CHAPTER 3.
Preparatory Operations

INTRODUCTION

By the fall of 1981, the essential characteristics of the 1982 Census of Agriculture had been decided and the Bureau began final preparations for the enumeration. These preparatory operations embraced four major activities: (1) finalization of the content of the census report forms, (2) compilation of the census mail list, (3) promotion of the census, and (4) printing and addressing the report forms for the initial mailout.

The preparation of the census mail list involved the acquisition, compilation, and unduplication of various lists of addresses believed to represent agricultural operations. The compilation was done in two major phases, the first involving not only the assembly of lists of addresses from various sources, but a major screening operation in early 1982—the Farm and Ranch Identification Survey—that was used to delete nonfarm cases from the list. The second phase involved both the list prepared in the identification survey and additional source lists not available for the construction of the initial list.

While the mail list was being assembled and unduplicated, the Bureau carried out a promotional campaign (see ch. 4) directed toward explaining the need for the census and encouraging cooperation and prompt response. Once again, the Bureau made extensive use of the broadcast media in its publicity campaign and encouraged dissemination of information about the census through vocational agriculture classes by distributing lesson plans and census guides to vocational agriculture teachers.

The content of the census report forms was finalized in January 1982. The content pretest had checked the capacity of the general design of the report form, but numerous changes were made after the pretest was completed. (See ch. 2 for a general description of report forms.) The report forms were printed by private contractors, who in some cases assembled the mailing packages. Once the initial census mail list was complete, the Bureau's Jeffersonville, IN, facility printed the address labels and affixed them to the mailing packages. Altogether, approximately 3.6 million packages were prepared for the initial census mailing in December 1982.

These activities are described in greater detail below.

ADDRESS LIST COMPILATION

Introduction

The Bureau of the Census introduced the mail collection of agriculture census information in the 1969 enumeration. The use of a mailout/mailback procedure requires an accurate and complete list of the addresses of agricultural operations meeting the Bureau's definition of a farm. Moreover, the list must not only cover the farms and ranches from which data are required, but must contain as few duplicate addresses as possible, since repeated requests for information increases both respondent burden and irritation.

For these reasons, the compilation of the census address list is a major part of the census. In the 1982 Census of Agriculture, the Bureau used a two-phase compilation—the Farm and Ranch Identification Survey and a final mail list compilation.

General Procedures

The Bureau compiled the agriculture census mail list from the records of previous censuses and from current administrative records supplied by a variety of Federal agencies and private associations. Names and addresses for specific agricultural operations frequently appeared on more than one source list, hence the various lists had to be matched to each other and "unduplicated" to delete multiple entries for a single agricultural operation. Two sources not previously used in the mail list compilation—the 1978 agriculture census out-of-scope and duplicate records files—were employed in the development of the 1982 address list to facilitate identification of agricultural operations and duplicate addresses.

The Bureau used nearly identical procedures in assembling the mail lists for the two phases of the address list compilation, with the major differences consisting primarily of modifications made to permit incorporation of the results of the Farm and Ranch Identification Survey to the main address list. Each list-construction program included five major operations: (1) formatting and standardizing all records in the source files, (2) matching and unduplicating employer identification number (EIN) and social security number (SSN), (3) geographic coding, (4) matching and unduplicating name and address, and (5) establishing controls and assigning identification codes. In addition, the Farm and Ranch Identification Survey involved a mail survey of addresses from the preliminary list whose status as agricultural operations was in doubt, while the final address

compilation program included matching the new source lists to the preliminary address file. The survey and the final address list compilation are described below.

Sources

Preliminary list—The Bureau began developing the identification survey mail list in the fall of 1981, with approximately 15.8 million addresses, using the main computer facilities at Suitland to compile and process the records. This list was a preliminary census mail list and was drawn from most of the sources that would be used for the final census mail address list. The sources used in the compilation of this list, and the number of addresses drawn from each, were as follows:

Source	Addresses
Total	15,835,422
1978 farms file	2,241,130
1978 multiunits and abnormals[1]	21,721
1978 nonrespondents	447,061
1978 census nonfarm file	1,290,677
1978 Farm and Ranch Identification Survey nonfarm file	1,948,564
1980 Agricultural Stabilization and Conservation Service (ASCS) file	4,974,853
1980 Statistical Reporting Service (SRS) list frame file for 31 States	1,861,737
1980 Business Master File (BMF)[2]	432,788
1980 Internal Revenue Service (IRS) form 1040C and 1040F files[3]	2,436,872
1980 SRS large operator list	66,138
Special lists	113,881

[1]Multiunits are companies or organizations with substantial agricultural operations at more than one location. Abnormal farms are farms operated by institutions, such as State agricultural research establishments, Indian reservations, and so forth.

[2]Includes IRS records for forms 1120, Corporation Income Tax Returns; 1065, Partnership Return of Income; and 941/943, Employers' Annual Tax Returns for Employees (941 coded SIC 01, 02, and 07 (Agriculture) for nonagricultural workers, 943 for agricultural workers).

[3]IRS form 1040C, Profit (or Loss) from Business or Profession (coded SIC 01, 02, and 07 (Agriculture)), and form 1040F, Schedule of Farm Income and Expenses, attached to form 1040, Individual Income Tax Return.

Final List—The preliminary mail list linkage process yielded a probable in-scope list of 1,916,000 addresses, while the 1982 Farm and Ranch Identification Survey identified 816,083 more in-scope records. These files, along with the names and addresses of that survey's nonrespondents, were combined with a number of new source files to assemble the proposed final unduplicated mail list of 6 million names and addresses. The sources used in the assembly of this unduplicated list, and the number of addresses drawn from each, were as follows:

Source	Addresses
Total	6,506,031
Preliminary list linkage, in-scope	1,916,000
1982 Farm and Ranch Identification Survey, in-scope	816,083
1982 Farm and Ranch Identification Survey, nonrespondents	522,422
IRS form 1040C and 1040F, 1981 tax year files	2,699,615
1981 Business Master File	491,071
1982 Farm and Ranch Identification Survey tenant/successor adds	34,840
Special lists	26,000

Format and Standardization

Prior to record linkage and the deletions of duplicates from the mail files, the Bureau had to establish a computer record format compatible with its processing programs and standardize the variety of computerized records assembled from the source lists. This involved a series of operations to identify each record's components, and reformat them as necessary, including the (1) assignment of a unique identification number, the source file number (SFN), to each record; (2) source record edit; (3) determination of name control; (4) insertion of surname locator; (5) address identification; (6) assignment of size codes; and (7) assignment of potential partnership or corporation (PPC) flags.

Assignment of source file number (SFN)—The format program assigned a unique identification to each computerized record to locate and identify it and the source from which it was drawn. Ranges of eight-digit numbers were reserved for each source used and the computer assigned numbers from these ranges to the appropriate records.

Source record edit—The basic edit program placed all records in a common format for processing. This format consisted of four data fields: (1) codes and data used in linkage and census processing, (2) primary and secondary names, (3) address, and (4) place. Each record was assigned an address priority code to identify the source from which it had been drawn and the comparative reliability of that source (for use in the linkage phase of the address list processing to retain the record most likely to contain the most recent address). The edit program also removed commas, periods, and certain special symbols (e.g., "&," "#") from the name and address field, inserted standard two-digit State abbreviations, and inserted a space between adjacent numeric and alphabetic figures. For example:

John A. Smith, Jr.		John A Smith Jr.
123 Main #201	became	123 Main 201
Doright, Idaho		Doright ID

Determination of name control—The name control (not the same as the name recode, which is discussed on p. 16) was normally the first four characters of the last name in the name field. The formatting program scanned the primary name field from *right to left* until it identified a nonnumeric word with three or more characters. The word was matched to a dictionary (called the "skip list") of 1,037 words and abbreviations to be ignored (e.g., "Farm," "Dairy," "Bros.," and so on). The first nonnumeric three-character word encountered that was not on the "skip list" was used to determine the name control. The first four characters (from left to right) of this word were inserted in the record's name control field. (If the word had four or fewer characters, the entire word was inserted in the field.) If the computer found no usable word after scanning the entire primary name field, the original name in the field was used as the control. A blank field was left blank.

