 minutes to complete, but about one in every six households will receive a "long" form which has 52 content questions and takes an average of 38 minutes to complete. Even though the national mail return response rate is expected to approximate 67 percent (51 percent for PR), the Census Bureau plans to take the following additional steps through the language program to motivate individuals to respond nationwide.

### A. Questionnaire Special Requests

1. Overview: The original operations plan for Census 2000 was to mail two questionnaires--one English and one other language--to households in selected neighborhoods believed to have a high proportion of linguistically-isolated households<sup>2</sup>. The Census Bureau's ability to target

---

<sup>1</sup> Census 2000 partners include any group that might help us reach our goal of raising response rates and improving accuracy in the census.

<sup>2</sup> Linguistically isolated households are those households where the language spoken is other than English and no one over the age of 14 speaks or reads English very well.

linguistically-isolated households was limited and the operational hurdles associated with implementing a dual mailing during the Census 2000 Dress Rehearsal were greater than anticipated. The revised Census 2000 language plan offers the following alternative: all households receiving an addressed advance letter will have the opportunity to request a questionnaire in Spanish, Chinese, Tagalog, Vietnamese, or Korean. Procedures for obtaining a Foreign-Language Questionnaire by return of the advance letter are included in Attachment A.

The Census Bureau selected the five additional languages listed above by reviewing the ten languages spoken by the largest populations in the United States. These ten language groups were then crossed with the Census Bureau's established "hard to count" index. The resulting largest top five linguistically isolated household populations speaking a language other than English (which have historically been the most difficult to enumerate) spoke Spanish, Chinese, Vietnamese, Tagalog, and Korean. In Puerto Rico, the largest linguistically isolated household population speaking a language other than Spanish spoke English.

2. Translation, Printing and Distribution: The Census 2000 English mailout/mailback short and long questionnaire will be translated and printed in Spanish, Chinese, Tagalog, Vietnamese, and Korean. These questionnaires will be available upon request to any stateside respondent receiving a Census 2000 advance letter. Respondents are asked to indicate which language questionnaire they would prefer, return their advance letter in an enclosed prepaid envelope, and hold the English questionnaire they will automatically receive through the mail. The advance letter will be imprinted with the Census barcode identifier associated with each specific household. Facsimiles of the advance letter and questionnaires are included in Attachment A. A copy of the printing contracts matrix describing all language forms and quantities is included as Attachment B. In Puerto Rico, the Update/Leave short and long questionnaires will be available in Spanish and English via the Telephone Questionnaire Assistance and Questionnaire Assistance Center procedures.
3. Processing: Specially requested questionnaires in languages other than English will be processed at the Census Bureau's National Processing Center (NPC) in Jeffersonville, IN. The non-English forms will follow all of the processing steps of English forms, and additional translation steps. Spanish language questionnaires are processed identically to English language questionnaires therefore no translation or special handling will be performed (except before coding).

All mailed back questionnaires will be checked in through the laser sorter at the NPC. The sorter will read the barcode on the questionnaire and send all non-English language forms for special handling. A processing ID, necessary for formation of the Decennial Response File (DRF), will be added to all "Be Counted" forms at check-in. Following this procedure, the "Be Counted" forms will follow the same process as the short and long forms. Non-English, non-Spanish language questionnaires will be scanned and the image retained as a back-up prior to sending the questionnaires to a special handling unit.

Clerical staff in the special handling unit will examine the questionnaires to determine whether the write-in responses are completed in English or a non-English language. If the responses are in English, a clerk will transcribe all of the responses on the questionnaire to a corresponding English questionnaire, add an appropriate barcode to identify the form as a replacement, and send the English questionnaire back to complete the processing steps. If the responses are in a non-English (non-Spanish) language, a clerk will log the form out of the questionnaire tracking system and send it to the appropriate sub-contractor for translation and transcription on a corresponding English questionnaire.

All non-English (non-Spanish) questionnaires will be delivered to the translation contractor on a daily basis. The contractor will translate and transcribe the information onto the corresponding English questionnaire and return both questionnaires to the NPC by the following day. Once returned, the new, transcribed English questionnaires will be logged back into the tracking system as replacements and processing will continue as for any other form.

## **B. Questionnaire Assistance Centers**

1. Overview: Questionnaire Assistance Centers (QACs) are facilities designed to assist local residents, including those needing assistance because of language barriers, in completing their census questionnaires. The QAC program increases the emphasis on Census 2000 community-based support efforts which take advantage of regional and local knowledge. The QAC facilities are an opportunity to strengthen the community-based language program given the concerns about functional literacy among some linguistically isolated households and the inaccessibility of telephones in some communities.

2. Questionnaire availability: Foreign language questionnaires will not be available at QACs. The Census Bureau will furnish each QAC with Language Assistance Guides in more than 40 languages to help respondents complete their official Census 2000 questionnaire. A Census 2000 Language Identification Flashcard also will be available for QAC staff and respondents to identify which language is spoken in the respondent household.

If a respondent is insistent about obtaining a replacement/alternative questionnaire, and a "Be Counted" questionnaire is not appropriate, special requests will be handled in the following manner.

- a. Staff will first ascertain the reason for a special request questionnaire, the mailing address of the requestor, and the Census ID printed on the original Census 2000 questionnaire if possible.
- b. If the requestor **can supply** the Census ID from the original questionnaire, staff will forward the request to the Local Census Office. These requests will be compiled and forwarded to NPC in Jeffersonville, IN. The NPC staff will assemble a new mailing package with an English questionnaire corresponding to the original Census ID type and a language assistance guide of the requestors choice. Staff will generate an address label with the correct address and Census ID barcode, address the new package, and mail to the respondent. Requests from Puerto Rico will be handled in the same manner.
- c. If the requestor **can not supply** the Census ID, staff will provide and/or assist the respondent with a stateside or Puerto Rico "Be Counted" questionnaire.

3. Detailed information: Detailed information on locations, staffing, criteria, deliverables, and so forth can be found in the Questionnaire Assistance Center Program Master Plan.

#### C. "Be Counted"

1. Overview: The "Be Counted" questionnaire will provide a means for people to be included in Census 2000 who may not have received a census questionnaire or believe they were not included elsewhere. "Be Counted" forms are intended to serve as an accessible alternative for households missed through the mailout or in update/leave and list/enumerate areas. They are not intended for informational purposes or as a substitute for the Census 2000 mailout, update/leave, or enumerator questionnaires.

