

## 1990 DATA COLLECTION AND PROCESSING DESIGN

Deborah A. Harner, Bureau of the Census 1/  
Washington, D.C. 20233

### INTRODUCTION

The purpose of this paper is to:

- identify the goals of the 1990 census
- summarize areas of major change in the census design from 1980 to 1990
- describe the proposed data collection and processing work flow for 1990 and
- summarize our remaining work

### GOALS

The design of the data collection and processing system used in the 1990 census must provide the setting for the Census Bureau to meet its stated goals.

Five major goals have been identified for the 1990 Decennial Census

- Meet constitutional and legal mandates.

We have a constitutional mandate to deliver counts for apportioning Congress to the President by December 31, 1990. This means that 9 months after Census Day, all of the state population totals must be final.

Starting with the 1980 Census, we have had a legal mandate to deliver counts to the States for their redistricting by April 1 of the year following the census. This means that population counts and some characteristics must be produced for very small geographic levels by April 1, 1991.

- Produce more timely data products.

We began planning our tabulation and publication program earlier than we did for 1980 and are committed to produce the data from the 1990 census on an earlier schedule than 1980. The information collected in the census is processed and tabulated at numerous geographic levels and released in a variety of forms.

- Maintain a high rate of overall coverage.

Our goal is to maintain a high rate of overall coverage, improve the accuracy of small area data, and reduce the undercount differential for population groups and geographic areas.

- Maintain strict confidentiality.

It is critical that we assure the real, as well as the perceived, confidentiality of the information we collect from individual people and households in the census. This commitment pervades all of our planning decisions and the design of the operations and systems we will use in the census.

- Strike an appropriate balance between the length of the questionnaire and the need for information by census data users.

The objective of assuring that the census questionnaires are of reasonable length requires a balancing act between the need for data and the number of questions we actually ask. In consultation with thousands of data users at all levels of the public and private sectors, we have identified many more valid and legitimate

data needs than we can possibly satisfy. In order to ensure the cooperation of the American people in 1990, we are pledged to asking only those questions that are legally required or meet important public needs.

The translation of these goals into a practical and cost efficient census design has been a major activity of the Census Bureau for the past several years.

In addition, a major objective in designing the 1990 census was to:

- Accomplish a dual strategy

This dual strategy requires that we obtain the best count possible, within timing and budget constraints, and be prepared to adjust the counts, if warranted.

All of these considerations guided the development of the 1990 census design.

### AREAS OF MAJOR CHANGE IN THE CENSUS DESIGN FROM 1980 to 1990

Advancements in technology have allowed us to develop automation tools to better meet these goals. New techniques and products from the ever changing field of information processing technology have allowed us to make existing computerized processes more efficient and automate several traditionally clerical tasks.

Increasing automation in the census can improve the accuracy of the data, lead to greater cost efficiencies, and give us more control over the entire census process. This control will better enable us to meet all cost, schedule, and data quality goals. Automation in a census context can mean many things. We have identified a number of functions that will be automated in 1990. We have tested these applications in the 1985, 1986, and 1987 test censuses. There are two major aspects of the Census that we will automate entirely in 1990: the geographic support system and the address control file (ACF).

#### Geographic Support System

The Census Bureau's geographic support system consists of three major parts: maps, address reference files, and master reference files. Maps graphically display physical and artificial features, such as streets, railroads, streams, and statistical boundaries. Census employees use these maps to follow up on unreturned questionnaires and to distribute and collect questionnaires where personal interviews are conducted.

Address reference files contain data on urban streets and the address range numbers for these streets. They allow the automated matching of specific urban addresses to specific geographic areas to facilitate the collection and tabulation of census data.

Master reference files contain a catalogue of all geographic and political areas. The files show the relationships between the lowest level geographic areas (blocks) and all other geographic areas (cities, counties, congressional districts, and states) for which data are gathered, tabulated, and published. In order to tabulate census results in a timely and precise manner, these three geographic support parts must contain consistent and accurate information.

