

## **The United States Census 2000 -- High Tech: High Expectations** (For presentation to the 18th Population Census Conference)

In reviewing and evaluating the results of the 1990 Census and preparing for the Year 2000 Census, we established three major goals:

- ◆ to count and profile the people and housing units of the United States and associated Island Areas as completely and accurately as possible
  
- ◆ to reduce or eliminate the historic pattern of accounting for different components of the population at unequal rates, and
  
- ◆ to do so at a cost the nation can afford.

A significant portion of the costs associated with the Year 2000 Census are those associated with technology. But we believe these costs will be well worth it in terms of the operations that will be made more efficient and effective. We are always looking at ways in which we can improve our speed and quality and limit the number of manual, repetitive, labor-intensive operations.

Census 2000 will be a highly computerized census, from data collection to releasing data on the Internet. And, in a sense, it will be the most accessible of the 21 censuses of population and housing ever, with an array of new techniques making it possible for an unprecedented number of people to construct a digital picture of America down to any one of seven million census blocks.

Today, I'm going to highlight just six major areas of **technological** improvements that we believe will allow us to meet the primary goals I mentioned earlier. Given the time and presentation limitations of this paper, I cannot include every advance in technology that we will be adopting for the census such as our plan to provide some field staff with laptop computers to facilitate the data collection and transmission of data. I have chosen to identify six major program or system areas that will be significantly affected by the implementation of advanced technology. The six areas include technological advances in:

- ◆ the Master Address File geographic program
  
- ◆ the payroll and hiring systems

- ◆ the telephone questionnaire assistance program
  
- ◆ the data capture system
  
- ◆ the computerized matching software system, and
  
- ◆ the data access and dissemination system

### **Master Address File Geographic Program**

To conduct the census, we must identify and locate an estimated 118 million housing units and special places like hospitals, college dormitories, and military barracks.

Therefore, one of our highest priorities this decade was the creation of a Master Address File (MAF) of housing units in the U. S. The MAF started with the 1990 list, but it will be greatly enhanced because it will be matched against the United States Postal Service (USPS) address list. This will be an improvement over the 1990 MAF as we will not be using private vendor lists as we did in 1990 when we found they were less accurate in low income areas. This was due to the fact that many of the 1990 lists were originally generated for marketing purposes.

Partnerships with state and local officials, community organizations, and tribal governments will also play an important role in making sure the MAF is accurate. And, of course, our enumerators will update the list as they work out in the field.

For the MAF to be useable in automated census taking operations, the TIGER (Topologically Integrated Geographic Encoding and Referencing) data base that we used in 1990 must be enhanced with new street features, street names, ZIP codes, and address ranges. Making these improvements to TIGER allows the automated linking of the MAF and TIGER based on mailing address.

At this time, and with current resources, we are focusing on improving the street name/address range/ZIP Code coverage in TIGER. Our efforts in developing a general digital exchange system has taken the approach of concentrating on road features and their attributes. Our efforts have evolved from developing a general digital exchange system to creating an automated system for transferring new road features and postal geography. The present system only attempts to match the TIGER roads to the roads in the spatial data base. The output from the match, together with the results from geocoding the clustered USPS's Delivery Sequence File of all mail delivery points against TIGER and against the local file, are used to

determine what updates to apply to TIGER. We have successfully implemented this system when local files have met our standards.

To use a file in our automated system, we convert it from an exchange format into a data base system. We have chosen the TIGER data base structure to take advantage of existing TIGER system software to process the file.

We also have made other great strides in the geographic program. For this census, we have the ability to do map imaging from each of our address listing operations to produce an updated set of maps using scanning equipment. Between the end of July and mid-December of this year, we will produce approximately three million maps and list about twenty-two million addresses.

We are greatly expanding the number of map sheets and map types to meet the needs of both internal and external users. Our major goals are to have the ability to deliver our user maps faster, to improve map quality over the 1990 census, and to increase device and storage options.

For this census, we will be distributing map production to the Regional Census

Centers and Local Census Offices. Staff will have the capability to specify the type of map they need for a particular operation and to include only those features that are required for that particular map request. For example, an important map for field operations is one that delineates an enumerator's assignment area. Our system will allow us to program into the computer, the area that the enumerator will be working in and identify only those map contents that are of use to an enumerator such as street names, neighborhood names, bodies of water, and the like.

Also new for this census is our Intranet map capability. This gives our employee users the ability to submit map requests locally and allows them to check on the status of the maps.