"Be Counted" questionnaires will be printed in English, Spanish, Chinese, Tagalog, Vietnamese, and Korean. The "Be Counted" operation will begin just before Census Day and will end just before the start of the nonresponse follow-up operation. Facsimiles of the "Be Counted" language questionnaires are included in Attachment C.

2. Detailed Information: Detailed information on translation, location, printing, processing, and so forth can be found in the Be Counted Program Master Plan.

#### D. Language Guides

1. Overview: Language guides are a positive complement to the Census 2000 language program. Guides will be user-friendly visual aids to assist respondents completing the Census 2000 questionnaires. Separate guides will be developed as reference for the Census 2000 short and long form mailout/mailback questionnaires. Guides are a useful, relatively low cost/low risk addition to the language program.
2. Languages: The Census Bureau developed a list of languages based on the 1990 Decennial Census of Population and Housing data of linguistically isolated households plus the most current immigration figures resulting in projected household populations for Census 2000. (Attachment D)  
Using that list as a baseline, the Census Bureau consulted with the Department of Commerce, the Director of Refugee Resettlement at the Department of Health and Human Services, and the Census Race and Ethnicity Advisory Committees to develop a final list of languages for Census 2000.  
Language assistance guides will be printed in the following languages.

Albanian	Amharic	Arabic	Armenian	Bengali
Burmese	Cambodian	Chamarro	Chinese	Creole
Croatian	Czech	Dari	Dinka	Dutch
Farsi	French	German	Greek	Hebrew
Hindi	Hmong	Hungarian	Ilocano	Italian
Japanese	Korean	Kurdish	Laotian	Polish
Portuguese	Roma	Romanian	Russian	Samoan
Serbian	Slovak	Somali	Spanish	Swahili
Tagalog	Thai	Tibetan	Tigrean	Tongan
Ukrainian	Urdu	Vietnamese	Yiddish	

The proposed language for language assistance guides in Puerto Rico will

be Haitian/Creole.

3. **Distribution:** The Census Bureau will furnish approximately 18 million language assistance guides to Questionnaire Assistance Centers, Local and Regional Census Offices, community groups and organization, partnership specialists, and Telephone Questionnaire Assistance staff. Language assistance guides will be distributed to Census partners in advance of Census Day. This will allow for broader dissemination and advertising in conjunction with community events. The current drafts of both the long and short form language assistance guides are included in Attachment E.
4. **Translation:** The Census Bureau will contract with an outside firm for the translation and validation of the language guides in the languages selected for Census 2000. Also, census regional census centers will work with census partners to translate and reproduce additional guides in languages localized to their specific communities. The Census Bureau also will recruit bilingual enumerators indigenous to the neighborhoods they will enumerate.
5. **Printing:** Unlike the questionnaires, language assistance guides can be developed and printed by census partners or community organizations in additional languages upon request in a very short time frame. Guides can be reproduced without compromising technical and data capture specifications.

#### **E. Telephone Questionnaire Assistance (TQA)**

1. **Overview:** The Census Bureau has identified potential constraints on the number of languages it can support through TQA. The current objective is to support the same languages for TQA as for "Be Counted" and the mailout/mailback questionnaires. The Bureau awarded the TQA contract to Electronic Data Systems (EDS), Herndon, VA on December 23, 1998. Census is in the process of finalizing specifications and determining whether additional efforts will be necessary to supplement the national program. For Puerto Rico, the TQA will support Spanish and English language calls.
2. **Responsibilities:** The volume of calls to the toll-free telephone assistance number in 1990 significantly exceeded all expectations. Six of the seven processing offices provided only English and Spanish telephone assistance. The San Diego Processing office also offered telephone assistance for English, Spanish, and six Asian languages. Given the Bureau's experience

during the 1990 census, the decision was made to rely on the professional call-center industry to provide telephone assistance for Census 2000.

3. Questionnaire availability: Foreign language questionnaires will not be available through TQA. A respondent request for a foreign language questionnaire by telephone will be handled in the following manner.
  - a. If the requestor **can supply** the Census ID from the original questionnaire, staff will forward the request to NPC in Jeffersonville, IN. The NPC staff will assemble a new mailing package with an English questionnaire corresponding to the original Census ID type and a language assistance guide of the requestor's choice. Staff will generate an address label with the correct address and Census ID barcode, address the new package, and mail to the respondent. Requests from Puerto Rico for English language questionnaires will be handled in the same manner.
  - b. If the requestor **can not supply** the Census ID, staff will obtain the address of the requestor. For house number/street name addresses, the information will be forwarded to the designated unit at the NPC. The NPC staff will follow a pre-designed 1-in-6 sampling scheme for stateside requests to determine which English form type (short/long) will be assembled with the language assistance guide of choice, and mailed to the requestor. Requests from Puerto Rico will be handled in the same manner. Staff will mail the Puerto Rico (Spanish) Update/Leave questionnaire corresponding to the 1-in-6 sampling scheme pre-designed for Puerto Rico with the language guide of choice or a Puerto Rico (English) Update/Leave questionnaire upon request. For addresses without a house number/street name, staff will conduct a reverse CATI interview that will use the stateside or Puerto Rico "Be Counted" questionnaire. However, this is to be conducted as a last resort procedure. All Puerto Rico telephone calls will be handled directly by a telephone operator.

#### F. **Education and Outreach**

1. Overview: The Census 2000 education and outreach plan is consistent with the overall Census objective to raise response rates and improve accuracy by increasing awareness of Census 2000 and educating the public. The plan will ensure that the Census Bureau reaches target audiences, especially those who learned a language other than English as their native language, with outreach and education materials in time to support the decennial programs. For each target audience, the plan identifies the group, the required materials, and the most effective means of conveying the decennial message.

Attachment F (Appendix A) includes an overview of all planned education and outreach products and the specific languages into which they will be translated.

2. Educational Materials: The content of Census 2000 educational materials will be determined by a review of the materials produced during the 1990 census, input from the Regional Offices, and other internal and external experts. The specifics of each document such as purpose, size, quantity printed, distribution, and the languages in which these documents are translated is available in the complete external communications plan for Census 2000 Publicity Office (C2PO). The author of each document will, whenever possible, be chosen based on knowledge of the topic area and will work closely with experts within the Census Bureau. The education and outreach materials planned for Census 2000 are broken down into the following four categories.

- a. Reference Materials: Reference materials include fact sheets, instructional manuals, and reference booklets. Most of these materials will be available in Spanish and many will be available in selected Asian languages for stateside distribution. These documents will be distributed primarily through Census headquarters and the regional offices or when they are aimed at easily accessible groups, they may be mailed directly to those groups or upon request to Census Advisory Committee members. These materials will be developed early in the census process and most will remain valuable after Census Day. Reference materials will be adapted as necessary and translated into Spanish for use in Puerto Rico (PR).