Surname locator—The surname locator was an indicator inserted in each record identifying the field position of the first letter (character) of the name control. If no name control existed (i.e., if the field was blank) the surname locator was set at zero, and the record could not be recoded for name and address linkage.

Address identification—In address identification, the formatting program extracted numeric characters from the address field for use in determining the status of possible match records. Box numbers, rural route numbers, and street address numbers were identified in the address field and placed in two special data fields, one for box numbers and street address numbers, the other for rural route numbers. To extract the numeric characters, the program scanned the address from left to right until a numeric

word (i.e., one or more numeric characters) was encountered. If it was the *first* word in the field, it was stored in the box/street field; otherwise the word preceding (to the left) the numeric word was matched to a dictionary of acceptable words (e.g., "Box," "Rural Route," "RR"). If the word appeared in the dictionary, the numeric then was stored in the appropriate field (words such as "No" or "Number" were ignored). If the word did not appear in the dictionary, the rest of the field was scanned and, if no values had been extracted for either of the storage fields but one or more numeric words had been identified, the program returned to the first of these and placed it in the box/street field. If no numeric words were found in the address field, the storage fields were left blank.

Size coding—The format program assigned a size code derived from size indicators in the records when received from the sources. This code was placed in a specific data field, depending on the source of the individual record. During the record linkage process, all of the size codes for any record were retained by transferring the code from any record deleted as a duplicate to the appropriate field of the retained record. Once linkage and deletion of duplicates was complete, the computer program scanned the size codes in each record; if multiple codes were present, the specific code to be retained for the record was determined on the basis of the the source priority code for each size code. The sources, and the size indicators used for each, were as follows:

Source	Indicator used
IRS forms 1040C, 1040F, 1065, and 1120	Gross receipts
1978 Census of Agriculture I/S (in scope)	Total value of products from 1978 census report
IRS forms 941/943	Cash wages code
1978 Census of Agriculture nonrespondents	1978 mail list size code
Special lists	Varied by list—value of sales, acreage, and so on
ASCS list	No size indicator present
Tenant/successor file	No size indicator present
SRS sample frame list	Total value of production (when available)
1978 Census of Agriculture O/S (out of scope) and 1978 Farm and Ranch Identification Survey O/S	1978 mail list size code

If no size indicator could be derived from any source, size code 13 (unknown) was assigned.

Assigning potential partnership or corporation (PPC) flags—To prevent deletion of matched partnership or corporate records and individual records, the computer identified and "flagged" certain records as possible partnership or corporation (PPC) cases. For example, John Jones might have had both a sole proprietorship operation and a partnership with Joe Smith. In such a case, the computer edit program might match the partnership record to Jones' individual record on the basis of *his* name and employer identification number (EIN), and delete one or the other record as a duplicate. A PPC flag on the Jones/Smith record would change the match status to "possible duplicate" and a clerical review of the case would determine final disposition. The specific method used to identify PPC cases varied, depending on the source list involved.

The format and standardization program identified agricultural services records and deleted them from the mail file. Records

with ZIP Codes for areas outside the 50 States also were removed from the file (Puerto Rico, the Virgin Islands of the United States, and Guam were part of the 1982 Census of Agriculture but were enumerated by field interview). The following computer files were established in order to facilitate computer processing of the mail file:

Agricultural services records (for deletion)
Records with ZIP Codes for areas outside the 50 States (for deletion)
Trace sample (see p. 17)
Records without an employer identification number (EIN) or social security number (SSN)
Records with EIN's or SSN's
"Short" records[1] with EIN's, with or without SSN's
"Short" records with SSN's only
"Short" post office name records for records without EIN's or SSN's
Tally file (by size and geographic code)
All serialized records (all records from the input files with their SFN's)

Geographic Coding

General information—The name and address linkage for the address list was carried out within ZIP Code number, but all the records in the input file had to be geographically coded before the linkage could be done. The Bureau developed the geographic coding system to ensure that every record entering the name and address linkage system had standardized and edited agriculture census geographic codes, i.e., State and county numeric codes, county "alpha" (letter) codes, and ZIP Codes.

Master Geographic Reference File—The information needed to update and standardize the geographic codes in the address records was obtained from the Master Geographic Reference File. This file was created by combining computerized information from the ZIP Code Reference File and the 1974 census in-scope files.[2] The ZIP Code file contained a list of all the post office names and ZIP Codes in the United States. Each post office name was entered in its standard full spelling and in many variations, as well as in a full recoded spelling, along with the State and county numeric and alpha codes, ZIP Code, and telephone area code for each. All ZIP Codes with fewer than six in-scope records were deleted from the 1974 in-scope file. After this, the Bureau matched the 1974 in-scope file to the ZIP Code Reference File to produce a master file of unique ZIP Codes with the proper and common variant spellings of most post office names and the "most probable" county location for each ZIP Code based on reported primary location of the majority of the farm records from the 1974 in-scope file. The county location of the post office was not used to code *all* records, since about 25 percent of the post offices served more than one county. ZIP Codes not matched to the 1974 in-scope file retained their original county code.

---

[1]These were formatted complete records minus the names and addresses. Using them saved computer time in sorting and linkage operations. After EIN/SSN linkage, they were matched to the master records in the mail file using the SFN's to identioofy duplicates.
[2]The post office names were recoded in the same fashion as the names and addresses of operators. See p. 16 for information on the name recode. The Bureau had to use the *1974* in-scope file because no majority county tally files were run for the 1978 census file; i.e., no tallies were made of farm ZIP Codes reported by county.

Processing the mail file—Once the Master Geographic Reference File was prepared, it was used to edit the census mail file records in a series of computer operations that (1) checked the validity of the ZIP Code/post office name match on each record; (2) inserted ZIP Codes, post office names, and State and county alpha codes into records missing these items; (3) standardized spellings of post office names; and (4) assigned State and county numeric codes.

Once the entire mail file had been edited and geographically coded, it was ready for record linkage.

EIN/SSN Record Linkage

General information—The EIN's and SSN's provided the simplest method of linking records from the various source lists. Approximately 89 percent of all the records in the initial mail file had an EIN, SSN, or both. The Bureau's computer programs matched the EIN's and SSN's on each record to other records in the file to identify individual records, possible duplicates, and positive duplicates. When a possible duplicate record was found, the suspect records were displayed (printed out) for clerical review; a positive duplicate (i.e., the EIN's and SSN's were identical, the name controls matched, and no possible partnership or corporation (PPC) flags were present) resulted in a computerized review of the address priority code, and the deletion of the record with the lower address priority. Record linkage was based on matching EIN's to EIN's and SSN's to SSN's; records with both went through two separate linking processes. While the EIN and SSN linkages were done separately, clerical review of possible duplicates from both was done in one operation.

EIN linkage—All records with an EIN were submitted to the EIN linkage process. The records were sorted by EIN by PPC flag, name control, and address priority code, and entered the matching cycle in code priority order. This meant that records to be deleted always entered the matching cycle *after* the record that served as the original—called the "deleting"—record.

Records were read from the sorted input file into two temporary storage "locations" for comparison, as diagramed:

		Output files	
	Record 1		
Sorted input file	Record 2	EIN only (plus all deletes)	EIN with SSN
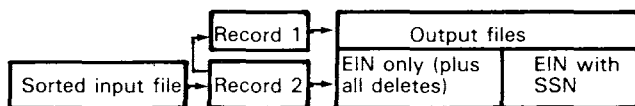
When records did not match—i.e., the EIN's were not the same—"record 1" was written out into the appropriate output file, "record 2" was moved to the record 1 location, and a new record from the input file was read into record 2 for comparison. When the EIN's of two records matched, the computer program reviewed the name controls and checked for a PPC flag; if the name control matched and there was no PPC flag, the records were declared a positive match. The sort done prior to the linkage ensured that record 2 had a lower address priority code than record 1; hence it was flagged for deletion. Its codes were transferred to record 1, and it was read into the appropriate output file while a new record moved into the record 2 location.