#### Address Control File

This paper will concentrate on discussion of the ACF. The ACF is a computerized file of the census address lists. With an automated address file it will be much easier to determine whether or not we included a specific address in the file. It will be possible to update the file with an address we missed in earlier operations. It will provide for improved collection and processing management control. In addition, with an automated address list we can take advantage of barcode check-in technology for automated check-in as well as update the list and use it in future Census Bureau operations.

A few specific changes envisioned for the 1990 census include:

- Flow processing
- Concurrent data collection/processing
- Automated check-in and Management Information System (MIS) reports

In 1980 work was organized in enumeration areas for both data collection and processing. Questionnaires in each enumeration area flowed through a sequence of clerical and automated processes. All data collection, editing, and follow-up for a district office had to be complete before questionnaires could be shipped on to data processing. Data collection and data processing were separated into two very distinct stages.

The Census Bureau has developed and tested the technology to replace control at a batch or enumeration area level with control of work at the address or serial number level. Serial number control allows for flow processing and is possible through the use of an ACF. The ACF allows us to accomplish concurrent collection/processing and this will allow for an earlier start and completion of data processing activities. This is a major improvement over enumeration area processing in 1980.

Barcode readers can replace the error-prone and tedious clerical questionnaire check-in operation. It is estimated that barcode check-in by "wandering" can process 900 forms per hour versus the 1980 production of 250 forms per day per clerk for clerical questionnaire check-in.

The MIS can be built into the ACF to generate reports to monitor and manage the progress of decennial activities. The manual control systems that were used in 1980 to monitor progress were often inadequate for efficient management of operations. In 1990 the MIS reports are expected to provide accurate and timely data for district office (DO), regional office (RO), processing office (PO), and Headquarter's review.

## 1990 DATA COLLECTION AND PROCESSING WORK FLOW

### Overview

The 1990 census work flow reflects the integration of technological advancements as well as coverage and content improvements.

The basic 1990 work flow can be divided into three major components:

- Precensus or address list compilation activities
- Census or data collection and processing activities and
- Postcensus or final processing, count preparation and evaluation activities

This paper limits discussion of the 1990 data collection and processing design to the methodology used in mailout/mailback areas.

Precensus activities in mailout/mailback areas include address list compilation and some coverage improvement operations. Precensus activities result in files that are used to address questionnaires for mailout. Once an initial list is created, it is updated by several postal casing and field canvassing operations. Precensus activities differ in rural versus urban/suburban areas.

In rural areas where we are unable by computer to assign addresses to census geography, we develop the initial address lists and simultaneously assign geography in an operation called prelist. For this reason, these areas are called prelist areas.

Non-prelist areas are called Tape Address Register (TAR) areas and are areas where we have geographic support system information that enables us to assign geographic codes to addresses. We obtain initial lists from vendors for these areas and assign geography, or geocode, using computer algorithms.

### Prelist Precensus Activities

Initial address lists in prelist areas are created when Census Bureau enumerators systematically canvass and list the mailing addresses and in some cases, the physical location description of all units they encounter. In addition, they identify the geographic location on maps. These lists are keyed and the addresses are printed on cards for the U.S. Postal Service (USPS). The USPS will update these lists in an Advance Post Office Check or (APOC).

During this check the USPS will identify missing addresses, undeliverable addresses, and duplicate addresses as well as make corrections to addresses. After incorporating these changes, the Census Bureau will complete an APOC reconciliation operation to resolve differences between USPS and Census Bureau mailing address designations. During this operation enumerators use maps, information from prelist, and information from the USPS to update and correct the address lists.

In summary, the prelist precensus activities include:

- Prelist
- Prelist APOC and
- APOC reconciliation

### TAR Precensus Activities

In TAR areas initial address files are purchased from commercial vendors. After these addresses are geocoded by the Census Bureau they are printed on cards for the USPS.

The USPS will update these lists in an APOC operation and, as in prelist areas, they will identify missing, undeliverable, and duplicate addresses, as well as make corrections to addresses.

The Census Bureau will create listings of these updated addresses and closer to census day, complete a prec canvass operation to further update and correct these address lists. During prec canvass, enumerators canvass an area using these listings, verify that the address and geographic information are accurate and complete, and make necessary changes.