- b. Drop-in Articles and Newsletters: Drop-in articles will be developed for distribution through general news releases and targeted organizational newsletters and newspapers. The timing of these releases will be designed to support activities that are occurring in the field. These articles will be coordinated with the fact sheets and other publications to enhance each other and posted on the Internet for easy access by census partners. This Internet access enhances the language program availability to many diverse communities by allowing partners and/or the general public to translate census articles into any language of choice. A newsletter, following similar standards, will be developed to keep census partners informed. These materials will be adapted as necessary and translated for use in PR.

- c. Videos: Videos will be developed to support partnership activities and distributed to regional offices for loan to partners. Videos have a potential for reaching a large audience at a low cost. Current plans are for all videos to be produced in English only, but census partners may choose to copy

and adapt videos with voice overs in their language of choice.

d. Posters: A few posters will be developed to build awareness and support activities before the census. However, most posters will be developed to motivate immediate response to Census 2000. A general confidentiality poster to reassure respondents that answering the census is safe will be available in English, Spanish, selected Asian, and American Indian languages for stateside distribution. The posters in this plan are differentiated from posters covered by the ad contract because they will be placed in areas where the Census Bureau will not pay a fee. Posters will be distributed primarily through Census headquarters and the regional offices or when they are aimed at easily accessible groups, they may be mailed directly to those groups. Posters have the potential for reaching audiences that could be missed by traditional means. Posters will be developed that are appropriate for use in Puerto Rico and will be provided in English and Spanish.

3. Printing and Distribution: The number of translated, printed copies and their distribution strategy will be different for each document depending on the accessibility of the target audience and resources available. While the messages will be translated somewhat differently for each language group, the objective of increasing awareness of Census 2000 and educating the public about its benefits will be the same in each document. Every document will contain a statement about the benefits of Census 2000 and reassure the reader that participating in the census is safe. Documents which stress the benefits and confidentiality of the census are crucial. The following three major audience groups have been identified for distribution:

a. Census 2000 Partners: Documents written for our census partners will serve as aids for them to educate their constituencies about Census 2000 or specific instructions on how they can participate in programs. These documents are written for a somewhat more sophisticated audience -- people who already know something about the census and understand its importance. These materials will be distributed early to aide in supporting Census 2000 activities and partners may translate any/all documents at their discretion.

b. General Public: Materials written for the general public will contain the broadest, most general messages and will be the most widely distributed. Individual documents will answer questions commonly asked about the census or provide instructions on how to participate in Census 2000. The distribution of these materials will be given priority because their content is

applicable to the widest variety of constituencies. Some of these materials will be translated into the same languages as the mailout/mailback questionnaires. All materials are created for the general public and may be translated by communities, churches, civic groups, or individuals as they deem necessary.

c. Target Audiences: These materials will address specific issues of interest to target audiences and our census partners working with these audiences. Distribution will vary depending on the size and accessibility of the target groups. Translation of these materials will depend upon the target audiences involved.

Documents and materials for these three audience groups in Puerto Rico will be adapted for use on the island and made available in Spanish as well as in English.

#### 4. Implementation Strategy:

The implementation strategy for the external communications plan will support the goal of increasing response to Census 2000 by carefully laying out the timing of releases, coordinating the development of various materials to enhance related products, and reaching respondents in a variety of different languages.

### G. Advertising

1. Overview: Critical to the success of Census 2000 is the communication and marketing strategy, combining public awareness, promotion, and outreach activities. The public awareness campaign must use powerful and effective advertising messages to motivate each household to fill out and return the census questionnaires as soon after receipt as possible. This is especially critical within the "hard to count" populations where languages other than English are predominantly spoken.

In the past, the Census Bureau has used public service messages to inform the public and to motivate people to respond. The 1990 census relied solely on pro-bono public service advertising, which failed to reach many people. Television announcements, for example, were run in time periods, such as very late at night when audiences were small. In addition, the Census Bureau received virtually no prime-time coverage. For Census 2000, the Census Bureau is planning a more targeted approach, using paid advertising placed to have the greatest impact and promotions in languages other than English.

The goal of the advertising campaign is to increase public awareness - thus increasing initial mail returns of census questionnaires nationwide and in Puerto Rico from the general public, targeted audiences, and historically undercounted populations. The advertising contractor must be able to translate materials into numerous languages (and possibly some dialects) and have the resources to provide conceptual as opposed to literal translations. The creation of powerful and effective themes and messages that can be modified to fit the circumstances of local, regional, and targeted groups is an essential element of the Census 2000 advertising contract.

2. Young and Rubicam Coordination: The Census Bureau announced in October 1997 that it had awarded the Census 2000 paid advertising contract to Young & Rubicam, Inc. (Y&R) and a consortium of four partner agencies. The consortium partner agencies include:

- The Bravo Group - a Y&R firm specializing in Hispanic outreach
- Mosaica - a Y&R company expert in advertising to Asian audiences
- Chislom Mingo - a firm that targets African American audiences
- G & G - a Native American company

The development of the advertising plan for Puerto Rico will be the responsibility of Y&R's subsidiary on the island.

An important focus of the Census 2000 advertising plan is outreach to traditionally hard-to-reach audiences. Therefore, the Census Bureau has a strong minority contracting component to reach those audiences. Overall, 28 percent of the total Census 2000 advertising contracting dollars are earmarked for small disadvantaged firms. The Census Bureau is committed to a Census 2000 advertising plan which includes an aggressive buying and placement strategy to reach hard-to-count audiences including linguistically isolated households.

The Census Bureau's "Request for Proposal - Advertising Services for Census 2000 (section C)" is included as Attachment H. This section provides a synopsis of the Census Bureau's advertising campaign needs and expectations with an overview of the following topics and their relationship to the language program needs:

- Objectives
- National Campaign Strategy
- Scope of Work

Description of Tasks  
Performance Measurement

**III. Cost Assumptions**

- A. Overview: The revised Census 2000 Language Program is estimated to cost 6.9 million more than originally budgeted, increasing the Language Program budget from \$35.6 million to \$42.5 million. This budget allocation is not a complete reflection of the costs associated with the Census 2000 Language Program. The additional cost burden incurred by other Census 2000 program areas because of components driven by the Language Program (such as TQA, "Be Counted", Advertising, and Publicity) are reflected in their respective Program Master Plans.
- B. Components (Table/Attachment H)

**IV. Schedule/Responsibilities**

- A. Overview: Census day is scheduled for April 1, 2000 and is preceded by a multiple mailing and/or respondent contact. Additional operations, including nonresponse will be conducted by bilingual enumerators in communities where Census partners or RCCs identify the need and availability. The complete operations overview is available in the Census 2000 Master Activity Schedule. (relevant language program portions included in Attachment I)
- B. Schedule: The following tentative schedule has been developed for key phases of the Census 2000 Language Program.