When EIN's matched, but name controls did not, or one or both records contained a PPC flag, the records were declared possible duplicates. No codes were transferred, but a possible-duplicate pair number was inserted in both records, linking them so that they could be displayed together for clerical review. If record 1 already contained one pair number, the same number was inserted into record 2, and record 1 was written into the appropriate output file, while record 2 was moved into the record 1 location. This cycle continued until the input file was

exhausted, all duplicates had been flagged, and all possible duplicates assigned pair numbers.

SSN linkage—The "EIN with SSN" output file from the EIN linkage operation was merged with the "SSN only" file to create the input file for the SSN linkage process. The input file was sorted by SSN, PPC flag, name control, and address priority code, and the basic matching and linkage procedures were the same as were used in the EIN linkage, except for the use of "dummy" file records and a slightly different procedure for assigning pair numbers.

"Dummy" records had to be used because there were records drawn from the IRS 1040C and 1040F lists that contained two social security numbers (usually, but not always, spouses) and some means had to be used to link records to both SSN's. The linkage itself was done on only one data field for each record, hence dummy records were created for form 1040C and 1040F records with two SSN's. The dummy records were duplicates of the master records except that the second SSN was substituted for the original, thus allowing linkage on both records. After the linkage procedure was completed, the dummy records were matched back to their masters, any codes picked up during processing were transferred to the master, and the dummy records were deleted.

Pair-number assignment in the SSN linkage phase followed the general procedures used during EIN linkage; the pair numbers assigned were a continuation of the sequence of the numbers used in the previous linkage operation. However, there were cases in which two records were possible duplicates in the SSN linkage, but each had a different pair number assigned from the EIN linkage. In such situations, the suspected duplicate cases retained their original pair numbers, and a secondary "collision" pair number was inserted into each record to tie suspected duplicates to the SSN-linked record.

Clerical resolution of possible duplicates—After EIN/SSN computerized linkage and deletion of positive duplicate records was completed, possible duplicates were sorted by pair number and displayed (i.e., computer printouts of the records were produced) in sets of two or more records—numbered sequentially within each set with a "label position number" (LPN)— with the same pair numbers. Clerks reviewed the records to determine whether or not the records within each set were duplicates. Records identified as duplicates were designated for deletion by circling the pair number and LPN and entering the "deleting record's LPN" (the DLPN) in the record for deletion so that the computer would transfer its identification codes to the deleting record. The pair-number/LPN/DLPN data were keyed for all clerical deletes and the results of the clerical review were processed by the computer program by matching these actions against the possible-duplicate file.

Name and Address Linkage

General information—All of the records not deleted from the mail file in in EIN/SSN linkage operation were subjected to a third matching operation using names and addresses. The name and address linkage involved the use of a SOUNDEX system[3] similar to that employed for the 1974 and 1978 censuses to compare the names and addresses on the records remaining in the file. Modifications were made to the 1978 system to improve (1) identification of name parts and secondary names, (2) linkage

---
[3]An index system based on the sound of the surname, rather than its spelling. In a SOUNDEX file, records for "Schumacher," and "Schoumacher," for example, would be indexed together to facilitate checking variant spellings of the name.

using first initial, (3) use of a middle initial, and (4) use of numeric address characters. (For a description of the 1978 procedures, see the 1978 Census of Agriculture *Procedural History*.)

The sequence of name and address linkage procedures for the Farm and Ranch Identification Survey, namely, (1) identification of name parts, (2) name and address recode, (3) name and address linkage and deletion of duplicate records, was carried over to the final mail list linkage operations.

Identification of name parts—Before the names and addresses in the records could be recoded for linkage, the name parts in the first and second name fields in each had to be identified. To do this, all of the words in each name field were compared to the "skip list" referred to above and were ignored if matched to any words on that list. The name fields then were scanned, and all the remaining characters and "character strings" (groups of characters) were classified as a surname, single letter, conjunction, nickname, or "other." Nicknames and conjunctions were identified by comparing each word in the field to another computerized dictionary and were classified accordingly. The classification codes used were:

Character string type	Code
Single letter	2
Surname	3
Conjunction	4
Other, including nickname	1

After each character and character-string in each field had been coded, the codes were retained in sequence and became the name pattern. The name pattern for each record then was used to identify each word or letter in the field. The fields were scanned from left to right—conjunctions were ignored and single letters merely identified as such—sequentially identifying each word as a first name, first initial (single letter), middle initial (single letter), or last name by comparing the name pattern to the computerized file of acceptable name patterns. There were 136 possible name patterns used for name-part identification. Approximately 99.3 percent of the name and address input file records were successfully matched to the pattern file. Records were rejected by the name-pattern matching program primarily because the surname locator code had been set at zero, or because a particular pattern did not match any of the acceptable ones. The latter situation occurred most frequently in multiple name strings, such as "Tom A Dick B and Harry C Smith."

Name and address recode—Once name-part identification was completed, the last name on the record was recoded by retaining the first letter of the name, deleting the second of each double character, and deleting all vowels (including "y"). The recoded name was left-justified and moved into a four-character storage cell (if the recode had fewer than four letters, the last character space(s) was left blank) with any excess characters (reading from left to right) dropped from the cell. For example:

HOLLAND became first, HOL-AND, then H-L-ND, and finally, HLND

SCHAEFER became first, SCH_ _F_R, then was truncated to SCHF

JONES became first, J_N_S, then was left justified to JNS_

The first name was recoded in the same way (e.g., JAMES became J_M_S, then JMS-; ROBERT, R_B_RT, RBRT, and so on), and first initials were identified and used alone. Middle names were not recoded, but the middle initial was identified and used. Once the first name was identified, it was checked against the "nickname dictionary" (a list of familiar names—"Dick," "Tom," "Bill," "Becky," and so on); if the name was found in the dictionary it was converted to the proper name ("Richard," "Thomas," "William," "Rebecca") and recoded. Nicknames that could represent several proper names (e.g., "Ed," "Ted") were recoded using the most frequently encountered proper name (e.g., for "Ed," "Edward"). Abbreviated names, such as "Geo." or "Chas.," were converted and had their proper names recoded.

A substantial number of records showed multiple name patterns—i.e., the record had a name following a conjunction such as "and," "or," or "&"— in the name field. When a record with a multiple name pattern was identified, dummy records were created for each possible name. Each dummy record carried all of the identification codes of its master record so that it could be matched back to that master after linkage was completed. Dummy records also were created for spouse names (except from IRS 1040C and 1040F source lists), names in the second name field, and partnership names.

For example, the name field for a record contained "John Jones & William Smith." The recode program identified the conjunction "&" following the middle name (in this case "Jones"), and determined that identifiable first and second alphabetic names followed it. The name field then was recoded with three possible combinations of names, "John Jones Smith," "John Jones," and "William Smith." If the character following the middle, or second, name was a conjunction followed by only a single name—e.g., "Smith Jones & Green"—all three names were recoded with individual dummy records, to identify partnerships that might change name order in different source file records.

The address recode actually was completed in the format and standardization processing, and involved identifying box/street and rural route numbers (see pp. 13-14 for details).

Record linkage—After recoding, master and dummy records were sorted successively by name and address recodes within each ZIP group, as follows: last name, first initial, PPC flag, dummy flag, box (number), rural route (number), first name, and source priority code. Once sorted, the file was ready for linkage.

The purpose of the linkage operation was to classify each record as (1) a duplicate, (2) a possible duplicate, or (3) a nonduplicate. Six key items were used to classify records: last name, first initial, first name, box/street, rural route, and middle initial. A match between the last names and first initials of any two records was required before further comparisons were made (records with matching last names but no first initials or names were sent through the entire linking cycle). If the last name and first initial matched, the records were compared successively on the other key items. Comparisons were made on all combinations of the match keys, and classification was based on the presence and extent of agreement among the various match keys.

As a result, the records were classified as nonduplicates, possible duplicates, or computer deletes (duplicates). Records in these classifications met the following descriptions:

Nonduplicates—Records matched on last name recode only, or records matched on last name recode but with different

first initials. Most records matching on last-name recode/first-initial/name recode, but with different *middle* initials, were classified as nonduplicates.