In summary, TAR precensus activities include:

- Procurement/geocoding of vendor lists
- TAR APOC and
- Precanvass

#### Final Precensus Activities

After the completion of these activities, additional coverage improvement and geocoding activities will follow. These activities include precensus local review and prelisting of special places. Questionnaire mailing packages will be labelled and provided to the USPS for a final postal coverage check 2 weeks prior to delivery.

Special place operations are designed for the enumeration of population groups whose living arrangements are different from the usual types of residences. Special places include places such as college and university dormitories, hotels, motels, hospitals, prisons, and the like. The local review operation is designed to provide all active, functioning local governments and American Indian reservations with an opportunity to review and identify major discrepancies between the census counts and locally prepared estimates.

#### Census Activities

Census activities begin when the USPS delivers questionnaires to all housing units on March 23, 1990. We will instruct households to complete the questionnaires as of April 1 and return them by mail. Both processing and district offices will receive mail returned questionnaires and check them into the ACF.

In high density urban areas (which are difficult to enumerate and in which we expect difficulties hiring enough staff) we will centralize the processing support as much as possible. For these offices, questionnaires will be mailed back to the PO for check-in so that the DO resources can be devoted to enumeration efforts. In other areas (which are easier to enumerate and where we can find adequate staff) we will decentralize processing activities. These DOs will check-in their own mail returns.

Approximately 3 weeks after check-in begins, the DOs will generate assignment lists for enumerators to use in nonresponse follow-up (NRFU). During NRFU we will visit each of the nonresponding addresses to obtain interviews. All addresses on the ACF lacking a check-in date will be scheduled for nonresponse follow-up. During NRFU we will visit each of the nonresponding addresses to obtain interviews. The NRFU is scheduled for 5 weeks in some offices and 6 weeks in other offices.

Concurrently, mail return questionnaires will be edited for content and coverage. Questionnaires received in the POs will be edited by computer and repaired clerically. Questionnaires received in the DOs will be clerically edited and repaired. Both the POs and DOs will conduct a telephone follow-up operation for questionnaires that failed content and coverage edits.

A field follow-up operation will recontact certain coverage and content failed edit cases and all units classified as vacant or deleted. This field follow-up operation will require 3 weeks. Complete, accepted questionnaires will flow to the POs for data capture and final editing.

We will use FACT 90 technology for data capture. This includes microfilming by specialized high speed cameras and FOSDIC scanning of the microfilm to convert responses to a computer readable form. We will compute local review counts and send these counts to local officials for a final round of review.

In summary, the census activities include:

- check-in
  - nonresponse follow-up
  - edit
  - field follow-up
  - postcensus local review
- Data capture and processing occur concurrent with these data collection activities.

#### Postcensus Activities

Postcensus activities fall into 2 general categories - activities related to the Post Enumeration Survey (PES) and final data processing activities. For the PES we plan to list addresses in sample blocks and conduct interviews before the DOs close. The PES data will be captured and matched to census data to produce undercount estimates.

During the final stages of processing the following activities will occur: edits and imputations, headquarters count review activities, searching and matching for coverage improvement and PES, creation of final edited data files, sample data processing and tabulation, and publication of data.

We will provide apportionment counts to the President no later than December 31, 1990 and finish production of redistricting counts by April 1, 1991.

#### REMAINING WORK

Before this proposed work flow is finalized we have some remaining work. The Census Bureau will continue to refine this 1990 work flow specifically looking at the timing of operations. We are currently in the midst of our 1988 Dress Rehearsal and we will use data from this test to finalize the 1990 data collection and processing design.

#### SUMMARY

To review, this paper has summarized the major census goals and how they lead to some specific changes in implementing the 1990 census. The work flow described in this paper reflects the integration of these goals and changes. The 1990 data collection and processing design relies largely on the strengths recognized in the 1980 decennial

design. The most significant area of change in the 1990 census will be the use and integration of automation. It is hoped that these changes will allow us to better meet our census goals.

1/ This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author(s) and do not necessarily reflect those of the Census Bureau.