Submit Census 2000 Clearance Package to OMB DSCMO/DMD/POP	June 30, 1998 Completed
Finalize Language Business Case Analysis DMD/DIR	July 31, 1998 Completed
Identify Potential Language Translation Contractors DSCMO/DMD	August 1, 1998 Completed
Issue Final TQA RFP DSCMO	August 14, 1998 Completed

Receive OMB Clearance for Census 2000	September 30, 1998 Completed
Start to Develop Procedures for Local Distribution of Language Materials FLD/DMD/DSCMO/C2PO	November 1998 Started
Award TQA Contract DSCMO	December 23, 1998 Completed
Advertise Language Print Contracts DSCMO/DMD	January 1999 Completed
Award Language Print Contract (Questionnaires) DSCMO	April 1999 Completed
Award Language Print Contract (Language Guides)  DMD	March 19, 1999 Completed
Provide Final Print Files to Contractor and Data Capture Program (Questionnaires and Language Guides)	April-August 1999

**V. Evaluation Requirements**

- A. Plans are currently not established concerning evaluation specifications resulting from the documentation of mail response rates for the Census 2000 Language Program.

**VI. Major Differences from the Census 2000 Dress Rehearsal**

- A. Dual Language Mailout: During the Census 2000 Dress Rehearsal targeted linguistically-isolated households were mailed both a questionnaire in English and a questionnaire in a second language. This operation was extremely time and cost intensive. In addition, the printing industry could not accommodate our request for an integrated, automated method to produce joint English/other language mailing packages with the correct household identifier on each questionnaire and maintain printing specifications for form tolerances. The creation of dual mailing packages with the same Census barcode identifier on both questionnaires became a manual clerical operation. Within the Census 2000 budget and time frame, operational objectives could not be successfully accomplished under this type of

scenario.

- B. **Questionnaire Special Requests:** To compensate for the shortcomings of the dual language mailout conducted during the Census 2000 Dress Rehearsal, the Census Bureau decided to offer all households which receive an advance letter the opportunity to request an alternative language questionnaire in one of 5 different languages. This option has the potential to benefit a larger component of the linguistically isolated community than a targeted mailing. The Census Bureau has doubled the number of mailout/mailback language questionnaires available to the public upon request compared to the Dress Rehearsal. Mailout/mailback questionnaires for the Dress Rehearsal were produced in English, Spanish, and Chinese - Census 2000 will have mailout/mailback questionnaires available in English, Spanish, Chinese, Korean, Vietnamese, and Tagalog.
- C. **Language Assistance Guides:** Providing "Be Counted" forms in multiple languages other than English created an incentive to use these forms as general purpose census "handouts" during the Dress Rehearsal. This would be operationally problematic for the full scale Census 2000. To curtail the use of "Be Counted" forms as questionnaire translations, the Census Bureau will create and distribute approximately 15 million language assistance guides in more than 30 different languages to assist respondents completing their English questionnaires.
- D. **QAC Paid Temporary Staff:** Local census partners had difficulty identifying volunteers to run some Questionnaire Assistance Centers (QACs) during the Dress Rehearsal. The Census Bureau therefore created 15,000 paid temporary staff positions for Census 2000 to supplement volunteers running QACs in neighborhoods that would benefit most.
- E. **Large Household Follow-up:** With the introduction of a six person form for Census 2000, plans to conduct a mailout/mailback Large Household Follow-up operation were canceled. Further information can be obtained in the Census 2000 Decision Memorandum Series.
- F. **Processing:** All "Be Counted" forms and all Asian language questionnaires will be returned to the National Processing Center. They will be scanned in upon receipt, then all information will be translated and transcribed onto the respective English questionnaire prior to image capture. Spanish language questionnaires generated from advance letter requests will be returned to the Pomona Data Capture Center (DCC) for check-in, translation, transcription, and data capture.
- G. **TQA Special Mail Requests:** Stateside callers requesting questionnaires in languages other than English through TQA will receive the language assistance

guide of their choice. In Puerto Rico, callers requesting questionnaires other than the Puerto Rico Spanish form will be offered the option of an English Puerto Rico form or the PR Spanish form with a language guide.



February 9, 2000

CENSUS 2000 INFORMATIONAL MEMORANDUM NO. 39

MEMORANDUM FOR      Distribution List

FROM:                    Susan M. Miskura *Susan Miskura*  
                              Chief, Decennial Management Division

Subject:                Program Master Plan: Census 2000 Long Form Sampling  
                              Operation

Contact Person:        Kathleen Stoner, Statistical Programs Branch, Decennial  
                              Management Division, Room 1422-2, (301) 457-8223

The Program Master Plan (PMP) for the Census 2000 Long Form Sampling Operation is attached.

Please forward any inquiries to the contact person for this PMP, Kathleen Stoner.

Attachment

**Census 2000 Long Form Sampling**  
**TABLE OF CONTENTS**

<u>Section</u>	<u>Page</u>
<b>I. INTRODUCTION</b>	<b>1</b>
<b>II. GENERAL DESIGN AND WORKFLOW</b>	<b>1</b>
<b>A. Inputs</b>	<b>2</b>
<b>B. Process</b>	<b>3</b>
<b>C. Outputs/Deliverables</b>	<b>8</b>
<b>III. COST ASSUMPTIONS</b>	<b>9</b>
<b>IV. SCHEDULE AND DIVISION RESPONSIBILITIES</b>	<b>9</b>
<b>V. EVALUATION REQUIREMENTS</b>	<b>11</b>
<b>VI. MAJOR DIFFERENCES FROM DRESS REHEARSAL</b>	<b>11</b>
<b>VII. PUERTO RICO</b>	<b>11</b>
<b>VIII. REFERENCES</b>	<b>12</b>

**Attachment A - List of Abbreviations**

**Attachment B - Flowchart**

**Attachment C - Initial Long Form Sampling: Summary Results**

## **I. INTRODUCTION**

The Decennial Management Division (DMD) is responsible for describing and defining the requirements, schedules, and workflow for each Census 2000 operation. This PMP describes the preparatory and operational activities required to conduct the Long Form Sampling operation for Census 2000 in the U.S. and Puerto Rico.