Possible duplicates—Records matched on first and last name recodes, but address information was absent or did not match. Records with only a first initial that matched on last name recode and address were classified as possible duplicates.

Computer deletes—Records matched on first and last names and on address information. (If a record in a set had a "Sr." or "Jr." attached to the name, the records were classified as possible duplicates and were displayed for clerical resolution, but if the two records had conflicting Sr./Jr. names (e.g., "James Jones Jr" and "James Jones Sr"), they were classified as nonduplicates.)

When a duplicate record was identified, identification codes were transferred from the case with the lower source priority code to the one with the higher priority, and the former was flagged for deletion. Possible duplicates were displayed (i.e., printed out) for clerical resolution. The clerks compared the linked records, determined whether they matched, identified the record to be deleted by circling the pair number and label position number of the duplicate, and entered the label position number of the deleting record on the duplicate. The duplicate/deleting record data for all clerical deletes were keyed into the computerized file and the results were processed in the computer by matching these actions to the file of possible duplicate records.

Final List Matching

The final address list compilation used a different matching scheme for the name and address linkage in which the *new* source records were matched to the preliminary mail file. All nonmatches then were unduplicated among themselves based on name and address procedures similar to those used in the preliminary compilation process. The residual records were added to the preliminary mail file to form the final census mail file.

Controls

General information—The Bureau established a system of checks and controls on the address fields in both the identification survey and final list phases of the compilation operation. These included a ZIP Code sample for possible testing of each phase of the processing, a trace sample for quality control of the overall operation, and control counts of records in the file at each processing step.

ZIP Code sample—A sample of records in the initial input files for both the identification survey and final mail list compilation was selected before the computer production runs began for possible testing of each phase of the computerized formatting, linkage, and deletion processes. All of the records in specified three-digit ZIP groups within various States became part of the sample. Plans originally called for processing the samples in the test runs so that any problems in the programs could be identified and corrected prior to actual production runs on the main files. However, while some testing and check operations were carried out, the entire planned testing program was cancelled because of time and staffing constraints.

Trace sample—A trace sample is a device used to observe the effect of processing on records. The Bureau assembled a trace sample for the mail list compilation by using the computer to flag every 1,000th record in the mail file prior to input to the format and standardization phase of the operation. When each record was selected for the sample it was displayed and reviewed by the research and analysis staff, this operation was repeated after each step in the processing of the mail file. This provided the Bureau's staff with files showing each sample record as it entered the compilation, and the changes made to it at each point in the processing cycle. The sample was used as a quality control tool and for research projects concerning the processing of the address file.

Control counts—At each stage of the processing cycle, the computer generated control counts of the number of records (1) in the input file, (2) in the output file, and (3) deleted. The counts served as numerical checkpoints at each phase of the processing. For example, the control counts for the EIN/SSN record linkage phase of the 1982 Farm and Ranch Identification Survey mail list compilation included the following:

Count	Records
Input to EIN/SSN record linkage	13,819,763
Computer deletes	6,488,703
Clerical deletes	299,473
Sets reviewed	318,000
Records reviewed	762,299
Output file	7,031,587

FARM AND RANCH IDENTIFICATION SURVEY

Introduction

The Bureau of the Census compiled the address list for the 1982 Census of Agriculture from a number of sources—primarily records from the previous agriculture census and other Federal agencies. The names and addresses on the "outside" source list generally represented individuals or places that were identified, for one reason or another, as having agricultural operations and qualified for participation in agriculture-oriented programs administered by the responsible agencies. The Bureau of the Census faced two major problems in using the lists compiled by these agencies: (1) duplication of address from one list to another, and (2) identification of agricultural operations that qualified as farms *under the Bureau's* definition.

The various matching procedures carried out during compilation of the list identified many duplicate addresses, as well as some of the operations that did not qualify as farms. However, a significant number of addresses on the list apparently represented farms under the Bureau definition, but they lacked sufficient confirming information for them to be positively identified. The Bureau introduced the farm and ranch identification survey in the 1978 Census of Agriculture as a means of attacking this problem; this proved so successful that it was included in the 1982 program as well.

Essentially, the 1982 Farm and Ranch Identification Survey involved the compilation and unduplication of a preliminary census address list, followed by a mail survey of operations whose status as agricultural was doubtful (usually because they appeared on only one source list). The results of the survey were used to identify out-of-scope cases for deletion from the final address file, as well as successors to persons who had discontinued agricultural operations and tenants not on the source lists.

Further, the information collected in the survey was used to update size information (i.e., acreage, volume of sales) for farm operators on the final mail list.

Pretest

Background information—The Bureau scheduled the 1982 Farm and Ranch Identification Survey for the early summer of 1982—several months before the final census mail list was established. While the report form used was similar to that employed in the 1978 survey, the Bureau considered a number of changes and tested the proposed survey report form format and content in the first quarter of 1981.

Report forms—The Bureau developed two report forms for testing. Form 80-A4(A)-T1 was designed to test sample groups consisting primarily of out-of-scope (O/S) cases. O/S cases were reports indicating that either the addressee was no longer associated with agricultural production, had already filed a return for the operation under another listing, or did not have sufficient agricultural production to qualify as a farm. This form incorporated a "skip" option in item 1, asking respondents if they had grown any crops or had any livestock or poultry during 1980. Respondents replying "yes" were instructed to answer questions about farm acreage, crops and livestock, value of sales, and type of organization. A "no" meant the respondents had only to check off why the farm was not in operation. Form 80-A4(B)-T1 was designed for use in cases where respondents were less likely to be O/S. An initial screening question was not used. After the respondents completed the basic agricultural questions, items 1 through 5, they were asked in item 6 whether they had reported any crops, livestock, or sales in items 3, 4, or 5. Respondents replying "yes" were instructed to skip item 7 (type of farm or ranch operation) and go on to item 8 (requesting Employer Identification Number (EIN)). Respondents replying "no" to item 6 were asked questions on land use and other agricultural-related activities. Both forms had remarks and identification sections.

The test sample—The Bureau selected a random sample of 2,320 addresses from the 1978 Farm and Ranch Identification Survey files and the 1980 Agricultural Stabilization and Conservation Service (ASCS) Survey nonrespondent lists for the test. Possible duplicate records, farms or ranches with sales of $100,000 or more, and addresses included in any of the 1978 census follow-on surveys were excluded from the sample. While the sample was geographically dispersed—i.e., the sample frame included files from each State—it was not a probability sample of the entire mail file and was not considered representative of the expected 1982 identification survey mail list. O/S and highly doubtful addresses (the ASCS survey nonrespondent list, for example, was considered a relatively unreliable source of in-scope (I/S) addresses) deliberately were included to test the capacity of the survey to obtain responses from, and identify, O/S addresses.

The sample was divided into nine subgroups to evaluate response rates achieved for addresses from various sources, then split into two files, one to be mailed the A4(A) form, the other the A4(B). The subgroups and the mail files by form type, were as follows:

Subgroup	Addresses	Form A4(A)-T1	Form A4(B)-T1
Total	2,320	1,142	1,178
1978 survey, O/S	308	308	—
1980 ASCS Survey nonrespondents	190	190	—
1978 survey I/S, census I/S, sales of less than $1,000	304	152	152
1978 survey I/S, census O/S, sales of less than 1,000[1]	283	142	141
1978 survey I/S, census I/S, sales of $1,000 or more	304	152	152
1978 survey I/S, census O/S, sales of $1,000 or more[2]	174	87	87
1978 census O/S	222	111	111
1978 census farms	307	—	307
1978 census criteria farms[3]	228	—	228

[1]Operations identified as farms, but which the census determined were not farms under the census definition.

[2]Operations that did not have sufficient sales in terms of *agricultural commodity production* to qualify as in-scope.

[3]Operations that normally would be expected, on the basis of inventories, acreage, and so on, to have sufficient volume of sales to qualify as farms, but did not do so, according to ASCS records for 1980.