### **Payroll and Hiring Systems**

In 1995 we took a long, hard look at our personnel and payroll system that will be used to pay more than 350,000 temporary employees of the Year 2000 census. Our challenge was to decrease inconsistencies and redundancies in administrative tasks while maintaining the integrity and confidentiality of the data. This approach resulted in a system called the Pre-appointment Management System and Automated

Decennial Management System or PAMS/ADAMS.

We have developed an integrated approach to the management systems for the Decennial Census. The PAMS/ADAMS system is a part of this integrated approach that provides the framework for field collection control activities at the Regional Census Centers and the Local Census Offices. Our integrated approach, or Decennial Field Interface (DFI), consists of seven principal components: Operations Control, Office Support Systems, PAMS/ADAMS, Geography Systems, Integrated Coverage Measurement Control, Administrative Control Systems, and the Management Information System. This approach uses state-of-the-art technology in an open system environment. The result will be a product that is simple enough for both our temporary and permanent employee population to maintain and support in the field.

PAMS processes job applications, tracks applicant status, assists applicant selection, processes FBI checks (allows for criminal history checks), and shares relevant information with ADAMS. PAMS processes applicants through ranking, selection, and hiring, and then provides this applicant data to the ADAMS systems.

ADAMS processes personnel actions, payroll, expense data, and, very importantly, progress and management reporting activities as a single entity. ADAMS receives the applicant data and sends personnel actions to the National Finance Center that is responsible for calculating salaries and transmitting data to the Department of Treasury for actual check issuance.

Throughout the census, this system will be expected to process approximately 1.5 million applicants, support the selection of the temporary employees needed to support Decennial operations, capture time and expense data daily, pay temporary employees weekly, and terminate temporary employees at the end of their service. Along with this, it will provide a variety of management reports and maintain historical data that will be very useful in planning for the Year 2010 Census.

### **Telephone Questionnaire Assistance Program (TQA)**

In the 1990 Census, temporary staffs provided paper-based questionnaire assistance to the public by telephone and through walk-in assistance centers. Technological advances continue to broaden the ways for furnishing such help and for facilitating the reporting of data. Our plan, which we replicated in our Dress Rehearsal during the spring of this year, is to acquire and implement a fully integrated nationwide network of call centers capable of supporting high-volume, limited duration inbound TQA calls from the general public. We also plan to make outbound follow-up calls to residential telephone numbers provided on census questionnaires. Our plan is not only to provide questionnaire assistance, but also to allow persons to provide their census responses.

We just conducted a vendor conference at Census headquarters for members of the call-center industry interested in Census 2000 plans for contractor-provided TQA. The call centers will combine 24-hour-a-day automated interactive voice response technology, featuring menus in at least two languages (English and Spanish), with live-operator services, between 8:00 a.m. and 9:00 p.m. (in all time zones), from March 3 to June 8, 2000.

Based on 1990 Census and 1995 Test Census data, we estimate that the TQA

program will handle an estimated nine to eleven million inbound calls. Inbound callers will be able to select from an automated menu that will supply answers to the most common questions such as, "How will the data be used?" and "How will you protect the confidentiality of my answers?" Some calls will require the assistance of operators who will be trained in customer awareness techniques, census concepts, how to complete computer- assisted census interviews via telephone, and confidentiality. In addition, the staff will perform at least one, and possibly more, outbound call operations to reconcile discrepancies or obtain missing information from mailed-back questionnaires. Both the contractor's managers and operators working in the call centers will be required to take a Census Bureau "oath of non-disclosure" of the data they collect. The service, which will cover all 50 states, the District of Columbia and Puerto Rico, will use the Federal Telephone System and include a TDD line for the hearing-impaired.

## **Data Capture System**

We have contracted with Lockheed-Martin Mission Systems (Lockheed-Martin) to help us use the best available data capture methodology for Census 2000 that allows us to design a more respondent friendly questionnaire and to minimize costly customization of hardware and software. Our system is called Data Capture 2000 (DCS 2000) . We have successfully tested the hardware and software that converts handwriting on the questionnaire into computer form with minor editing by technicians. The system will use off- the-shelf hardware and software to record a full electronic image of every questionnaire. We will sort mail-return questionnaires automatically to ensure timely conversion and capture of critical information needed prior to nonresponse follow-up. We also will use Optical Mark Recognition (OMR) to capture check-box items, Intelligent Character Recognition (ICR) and key-from-image to capture write-in character based data items. The images will be processed by software capable of recognizing an infinite variety of hand strokes as either alphabetic or numeric characters.