The major objective of this operation is to determine which households will receive the long-form census questionnaire, and to produce a file by which these households can be identified for the mailing of the questionnaires. For areas in which no mailout is done, the objective is to correctly sample and obtain the long form data. The goal of this operation is to obtain long form sample data for approximately 17% of all addresses in the census. The data collected from the long form sample are used to estimate various demographic and socioeconomic characteristics for the U.S. population.

The major activities for the long form sampling operation have been completed. These activities include refinement of long form sampling methodology, preparation of long form sampling specifications, review and approval of long form sampling output, and the preparation of sample tolerance check (STC) specifications for the list/enumerate (L/E) areas. Section VIII of this document contains a list of references used in creating this PMP. The weighting and variance estimation procedures are currently being prepared, so a description of these will be included in a separate PMP.

The following divisions are responsible for the planning and/or implementation: the GEO, the DSCMO, the DSSD, the DSMD, and the DMD.

## **II. GENERAL DESIGN AND WORKFLOW**

In mailout/mailback and update/leave enumeration areas where questionnaires are mailed back, long form sample selection is conducted before questionnaire mailout (or delivery). A sampling file is created from the Decennial Master Address File (DMAF). Input on governmental units and size and occupancy rates from other files is applied and measures of size for each block are calculated. The estimated size of each block is used to determine the size of all associated LFSEs and census tract, which then determine the sampling rate for the block.

For list/enumerate areas, where questionnaires are not mailed, special place and group quarters enumeration, and adds in mailout/mailback and update/leave areas, long form sampling is done in the field by enumerators following established procedures for non-mailout operations. This includes sampling done by following patterns in pre-numbered listing books, and on-the-spot sampling for various military and service-based enumerations. The various operations and corresponding sampling methods are detailed in Section II.B.3.

The list of addresses that will receive the long form questionnaire will generally be

chosen by taking a systematic, variable rate sample of addresses. Sampling rates will be applied to each census block. Determination of the sampling rate for a block is based on the measure of size (MOS) for the long form sampling entity (LFSE) or census tract that the block is in. The target sampling rates (TSRs) that will be used are 1-in-2, 1-in-4, 1-in-6, and 1-in-8.

The sampling strata and their cutoff points are:

- 1-in-2 for LFSEs < 800 addresses
- 1-in-4 for LFSEs between 800 and 1,200 addresses; and if not 1-in-2 or 1-in-4; then
- 1-in-6 for census tracts < 2,000 addresses; and
- 1-in-8 for census tracts  $\geq$  2,000 addresses.

For blocks that fall into more than one sampling stratum, the higher sampling rate will be applied.

The following are treated as LFSEs for sampling purposes: states, counties, cities, school districts, incorporated places, Minor Civil Divisions (selected states only), American Indian Reservations (AIRs), Tribal Jurisdiction Statistical Areas (TJSAs), Alaska Native Village Statistical Areas (ANVSAs), and census designated places (Hawaii only). Trust Lands will follow the designation of their associated reservation.

Census tracts are the interim census tracts defined for Census 2000 data collection purposes.

For operational convenience, neighboring blocks may be combined to create Assignment Areas in Update/Leave, Update/Enumerate, and List/Enumerate areas. In Assignment Areas that have blocks with different sampling rates, adds obtained through the field updating will be sampled at the higher rate. Due to operational constraints, List/Enumerate Assignment Areas initially assigned 1-in-6 or 1-in-8 sampling rates will be sampled at a 1-in-6 rate, while those initially assigned 1-in-2 or 1-in-4 rates will be sampled at a 1-in-2 rate.

#### **A. Inputs**

##### **1. Master Address File (MAF) Extract**

The GEO prepares this extract according to specifications written by the DSSD and delivers it to the DSCMO for use in decennial census operations.

##### **2. Decennial Master Address File (DMAF)**

The DMAF is the address file (prepared by the DSCMO) used in all decennial census operations and includes addresses from the MAF extract. The DMAF is created by adding various operational codes, counts, and variables to the initial MAF extract.

3. Geographic Reference File - Codes (GRF-C)

The GRF-C is a file containing collection and current geographic codes for each collection block.

4. Geographic Reference File - Names (GRF-N)

The GRF-N is a file containing name and attributes for geographic entities.

5. School District Equivalency File (SDEQ)

The SDEQ lists all school districts containing each 2000 collection block.

6. Accuracy and Coverage Evaluation (A.C.E.) Universe Files

Long form sampling uses the A.C.E. universe files to obtain the 1990 occupied and total HU counts and 1990 American Indian population and total population counts for each 2000 collection block.

**B. Process**

1. MAF Extract Preparation

a. Identification of Long Form Sampling Entities

The MAF extract includes all places, but in preparation for long form sampling, the GRF-N is used to identify which places are long form sampling entities. (Minor civil divisions (MCDs) are treated as long form sampling entities in states that contain them, and census designated places (CDPs) are used as long form sampling entities in Hawaii.) Then the SDEQ is used to add school district codes to each address record. Other flags are assigned to the address record according to whether it is in an American Indian Reservation (AIR), Tribal Jurisdiction Statistical Area (TJSA), or Alaska Native Village Statistical Area (ANVSA).

b. Calculation of Measure of Size

The basis for all measures of size is an address count. For most areas, this will be based on a count of the addresses in the July 1999 Master Address File (MAF) extract. For areas that do not have current counts available, 1990 housing unit counts will be used. These include the List/Enumerate and Remote Alaska types of enumeration areas.

Although address counts are a good starting basis for measures of size, they cannot be used directly. Since the primary interest for long form estimation is the reliability of person level estimates, we would ideally like to use population counts as our measure of size. However, unlike address counts, updated population counts are not available. Therefore, we use the address counts as a proxy for the population count. A risk associated with this approach is that if an area has a high vacancy rate, it may be assigned a lower sampling rate than it would were the size based on the population size. To compensate for this, we will apply the proportion of 1990 occupied housing units to the MAF extract counts in determining the measure of size for an area.

Using the A.C.E. universe files and the MAF extract, the measure of size components and the MOS will be calculated for each LFSE and interim census tract. Each 2000 collection block will be assigned a governmental unit MOS that corresponds to the smallest MOS of all LFSEs in which the block is located and a tract MOS that corresponds to the MOS for the interim census tract in which the block is located.