Mailout and followup—The test packages, each containing the appropriate report form, cover letter, and a return envelope, were assembled at the Jeffersonville, IN, facility and mailed in two "waves." The first wave on January 30,1983, consisted of 1,178 form 80-A4(B)-T1 packages. Mailout was completed on February 2, with the 1,142 A4(A)-T1 packages.

The Bureau carried out a single mail followup on February 19, mailing duplicate complete test packages to *all* addresses in the test file. Receipts were closed out on March 15, 1982.

Response and processing—The test achieved an overall response rate of 70.1 percent, slightly lower than in the 1978 identification survey pretest. Sample group B (ASCS nonrespondents) had the lowest response—only 26.3 percent—and was deleted from the totals so as not to distort the results. (This reduced the total A4(A) mail file to 952 addresses.) The Bureau received a total of 702 A4(A) report forms, yielding a response rate from the A4(A) file of 73.7 percent (excluding the ASCS addresses), while respondents returned 874 A4(B)'s, achieving a 74.1-percent response rate. With the aberrant group B response deleted, the Bureau considered the differences in response rates between the two forms to be negligible.

Upon receipt at Jeffersonville, the test forms were checked in and given a technical review. Report forms were classified as either I/S or O/S based on the respondents' reported crops, livestock, and expected agricultural product sales value. In some cases a "point" system was needed to determine if an operation was I/S. Points were assigned based on the acres of crop land harvested and inventory of livestock. Total points then determined whether or not an operation met the farm definition.

A random sample telephone followup of respondents and nonrespondents was made to check data accuracy and determine reasons for nonresponse. A total of 25 I/S and 46 O/S cases were contacted and their status confirmed, although several O/S cases were found to have been misclassified because respondents did not follow directions. Fifty-five nonrespondents were contacted, most of whom indicated they had not realized the survey applied to their operations.

Results—Pretest response was considered good, with only a negligible difference between form 80-A4(A)-T1 and form 80-A4(B)-T1. Less than 29 percent of the total respondents

correctly followed the skip question. Over two-thirds of the respondents either left the skip question blank or followed the skip pattern incorrectly. Over one-quarter of the report forms received required application of the point system to determine whether the operation was I/S or O/S.

Since the difference in response between the two test forms was negligible, the Bureau adopted a modified version of the A4(B)-T1 report form for the 1982 Farm and Ranch Identification Survey. This form used the skip questions, immediately followed by reasons for nonfarm status, and greater detail in the crops and livestock sections, enabling the Bureau to make more accurate classification using the point system.

Development of Farm and Ranch Identification Survey Mail List

The initial list assembled for the 1982 Farm and Ranch Identification Survey was half-again as large as the file collected for the 1978 survey, and it was subjected to record linkage and unduplication procedures similar to those that would be used in preparing the census address list. (These procedures are described in more detail on pp. 15-16.) For the Farm and Ranch Identification Survey mail list only, one additional clerical operation, designated "Phase III clerical review" also was carried out. In this procedure, all records with "2+" numbers (those multiple record sets identified in the 1978 Census of Agriculture) and their associated linkages from the EIN/SSN and the name and address linkage operations were displayed for an additional clerical review. Sets including a partnership or corporate record were displayed and considered for inclusion in the identification survey. All possible duplicate pairs also were displayed and reviewed for possible inclusion.

After (1) linkage and (2) deletion of duplicate and unmatched 1978 O/S records, the preliminary mail file contained 4,969,809 addresses.

Based on previous census results and research, the Bureau considered addresses on the unduplicated preliminary list that had been found on only one source list, or on two or more selected lists that had high proportions of nonfarm addresses, to be candidates for the survey, including those from the following sources or combinations of sources:

1978 Census of Agriculture I/S, size code 13 (sales of less than $1,000)
1980 ASCS Survey only
1980 IRS form 1040F only (less than $100,000 in receipts)
1980 Business Master File (BMF) only (less than $100,000 in receipts)
1978 Census of Agriculture I/S only, size codes 4-12 (sales under $100,000)
1978 Census of Agriculture I/S, size codes 4-12, and 1980 ASCS Survey
1978 Census of Agriculture nonrespondent only
1978 Census of Agriculture nonrespondent and 1980 ASCS list
1978 Census of Agriculture nonrespondent and 1980 SRS list
1980 SRS list only
1980 SRS and 1980 ASCS Survey lists
1978 Farm and Ranch Identification Survey or 1978 Census of Agriculture I/S and "miscellaneous" 1980 source Special lists only
Possible duplicates identified during clerical review
1978 Census of Agriculture I/S and 1980 SRS lists; 1980 SRS and ASCS and 1978 Census of Agriculture I/S size codes 4-12

A computer selection of the preliminary file identified addresses from these sources for inclusion in the identification survey. The resultant file contained 3,053,806 addresses. In December 1981, the Bureau prepared computer tapes for use in the production of address labels for the mailing packages and began preparations for the mailout from Jeffersonville.

Mailout and Followup

Preparations and initial mailout—The identification survey mail file was split into three "waves" to distribute labeling production and package assembly workload. Each wave contained approximately 1 million addresses drawn from the States of two or more census divisions. The divisions and States covered by each wave were as follows:

Wave	Census divisions	States
1	1	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
	2	New Jersey, New York, Pennsylvania
	3	Illinois, Indiana, Michigan, Ohio, Wisconsin
	6	Alabama, Kentucky, Mississippi, Tennessee
2	4	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
	5	Delaware, Georgia, Florida, Maryland, North Carolina, South Carolina, Virginia, West Virginia
3	7	Arkansas, Louisiana, Oklahoma, Texas
	8	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
	9	Alaska, California, Hawaii, Oregon, Washington

The Bureau prepared computer tapes of the addresses on the mail list and printed address labels in ZIP Code sequence using the high-speed printers at the Jeffersonville office.

The Jeffersonville staff assembled, labeled, and mailed the survey packages as the labels were produced and delivered. Approximately 2.9 million packages were assembled for the initial mailout consisting of the report form 82-A4, 1982 Farm and Ranch Identification Survey, a form 82-A4(L1) transmittal letter explaining the need for the survey and requesting prompt response, and a form EC-4A return envelope, overprinted "AG-FR" in the lower left-hand corner. About 200,000 more packages were prepared that included all of the above contents plus a form 82-A4(I) insert sheet requesting respondents who received more than one report form to either complete one copy and mark the others "duplicate" (returning all to the Bureau in the same package) if they were all for the same operation, or to complete a separate report form for each *separate* operation. (The A4(I) insert sheet was used only in the first mailout for possible partnership and/or multiunit operations identified during phase III clerical review [see above].)

The initial mailout to 3,053,325 addresses was carried out on a flow basis (by wave and in ZIP Code sequence within each wave) from March 5 through March 10. The details of the mailing were as follows:

Inital mailout	Began	Completed	Contents	Packages
General	3/5/82	3/10/82	A4 report form, A4(L1) transmittal letter, EC-4A return envelope	2,860,541
A4(I) Insert[1]	3/5/82	3/10/82	A4 report form, A4(L1) transmittal letter, form A4(I) insert, EC-4A return envelope	192,784

[1]After the initial mailout, cases that originally received packages including the A4(I) insert were treated as regular cases for followup purposes.

Followup mailings—The first of five followup mailings for the survey began approximately 5 weeks after the initial mailout was completed, on a flow basis beginning with wave 1 addresses and working through wave 3. The first followup employed a 4" x 6" reminder card, form 82-A4(P), to request replies to the survey from nonrespondents. The remaining four mailings involved complete survey packages identical—except for the transmittal letter—to that used in the initial mailout.