While scanners have been used for decades to recognize the marks made in answering standardized forms, the strokes that were read by the scanners were

confined to defined ovals or boxes, with the markings restricted to Number 2 lead pencils. By contrast, the optical scanners that will be used in this census will recognize and decipher hand-written responses made by both pens and pencils. After the characters are translated into computer code, the respondent data will be transmitted electronically to headquarters for statistical processing and analysis. Our goal is to minimize expensive manual keying and processing operations.

Our system will reduce the logistical burdens associated with handling large volumes of paper questionnaires. Once original questionnaires are checked-in, prepared and scanned, subsequent operations will be accomplished using the electronic image and captured data.

We also have contracted with TRW to build out, staff, and manage the operations of four data capture centers that will house the DCS 2000 system. More than 120 million questionnaires will flow through the DCS 2000 systems at four centers located in Jeffersonville, IN; Baltimore County, MD; Phoenix, AZ; and Pomona, CA. Each of the data capture centers will employ between 1,500 and 2,000 people at the height of the census between March and July, 2000.

## **Computerized Matching Software System**

One of our main goals is to make it simpler for people to be counted by having census forms available in public locations, provided in multiple language translations, and mailed out twice in mail-out/mailback areas. Responses to the census also will be accepted over the telephone. Providing these response options will make it easier for everyone to be counted, but will increase the possibility of multiple responses for a given person and household.

We believe modern technology will allow the Census Bureau to spot and eliminate multiple responses from the same household. Major automation of the unduplication efforts was not feasible in past censuses because available technology and costs permitted only a small subset of person names to be data captured and unduplicated. This software also will be useful in helping us to implement other census follow-up coverage measurement operations.

## **Data Access and Dissemination**

In each decennial census, the Census Bureau's role as a data deliverer has grown steadily. For Census 2000, the Census Bureau has taken steps to ensure that the payoff--the delivery of data-- is commensurate with the scale of its fact finding effort.

The Census Bureau intends to give users additional data not available to them before and greater flexibility in using the data while controlling costs and assuring confidentiality.

The method of tabulating and disseminating data for Census 2000 will differ significantly from previous censuses. For the first time, the Data Access and Dissemination System, or DADS, will be on the Internet and will be the **primary** vehicle to disseminate decennial census data. Still in its developmental stages, DADS is an interactive electronic system that will allow data users access to predefined data products, metadata, and on-line help, as well as to build custom data products on-line.

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:

the Decennial Censuses, including the 1990 Census, the 1998 Dress Rehearsal and the Census 2000,

the 1997 Economic Censuses,

the American Community Survey (ACS), and

the 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.

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 must 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.

We have kept the end-user in mind for all aspects of DADS. We plan to provide them with access to Census Bureau data and allow creation of customized products. Inherent in our support of users is the ability to provide them with on-line help for using the system and accessing census data. Since we know that the Census Bureau is not the provider of every significant source of demographic and economic information, we will be providing links and pointers to relevant non-census federal data sets and non-federal data sources. The implementation of an on-line feedback system for evaluations and suggestions for products and the DADS system is another essential component of DADS.

To implement our users' requirements, our staff adopted an iterative development process that has the major benefit of staying current with changing technology over a five year period that allows for incremental expansion in the areas of function, content, and architecture. We are working jointly with IBM as a major partner to help us in the designing, building, and testing of the DADS system. A complete systems load testing will be conducted prior to actual deployment and release of this system. In turn, the results of the 1998 testing process of DADS will be used to help improve the DADS 2000 system.

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.

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 outgoing results. Technically difficult to implement, this option of DADS will require extensive internal testing and evaluation prior to release to the public.

Other miscellaneous systems that will be automated for the Year 2000 that you may be interested in include the Management Information (MIS) and Office Automation systems. We are automating both the scheduling and cost- and- progress components of our MIS using PRIMAVERA software . The scheduling component contains a data base of activities that is linked with date and duration information, while the cost and progress system will provide the real-time status of operations as they occur.

In the area of Office Automation, we are planning on using an automated document control system called PC DOCS. In addition, we have successfully developed an Intranet system of electronic mail so that we can more effectively communicate with the myriad decennial-related division employees, including our Field and processing offices, that must be brought into the messaging loop.

In summary, I'd like to note that the use of keypunch cards in the 1890 census ushered in an era of business machines, tabulators, and cash registers while the computer age flowed out of UNIVAC I that was used to process the 1950 census. With Census 2000 making it possible for anyone with a personal computer to log on to the wealth of data used by government agencies, researchers, and interest groups, we hope to vastly increase our audience for census data. This has been dubbed by many as the Age of Computer Technology, and we at the Census Bureau are going to do our utmost to use these tools to carry out the best census ever.