The 2000 A.C. E. universe files and block statistics collected during the 1990 Census are used to calculate estimated block sizes. The variables used are:

1. Estimated number of American Indian and Alaskan Native renters in the 2000 collection block in 1990
2. Estimated number of American Indian and Alaskan Native owners in the 2000 collection block in 1990
3. Estimated number of 1990 housing units in the 2000 collection block

4. Estimated number of 1990 occupied housing units in the 2000 collection block
5. Estimated 1990 total non-group quarters population in the 2000 collection block

These variables are used to calculate estimates of 1990 occupancy rates, the 1990 American Indian/Alaska Native population, and the 1990 total population. An algorithm using all these estimates is employed, resulting in a measure of size estimate for each block. **The sampling rate for each block is determined by the measure of size of its associated LFSEs and census tract.**

## 2. Sample Identification (HQ)

The primary source of addresses for Census 2000 is the DMAF. Addresses added to the DMAF after its initial creation for inclusion in the mailback universe will be sampled separately using the sampling rate already determined for each block.

### a. Assignment to Sampling Strata

Each collection block in the 50 states, the District of Columbia, and Puerto Rico will be assigned to a sampling stratum and assigned a target sampling rate based on the block's governmental unit or tract MOS. A block will be assigned to only one sampling stratum and therefore only one TSR will be associated with each collection block. In general, the smaller the block, the higher the sampling rate for that block.

### b. DMAF Frame Address Sampling

For blocks not in list/enumerate or remote Alaska areas, a random start, systematic sampling procedure will be used to select the sample. Additional addresses may be added after initial mailout, and they will be sampled at the same rate as the rest of the block.

## 3. Sample Identification (FLD - LCOs)

- a. Adds in Update/Leave, Update/Enumerate, Urban Update/Leave, Urban Update/Enumerate Areas, and adds picked up in NRFU for any of these operations

Addresses may be added during any of these operations. In these cases, the sampling rate for the assignment area (AA) will be used; we will not use the sampling rates for individual blocks. The AA sampling rate will be the highest sampling rate that occurs in any block in that assignment area. For example, if the 1-in-4 sampling rate and the 1-in-6 sampling rate are the only ones used in the assignment area, the higher sampling rate, 1-in-4, will be used for added addresses.

b. L/E and Remote Alaska

The address listing books for these areas will be preprinted with the appropriate sampling patterns.

c. Group Quarters

A systematic sample of persons enumerated in group quarters (GQs) will be selected for the long form sample. A fixed 1-in-6 sampling rate will be used in GQs.

d. Service Sites (Shelters and Soup Kitchens)

Enumerators will administer the questionnaire to the populations found in these service sites. A systematic 1-in-6 sample will be selected at the time of enumeration.

e. Mobile Food Vans and Targeted Nonsheltered Outdoor Locations Enumeration (TNSOLs)

Clients of mobile food vans will be interviewed using the D-15A short form ICQ (Individual Census Questionnaire) only. Individuals enumerated at TNSOLs will be interviewed using the D-20A ICR (Individual Census Report). No long form interviews will be conducted at either type of location.

f. Telephone Questionnaire Assistance (TQA)

The TQA operation has two main functions: to mail questionnaires to those who call to request them, and to take interviews over the phone from respondents who wish to provide them this way.

Respondents who phone to request that a form be mailed will receive either their designated form type (if they have their census

ID number) or be subject to a 1-in-6 sampling rate (if they do not have the census ID number).

Respondents who wish to provide an interview by phone are not eligible for the long form. The short form is always used for those cases.

g. **USPS (United States Postal Service) Updated Address File Delivery**

In February 2000, an updated address file from the USPS will be delivered. This file will include addresses which were added by the USPS since the delivery of the November Delivery Sequence File. Added addresses which are geocoded in time will be included in the February 28, 2000 DMAF and will be sampled by continuing the appropriate sampling pattern for the block in which they are located.

h. **Military Bases and Maritime Vessels**

People living in housing units at military installations receive questionnaires by mail and are sampled using conventional procedures. Others that are enumerated at their work stations (those living in barracks or aboard maritime vessels) with Military Census Reports (MCRs) or Shipboard Census Reports (SCRs) will be sampled at a 1-in-6 rate, using the last 4 digits of the respondent's Social Security number.

i. **Quality Assurance - Sample Tolerance Check**

This procedure is implemented in L/E areas to detect bias in the long form sample stemming from enumerator error. Enumerator bias in L/E AAs arises when enumerators ignore the systematic sampling procedure and administer the long form questionnaire to a disproportionate number of vacant housing units (HUs), or to HUs containing predominately small or predominantly large families, thus adding a bias to the long form sample.

In general, the procedure consists of identifying areas with a statistically significant difference between a simplified long form estimate of the population and the population count obtained from the initial phase of the census enumeration. The areas for this operation are defined as two consecutive assignment areas (AAs)

completed by a single enumerator. Batches of consecutive AAs will be created by software run after all L/E areas for an LCO are complete. The LCO staff will capture some basic data from the listing books and run the STC computer software developed by the Technologic Management Office (TMO). If a batch fails the STC, both AAs within the batch will be resampled independently and returned to the field for additional interviewing during the Coverage Improvement Follow-Up (CIFU) operation. In some instances, AAs that pass the STC, but are sampled at a lower rate than designated will be subject to a supplemental sampling of addresses. Refer to the List/Enumerate PMP for details of this sampling operation.

### C. Output/Deliverables

All of the files and counts listed below will be supplied to the DSSD by the DSCMO. They will be supplied after initial sample selection, and throughout the long form sampling process at specified intervals. The main purpose for supplying these files to DSSD is so that DSSD staff may evaluate the sampling results (to make sure the long form sample is of sufficient size, and that the correct sampling rate has been applied to each block) for quality assurance purposes.

#### 1. Block level file with sampling results.

The DSSD will supply a suggested record layout. This file will contain information about the governmental unit code, unit type, functional status, and measures of size for each governmental unit. It will also include information about the random start, take-every, probability of selection, number of selected units, sampling rate, and so on.

#### 2. Listing of specified address records.

This listing must include address records from the long form sampling file, including all variables with appropriate geographic identifiers, as specified by the DSSD. This file is to be delivered after the completion of sampling for each state, and will include a sample of address records to be specified by the DSSD.

#### 3. Specified governmental unit (GU) counts

This listing must include counts of sampled and non-sampled addresses organized by GU, separated by sampling rate. Other pertinent information for this file will be specified by the DSSD.

4. Specified census tract counts

This listing must include counts of sampled and non-sampled addresses organized by census tract, separated by sampling rate. Other pertinent information for this file will be specified by the DSSD.