The cutoff date for response to the first mailout was April 13. A new computer tape of nonrespondent and postmaster return (PMR) addresses as of that date was prepared and used to print mailing labels, each of which included codes identifying the address as a nonrespondent or PMR. The packages were preassembled at Jeffersonville so that the staff needed only to affix the new address labels. An identical routine was followed after each cutoff date, although not all PMR's were remailed (see below). The period between the completion of each mailing and the closeout date for response tended to shorten as the survey proceeded. The mailing dates, package contents, and total number of packages mailed for each mailout were:

Mailout	Began	Completed	Contents	Packages
1st followup	4/13/82	4/22/82	A4(P) reminder	1,313,375
2nd followup	5/11/82	5/21/82	A4 report form, A4(L2) transmittal letter, EC-4A return envelope	1,159,000
3rd followup	6/2/82	6/10/82	A4 report form, A4(L3) transmittal letter, EC-4A return envelope	678,357
4th followup	6/22/82	6/24/82	A4 report form, A4(L4)transmittal letter, EC-4A return envelope	417,323
5th followup	7/13/82	7/15/82	A4 report form, A4(L5) transmittal letter, EC-4A return envelope	289,477

Telephone survey of nonrespondents—The Bureau carried out a telephone survey of nonrespondents to the identification survey to (1) determine reasons for nonresponse, and (2) develop ratios of in-scope to out-of-scope cases for all nonrespondents and for specific source groups of nonrespondents. The Bureau drew a random 1-in-1,000 sample of the survey nonrespondent file as of July 15, 1982. This yielded 562 cases, including 91 A4(I) cases that were deleted, reducing the actual sample to 471 addresses.

An address label was prepared for each sample case and was attached to an A4 report form, together with a cover sheet containing a number of questions on the clarity of items on the A4 and of the purpose of the survey. The Agriculture Division staff at Suitland tried to contact each addressee in the sample and complete a report form for his/her operation, as well as obtain responses to the cover sheet survey questions. A number of problems were encountered in the study, not the least of which was the fact that over a third of the sample nonrespondents (184 in all, or 39 percent) had unlisted telephone numbers and could not be reached. Contacting the farm *operator* proved a problem as well; frequently a spouse answered the telephone, and many operators held off-farm jobs and could not be reached during normal working hours.

The Bureau managed to contact 169 persons during the survey, 78 of whom said they had received a report form, while 31 asserted that they had not and 50 "did not know." About half of those who said they had received a form claimed to have completed it. The reason given most often for not completing

and returning the report was that the respondent was "not in the farming business." Of the cases classified in the nonrespondent survey, 75.5 percent were identified as out of scope, indicating a high probability that nonrespondent addresses did not represent farms under the Bureau's definition.

Postmaster returns (PMR's)—Inasmuch as the addresses selected for the identification survey were those on the census mailing list considered least likely to represent farms, the Bureau decided that it could not justify extensive followup efforts to addresses that the Postal Service could not locate. Therefore, postmaster returns (PMR's) were remailed only in the first mail followup to the identification survey, and PMR's with SRS-only or ASCS-only source codes (some 277,245 cases) were not remailed at all.

New mailout packages for PMR's to be remailed in the first followup were assembled at Jeffersonville. These were identical to the initial mailout packages except that the outgoing envelope was overprinted "Please complete and return within 7 days," in the lower left corner, and "M2" in the lower right. PMR's with an address change, or with an indication that the addressee was deceased, were referred to the correspondence unit for remailing or for successor search. New address labels bearing primary and secondary codes of "2-1," "2-2," and "2-3" (for addresses from waves 1, 2, and 3, respectively), were prepared and affixed through the windows of the outgoing envelopes. The PMR remailing to 135,282 addresses was carried out simultaneously with the mailing of the reminder cards during the third week of April 1982. Some second-time PMR's were received, but no further mailings were done.

Processing

Receipt, batch and check-in—Upon receipt at Jeffersonville, the unopened packages were sorted into (1) nonagricultural mail, (2) A4 receipts with visible barcodes, (3) PMR's, (4) census-originated correspondence (AGCOR), post office address corrections (including PMR's with corrections), and (6) other receipts (including A4 receipts without visible barcodes). The sorted mail was referred as follows:

Receipt type	Referred to—
Nonagricultural mail	Mail room for distribution
A4 report forms with visible barcodes	Batch for check-in
Postmaster returns (PMR's)	PMR sort
AGCOR	Correspondence unit
Postal address corrections (includes PMR's with corrections)	Name and address keying (for mail list updating)
Other receipts (includes A4 receipts without barcodes)	Sorting—receipts without visible barcodes

The batching unit grouped A4 packages with visible barcodes into work units of about 500 forms each, placed them in a plastic bag with a form A402 Check-In Work Unit Cover Sheet (the A402 contained the work unit number assigned to the batch, along with the date, primary and secondary status codes, and the count), and sent them to the check-in unit for barcode reading and automatic check-in. At the check-in unit the A4 packages were removed from their plastic bags, the data on the A402 were entered on computer tape, and the unopened A4 packages were placed on a conveyor for barcode reading. The barcode reader used a laser device to scan the barcoded CFN's on each mailing label and transferred the data to computer tape. Barcodes rejected by the automatic scanner were checked in using hand-held laser "wands," which also captured and transferred the

CFN information to tape. Data from the computer tapes were transmitted daily to Suitland and were used to update the check-in control file. After confirmation from Suitland that the tapes had been received and loaded into the control file, the checked-in A4 receipts were released for clerical review.

Receipts without visible barcodes were referred to a manual sorting unit where each package was opened. The CFN, if any, was identified, keyed into the computer, and the case was referred to one of the units according to its contents. Nonagricultural mail was referred to the mail room for distribution. Respondent-originated correspondence and report forms with correspondence attached were sent to the correspondence unit, and A4 receipts were checked for (1) presence of a CFN, (2) acreage reported, (3) tenants indicated (i.e., any reported on the place), and (4) remarks. Congressional correspondence (correspondence (1) from a member of Congress, or (2) from a respondent who indicated he or she had contacted their Representative or Senator about the census) was sent immediately to the supervisor. Report forms with CFN's were sent to batch for check-in, and respondent-originated correspondence with CFN's were sent for batching for check-in keying.

Clerical review—A clerical review of the A4 report forms determined whether or not a given operation was I/S or O/S. If the data on the A4 indicated an agricultural operation, the reviewing clerk wrote "I/S" in the label area of the form and referred the case for keying. O/S cases were reviewed to determine if the respondent named a tenant, successor (the A4 requested respondents with no agricultural operations of their own to provide the names and addresses of any tenants or successors that might have such operations), partner, or another person who may have filed a report form. The appropriate code was written in red ink at the center of the upper margin of the face of the report form as:

    T = Tenant
    S = Successor
    P = Partner
    CF = Claims filed

These cases then were referred for research.

Report forms for cases determined to be O/S were coded "O/S" in red ink in the upper right-hand corner of the face of the form, along with a numeric code indicating the reason the case was identified as O/S. These codes were:

    O/S-2 Deceased
    O/S-3 Landlord only
    O/S-4 Nonagriculture without land
    O/S-5 Nonagriculture with land
    O/S-7 All others not covered by the above categories

As the individual report forms were reviewed, they were grouped by code for further processing. When more than one code was present (e.g., "I/S" and "2+") disposition was determined by the priority assigned to the codes. The codes, in priority order, and the disposition of the report forms after clerical processing, were:

Code groups by priority	Disposition
1. R	Problem-solving analyst
2. 2+	"2+" analyst
3. Form letter assigned	Correspondence typing
4. S, T, P, or CF	Research
5. O/S without address change	Laser wand check-in grouped by reason O/S code
6. O/S with address change	Laser wand check-in grouped by reason O/S code
7. I/S	Batch for data keying

Technical review and interactive research—The technical review unit was responsible for resolving problem cases referred to it from the receipt-and-sort, data-keying, correspondence, clerical-review, and research units, as well as for checking "large" farm cases (i.e., places with total values of products (TVP) sold of $100,000 or more, or with 1,000 acreas or more of land) and "2+" reports. Cases were referred to technical review because (1) I/S or O/S determination could not be made, (2) names and addresses of tenants or successors had not been adequately reported, (3) reported tenants and successors were possible matches to the census mail file, (4) attached correspondence was in conflict with reported data or the content of the correspondence was not understood, or (5) referrals from the clerical unit required O/S review. Report forms were reviewed in more detail than in the clerical review and responses were analyzed to determine the status of each case and to identify tenants or successors and so on. Cases were designated I/S if the reported data indicated $750 in estimated value of agricultural production in 1982, or if the data could be assigned a specified number of "points" based on acreage, production of various crops, or livestock inventories. In addition, any respondent producing crops, livestock, or poultry under contract was considered I/S. Other problem cases (e.g., large acreages of idle land, blank reports, refusals, abnormal farms, and so on) were referred to senior analysts for detailed review, while Alaska and Hawaii cases were sent to subject-matter specialists at Suitland.