5. Long form sampling universe (for unduplication with American Community Survey sample)

Fifty-one files (one for each state, one for Washington, DC) will be supplied to the DSMD. (Puerto Rico is not in the ACS sample for 2000). They will list the long form sample universe and the addresses designated to be in sample. American Community Survey (ACS) questionnaires will not be sent to addresses in the long form sample.

6. Specified address counts for American Indian Reservations (AIR), Tribal Jurisdiction Statistical Areas (TJSA), or Alaska Native Village Statistical Areas (ANVSA).

This file is only created for states that contain any part of an AIR, TJSA, or ANVSA. It includes the same information as shown for 2. above, but organized by appropriate AIR, TJSA, or ANVSA.

A table showing a summary of the sampling rates for each state is shown in Attachment C.

### III. Cost Assumptions

There are no direct costs associated with this operation; all functions are performed by headquarters personnel or by field personnel as part of other operations.

### IV. Schedule and Division Responsibilities

#### A. Schedule

The following is the schedule of major activities for the long form sampling operation as reflected in the master activity schedule (MAS).

	<u>Start</u>	<u>Finish</u>
Refine long form sampling methodology	01/05/98	04/01/99
Prepare long form sampling specifications	02/16/99	04/20/99
Review and approve long form sampling output	07/07/99	08/25/99
Prepare sample tolerance specifications (L/E)	04/13/98	04/23/99

**B. Division Responsibilities**

The following is a list of responsibilities, by division, for this operation.

**1. Geography Division (GEO)**

The GEO will:

- a. Deliver the MAF to the DSCMO.
- b. Deliver the GRF-N, GRF-C, and the SDEQ files to the DSCMO.

**2. Decennial Management Division (DMD)**

The DMD will:

- a. Document program requirements in the Long Form Sampling Program Master Plan (PMP).
- b. Maintain the Master Activity Schedule (MAS).
- c. Monitor progress of the operation.

**3. Decennial Systems and Contracts Management Office (DSCMO)**

The DSCMO will:

- a. Provide the DMAF.
- b. Deliver all specified listings (sampling results and other output) to the DSSD.
- c. Provide the long form sample file to DSMD for unduplication.

**4. Decennial Statistical Studies Division (DSSD)**

The DSSD will:

- a. Provide long form sampling specifications.
- b. Provide the A.C.E. universe file to the DSCMO for use in preparing occupancy rates.
- c. Perform operational analyses of long form sampling.

**5. Field Division (FLD) - Headquarters**

The FLD will:

- a. Develop (either in-house or via contractor) all training materials, office and field manuals, and enumerator materials. Long form sampling will be included in these manuals as an integral part of various data collection operations.

**6. Field Division - LCOs**

The FLD LCOs will:

- a. Perform long-form sampling as an integral part of various data collection operations.

**7. Demographic Statistical Methods Division (DSMD)**

The DSMD will:

- a. Perform an unduplication of the American Community Survey (ACS) files using the long form sample file supplied by the DSCMO.

**V. EVALUATION REQUIREMENTS**

Evaluation requirements are undefined at this time.

**VI. MAJOR DIFFERENCES FROM DRESS REHEARSAL**

In dress rehearsal (DR), if a respondent called into telephone questionnaire assistance (TQA) and wanted to give a phone interview, that respondent was sampled for long form. This will not be the case in the 2000 operation; only a short form interview will be conducted.

In dress rehearsal, the sample included an extra ten percent to cover sample loss. This will not be the case in 2000.

## **VII. PUERTO RICO**

All blocks in Puerto Rico will be sampled at a rate of 1-in-6. No variable sampling rate will be used.

The Puerto Rico and Island Areas Branch in DMD is responsible for adapting and translating field and office manuals and training materials. Long form sampling will be included in these manuals as an integral part of various data collection operations.