After technical review, the coded report forms were either sent to the control unit for return to the processing cycle, or to special units (senior analysts, correspondence, or telephone) for further disposition. Respondents who had recently sold or rented their property were asked to report the successors' and/or tenants' names, which then were referred to the interactive search unit for matching to the preliminary mail list using computer terminals and the interactive SOUNDEX name-search system (see p. 15). Tenant/successor names that did not appear on the list were keyed during the data entry phase of the survey operation and were considered new operators for the final mail list compilation.

Data keying and computer processing—Report forms were batched into work units of 100-200 forms each for data keying and for keying of tenant/successor names and addresses ("adds"). The electronic key-to-disk system at Jeffersonville was used and allowed data to be keyed to disk and then transferred to computer tape for transmission to the main computer facility in Suitland using the telephone data link system. The general I/S file from the identification survey was assembled for use in the compilation and unduplication of the final census mail list.

Results

Response to the survey—The 1982 Farm and Ranch Identification Survey achieved an overall response rate of 82.9 percent, including approximately 14.6 percent PMR's. This was considered fairly good response, although somewhat lower than was attained in the 1978 identification survey. (In 1978, the overall response rate was 89.9 percent, with 9.3 percent PMR's. The difference in both the size of the survey mailout [slightly over 4 million for 1978 and slightly more than 3 million for 1982] and the lower response rate for 1982 were partly the result of improved identification and deletion of I/S operations in compiling the survey mail list.) Response rates varied considerably depending on the source list for any given address. The highest rates of response were achieved for addresses drawn from various combinations of sources involving the 1978 I/S list,

ranging from a low of 87 percent for addresses from the 1978 "I/S only list," to 93 percent for the "1978 I/S and others" list. The lowest response rate attained was for addresses drawn from the 1978 census nonrespondent list—only 48 percent. Respondents returned 2,085,728 report forms to the Bureau, and 445,175 PMR cases were received.

Results—The Bureau classified 1,269,645 (60.9 percent) of the report forms received from the identification survey as O/S. These cases included landlords with no agricultural operations of their own, ex-owners, duplicate addresses, and so on, as well as persons who simply had no connection with agriculture. The O/S names and addresses, and the identification survey PMR cases were deleted from the final mail file. Corrections to names and addresses, acreage and other size codes, and other data from the 816,083 I/S cases identified in the survey were used to update the final census mail list.

FINAL MAIL LIST PREPARATION

General Information

The Bureau carried out a second address list compilation and linkage process between September and December 1982 to prepare a final mail address list for the census proper. The in-scope lists from the preliminary address linkage, the 1982 Farm and Ranch Identification Survey, and additional source lists not available until the late summer and early fall of 1982 were incorporated into a proposed final address list for linkage and preparation for the census. The principal operations involved in the assembly and linkage of the final address list were:

1. Assembly of the mail list from sources.
2. Formatting and standardization of all records.
3. EIN/SSN record linkage and deletion of duplicate records.
4. Geographic coding.
5. Name and address record linkage and deletion of duplicate records.
6. Establishment of controls and assignment of identification codes.

These procedures were similar to those employed in the compilation of the preliminary list from which the Farm and Ranch Identification Survey mail list was drawn, except that a variety of new source lists were included in the final file and the name and address record linkage was modified to reduce costs and time required for processing. (See pp. 12-17 for details.) The final mail list preparation included several additional steps:

7. Assignment of census file numbers (CFN's).
8. Identification of "must" and "certainty" cases.
9. Mail list sampling.

These activities are described below.

Census File Numbers

Final preparation of the mail lists for both the identification survey and the census proper included the assignment of a unique identification number, the census file number (CFN), to each record. The CFN consisted of 11 digits arranged in three groups. The first five digits were the State and county codes for the address on the record, the second five digits comprised the serial number identifying the specific operation within its county, and the eleventh digit was a check digit. The check digit provided a mathematical check for quality control during data keying of the returns.

Must and Certainty Cases

"Must" cases were agricultural operations so large that failure to include their data would significantly distort the census statistics, or which required special handling, such as multiunits. "Must" cases were selected by computer after record linkage was completed for the final mail list. Using size codes and lists of multiunits from the 1978 census in-scope list, and other size indicators from the mail files, the computer program selected addresses (1) the Bureau believed represented operations so large that some data had to be obtained, rather than imputed, in cases of nonresponse; (2) for which an explanation was needed of *why* the addressee was not engaged in agricultural operations; and (3) for which there were indications the census return would require a special analyst's review. These broad categories embraced the following types of addresses:

Multiunits—Multiunits were companies or organizations with substantial agricultural operations at more than one location. In general, two or more report forms were required for each organization, and each establishment was considered a separate farm. Separate mail files were maintained for each master (company/organization) record and each associated establishment. Multiunits identified before the census mailing had multiunit identification numbers assigned in the alpha/plant field of the address label, indicating whether the report form was for the master or an associated establishment.[4]

Abnormal farms—Abnormal farms were farms operated by institutions such as State agricultural research establishments, prisons, Indian reservations, and so forth.

Other farms—The "other farms" category included addresses believed to represent individual agricultural operations large enough to qualify as "must" cases. The criteria used for determining status as a must case, the expected total value of products (TVP) sold and/or total acreage, varied among States. The minimum criteria for assignment as a must case for Florida, for example, were a TVP of $500,000, or a total of 2,000 acres or more. On the other hand, a must case for West Virginia had to have a TVP of $100,000 or more, or at least 1,000 acres of land. The requirements for most other States were TVP's of $200,000 or more, and minimum acreages varying from 1,000 to 10,000. (Acreage and TVP requirements tended to be lowest in the Eastern States.)

"Certainty" cases were cases that did not qualify as "must" cases on the basis of size or type of farm, but were considered sufficiently important to justify intense followup. A portion of the "certainty" cases were selected on the basis of acreage and TVP. The minimum acreage requirement generally was the same as for "must" cases—i.e., ranging from 1,000 to 10,000 acres. Minimum TVP (based on historic and mail source list data) varied from $40,000 to $100,000. The "certainty" cases also included *all* farms in counties that had fewer than 100 farms enumerated in the 1978 Census of Agriculture.

---

[4]The numeric "alpha" code identifying the company was a six-digit number in the alpha field of each of the various establishments' records. The "plant" code was a four-digit establishment identifier. The master record for a multiunit would have the company identifier in the alpha field and four zeros in the plant field, while each associated establishment had the company identification plus an identifying number in the plant field. Each report form for a master or an associated establishment was assigned a unique serial number, the associated establishments being assigned numbers in immediate sequence following the master.

Mail List Sampling

The Bureau reintroduced sampling as an agriculture enumeration technique in the 1978 Census of Agriculture (it previously used sampling in the 1945 Census of Agriculture, but had restricted it primarily to follow-on surveys since that time) as a means of collecting selected data while keeping respondent burden to a minimum. The technique was very successful and the Bureau decided to use it again for the 1982 census. The sample employed for the 1982 enumeration was drawn in essentially the same manner used for 1978; it included all "must" and "certainty" cases, and a sample of all other addresses on the census mail list.

Sampling of noncertainty cases—i.e., the "regular" sample—was done on a county basis, with the sampling rate determined by the number of farms enumerated in each county in the 1978 census. For counties with fewer than 200 farms, but more than 100, 1 in every 2 farms was added to the sample, while in counties that had more than 200 farms, 1 in every 6 farms was selected. The "must" and "certainty" cases drawn from the final census mail list totaled 326,131 addresses, and the "regular" sample of all other farms added 573,052 to the sample list. These 899,183 addresses comprised an approximate 25-percent sample of the 3.65 million records on the final mail list, one large enough to provide reliable county-level estimates for the sample items. The sample addresses—"must," "certainty," and "regular"—would be mailed report forms that included both the "core" items asked of all farms and the sample items. (These sample report forms reflected the Bureau's questionnaire-regionalization policy in that the sample forms in each region were adapted from the regionalized nonsample questionnaire for that region.)