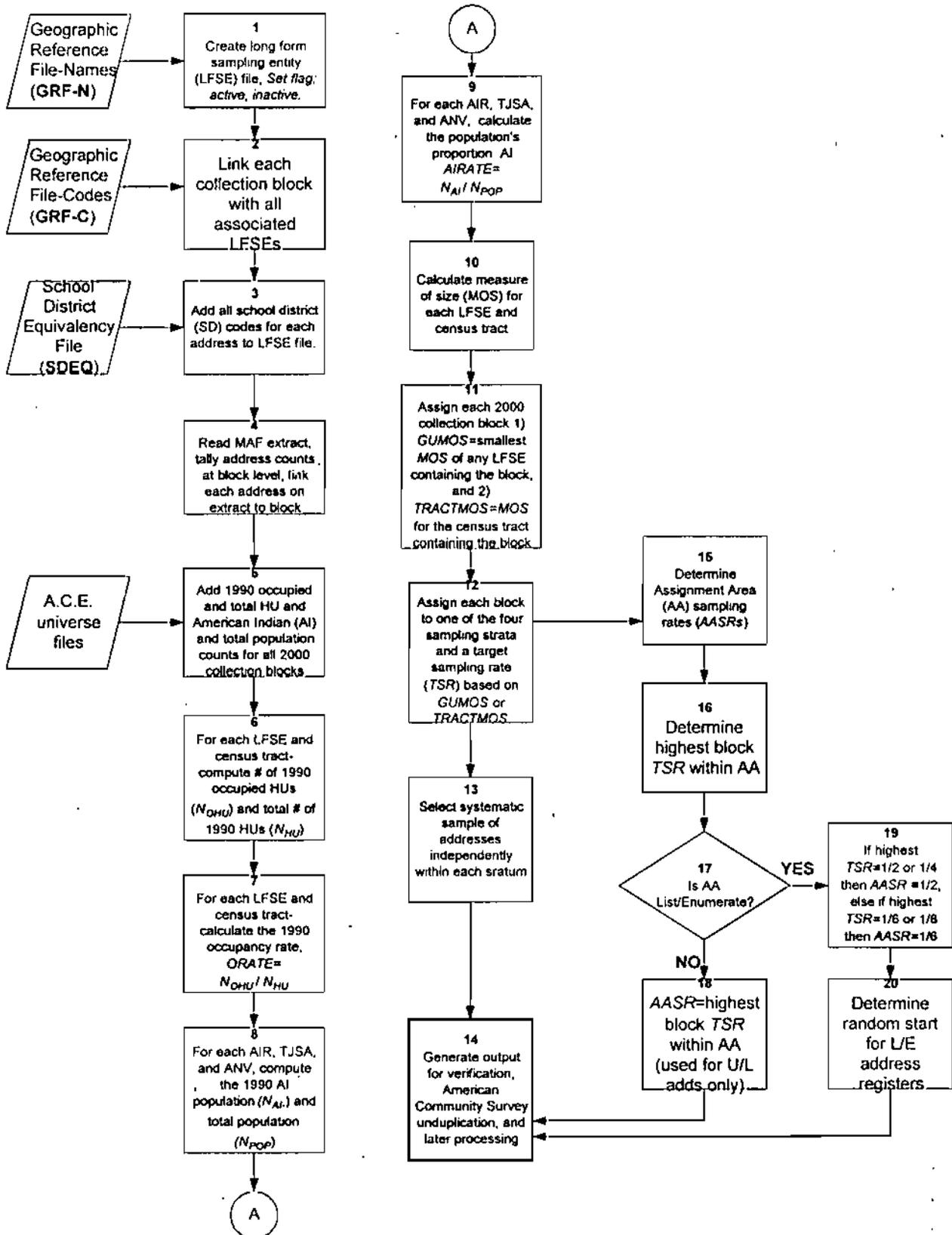
## **VIII. REFERENCES**

1. DSSD Census 2000 Procedures and Operations Memorandum Series LL-1, "Sample Tolerance Check Specifications for Census 2000," April 23, 1999.
2. DSSD Census 2000 Procedures and Operations Memorandum Series LL-2, "Projected Long Form Sample Size for Census 2000," May 17, 1999.
3. DSSD Census 2000 Procedures and Operations Memorandum Series LL-3, "Approval of the Sample Tolerance Check Software for Census 2000," June 15, 1999.
4. DSSD Census 2000 Procedures and Operations Memorandum Series LL-4, "Long Form Conversion Processing for the Sample Tolerance Check in Census 2000," August 5, 1999.
5. DSSD Census 2000 Procedures and Operations Memorandum Series LL-5, "Long Form Sampling Specifications for Census 2000," November 17, 1999.
6. DSSD Census 2000 Procedures and Operations Memorandum Series LL- , "Requirements for Measures of Size to Assign Long Form Sampling Rates," DRAFT June 21, 1999.
7. DSSD Census 2000 Procedures and Operations Memorandum Series LL- , "Census 2000 Long Form Sample Design," DRAFT February 26, 1999.
8. DMD Census 2000 Informational Memorandum No. , "Requirements for Assigning Long Form Sampling Rates in Update/Leave and Rural Update/Enumerate Areas," DRAFT May 27, 1999.

## LIST OF ABBREVIATIONS

A.C.E.	Accuracy and Coverage Evaluation
AA	Assignment Area
AASR	Assignment Area Sampling Rate
ACS	American Community Survey
AIR	American Indian Reservation
ANVSA	Alaska Native Village Statistical Area
CDP	Census Designated Place
DMAF	Decennial Master Address File
DMD	Decennial Management Division
DR	Dress Rehearsal
DSCMO	Decennial Systems and Contract Management Office
DSMD	Demographic Survey Methods Division
DSSD	Decennial Statistical Support Division
GEO	Geography Division
GQ	Group Quarters
GRF-C	Geographic Reference File - Codes
GRF-N	Geographic Reference File - Names
GU	Governmental Unit
GUMOS	Governmental Unit Measure of Size
HU	Housing Unit
L/E	List/Enumerate
LFSE	Long Form Sampling Entity
MAF	Master Address File
MAS	Master Activity Schedule
MCD	Minor Civil Division
MOS	Measure of Size
ORATE	Occupancy Rate
PR	Puerto Rico
SDEQ	School District Equivalency File
STC	Sample Tolerance Check
TJSA	Tribal Jurisdiction Statistical Areas
TQA	Telephone Questionnaire Assistance
TRACTMOS	Tract Measure of Size
TSR	Target Sampling Rate
U/L	Update/Leave

### Census 2000 Long Form Sampling Procedure



Initial Long Form Sampling: Summary Results

State	Total	Designated	Percent Designated
Alabama	2,042,456	354,931	17.4
Alaska	237,341	45,147	19.0
Arizona	2,250,181	337,897	15.0
Arkansas	1,189,101	246,904	20.8
California	12,730,706	1,840,968	14.5
Colorado	1,862,996	310,878	16.7
Connecticut	1,458,730	221,245	15.2
Delaware	355,589	60,422	17.0
DC	290,082	44,797	15.4
Florida	7,603,796	1,058,752	13.9
Georgia	3,490,749	540,378	15.5
Hawaii	521,789	87,504	16.8
Idaho	542,042	107,798	19.9
Illinois	5,360,400	902,463	16.8
Indiana	2,680,879	452,110	16.9
Iowa	1,263,962	283,327	22.4
Kansas	1,162,922	230,677	19.8
Kentucky	1,797,696	316,796	17.6
Louisiana	1,949,648	332,473	17.1
Maine	583,096	131,581	22.6
Maryland	2,228,934	339,484	15.2
Massachusetts	2,747,639	406,602	14.8
Michigan	4,404,002	901,739	20.5
Minnesota	2,130,213	524,719	24.6
Mississippi	1,181,905	210,171	17.8
Missouri	2,538,736	492,756	19.4
Montana	400,731	99,478	24.8
Nebraska	732,960	174,934	23.9
Nevada	805,053	108,527	13.5
New Hampshire	513,173	93,951	18.3
New Jersey	3,460,295	534,821	15.5
New Mexico	797,016	137,793	17.3
New York	8,111,938	1,355,831	16.7
North Carolina	3,582,596	615,768	17.2
North Dakota	293,300	79,295	27.0
Ohio	4,947,406	831,921	16.8
Oklahoma	1,545,119	344,302	22.3

Oregon	1,514,247	248,986	16.4
Pennsylvania	5,541,265	1,079,083	19.5
Rhode Island	464,467	68,147	14.7
South Carolina	1,814,125	305,123	16.8
South Dakota	321,038	83,463	26.0
Tennessee	2,501,196	393,823	15.7
Texas	8,308,816	1,390,311	16.7
Utah	797,383	139,361	17.5
Vermont	258,433	70,327	27.2
Virginia	2,941,894	444,856	15.1
Washington	2,578,680	410,333	15.9
West Virginia	859,037	163,014	19.0
Wisconsin	2,400,119	599,976	25.0
Wyoming	194,797	38,253	19.6
<b>U.S. Total</b>	<b>120,290,674</b>	<b>20,594,196</b>	<b>17.1</b>
Puerto Rico	1,357,202	226,200	16.7
<b>Total</b>	<b>121,647,876</b>	<b>20,820,396</b>	<b>17.1</b>