PRINTING AND ADDRESSING REPORT FORMS

General Information

For the 1982 Census of Agriculture, the Bureau employed printing and package assembly procedures generally similar to those used for the 1978 enumeration. Private contractors printed all the census materials and assembled the mailing packages, and forwarded the packages to the Bureau's Jeffersonville facility for final preparation and mailout.

Address Labels

In the 1978, census the Bureau used a private contractor to print the address labels for the census mailings. The need to maintain the confidentiality of the census mail file required extraordinary security measures, and the Bureau decided to avoid these complications in the 1982 census by preparing the mailing labels "in-house." Bureau headquarters prepared computer tapes of the mail file and sent them to Jeffersonville by courier. The Jeffersonville computer facilities included six Printronix high-speed printers, which had all the capabilities needed to print the address labels—including machine-readable bar codes—and these were used to produce the address labels for the initial and followup mailings.

The first of approximately 200 reels of computer tape containing the final mail file for the initial census mailout were sent to Jeffersonville at the end of August 1982, with the rest arriving on a flow basis as the file was completed. Printing of the labels began in September, and the last of the 3.65 million needed for the census mailing were delivered in mid-November. After each followup mailing closeout, the Bureau updated the nonrespondent list and dispatched computer tapes with the new mail file to Jeffersonville for label preparation. Separate address labels were used for all mailings except the first followup, which employed the form 82-A01(L2) reminder card. The computerized address files and the Printronix printers were used to print addresses directly onto the face of the cards.

Printing, Assembling, and Addressing

General—Eight private contractors printed the report forms and the other census mailout materials, and assembled the mailing packages according to Bureau specifications. Each contractor printed *all* the materials for a given package (e.g., the package for the initial mailing to nonsample addresses in a particular geographic region), assembled the packages, and sent them to Jeffersonville. The printing contractors also printed and assembled (1) the mailing packages for the followup mailings; (2) additional nonsample, sample, and "must" packages for mailing to postmaster return (PMR) cases and "adds;" and (3) all types of report forms as general reference materials and for mailing to respondents on request. At the Jeffersonville facility, Bureau personnel carried out quality-control checks (see below), added special instruction sheets to some packages (about 75,000 addresses were identified as requiring special instructions—e.g., bee and honey producers, contract poultry producers, or worm producers), and applied the address labels for the mailout.

Quantities—The total number of standard report forms printed, and those used for the initial mailout, by form type and geographic region, are shown in table 1.

Table 1. **Standard Report Forms Printed**

Region	Report forms					
	A01 (nonsample)		A02 (sample)		A03 (must)	
	Total	Initial mailout	Total	Initial mailout	Total	Initial mailout
Total[1]	6,246,300	3,234,000	2,393,000	1,023,000	411,900	157,400
01	430,000	245,000	162,000	82,000	32,200	15,400
02	1,307,000	724,000	422,000	209,000	55,300	27,500
03	604,000	327,000	218,000	106,000	27,200	12,000
04	1,317,000	689,000	474,000	227,000	47,100	22,000
05	116,500	60,000	38,500	18,000	14,500	5,500
06	963,000	481,000	319,000	146,000	48,300	23,000
07	500,300	242,000	155,000	72,000	27,100	12,000
08	249,500	135,000	102,000	47,000	50,600	11,000
09	159,500	80,000	64,000	30,000	19,000	9,000
10	224,500	123,000	72,000	35,000	19,000	9,000
11	244,000	128,000	92,000	42,000	26,600	11,000
12[2] ...	—	—	20,500	9,000		

[1]Includes quantities of each type, for each region, *not* included in packages. These, by type, were: 82-A01 (nonsample), 131,000; 82-A02 (sample), 254,000; 82-A03 (must) 45,000.
[2]Hawaii.

A facsimile of a representative report form is included in appendix G.

Other items ordered for the data collection mailings included the following informational materials, letters, and envelopes.

Form number	Description	Quantity
82-AO1(I)	Information sheet	8,191,200
82-AO1(B)	Brochure	4,663,400
82-AO1(L1)	Transmittal letter	5,091,400
82-AO1(L2)	Reminder card	2,650,000
82-AO1(L3)	Followup letter	2,152,000
82-AO1(L4)	Followup letter	1,452,000
82-AO1(L5)	Followup letter	1,106,325
82-AO1(L6)	Followup letter	1,741,000
82-AO1(L7)	Followup letter	250,000
82-A31	Special instruction sheets	112,500
82-A31A	Abnormal farms—institutional	4,000
82-A31B	Abnormal farms—grazing associations	500
82-A31C	Contract poultry producers	40,000
82-A31D	Bee and honey producers	2,400
82-A31E	Laboratory animal producers	200
82-A31F	Feedlot operators	10,100
82-A31G	Nursery and greenhouse products	50,000
82-A31H	Worm producers	900
82-A31I	Fish and aquaculture	1,400
82-A31J	Multiunits	3,000
82-A7A	Outgoing envelope	4,850,000
82-A7A (Rev.)	Outgoing envelope	200,000
82-A7B	Outgoing envelope	3,400,000
82-A7C	Outgoing envelope	300,000
82-A7D	Outgoing envelope	600,000
82-A8	Return envelope	8,900,000
82-A9	Letter envelope	3,000,000

Facsimiles of the information sheet, transmittal letter, reminder card, and followup letters are included in appendix G.

Quality control—The contents of the initial mailing packages for nonsample, "must," and sample cases are given in table 2.

Upon receipt in Jeffersonville, each shipment of assembled packages was submitted to a quality control check. This involved a manual, random sampling of three packages from each carton if the shipment comprised fewer than 10 cartons, from 5 randomly selected cartons for shipments of from 10 to 150 cartons, and from 20 cartons for shipments of more than 150 cartons. The packages pulled for the check were opened and their contents inspected to ensure that all of the required items were included, in the right order, and that the label area of the report form was visible through the open window of the mailout envelope. If everything was correct, the cartons and shipment were accepted; if not, the remaining packages in any carton with a defective package were inspected either manually or by shadowgraph. If 2.5 percent or more of the packages within a carton were defective, all the cartons in the shipment were subjected to 100-percent inspection. Once all defects had been corrected, the packages were returned to the mailing preparation cycle.

Multiunits and abnormals—Multiunits with agricultural operations were mailed agriculture census "must" forms in the general mailout to multiunits carried out by the Economic Census Staff as part of the 1982 Economic Censuses. During preparation of the final mailing list for the agriculture census, the Agriculture Division established a separate computerized file of multiunits with agricultural operations drawn from its source lists—approximately 5,800 addresses. In August 1982, Agriculture Division sent this file to the Economic Census Staff for matching to the latter's list (drawn from the Standard Statistical Establishment List [SSEL]) of multiunits with possible agricultural operations. After matching, the agriculture census multiunits were treated as part of the general multiunit file. Agriculture report forms for multiunits were mailed by the Economic Census Staff, either as individual company packages to companies involved only in agricultural operations, or as part of larger, multiform packages to companies with a variety of economic activities.

Abnormal operations were handled as "must" cases and were sent the appropriate report forms as part of the initial agriculture census mailout.

Labeling—Mailing labels for the census packages were printed by form number in ZIP Code sequence. All of the report forms were labeled by machine. Each package was addressed by applying the appropriate label to the report form through the open window of the form 82-A7A outgoing envelope. Four labeling machines at Jeffersonville carried out this task, at the rate of 10,000 labels per hour each, during the last two weeks of November. (The Bureau carried out the initial mailout for the agriculture census in the last week of December 1982; see ch. 5.)

Table 2. Contents of Initial Mailout Packages

Type	Outgoing envelope	Report form	File copy	Information sheet	Return envelope	Brochure	Cover letter
Nonsample	82-A7A(1st class)	82-AO101 through 82-AO111 (as appropriate) "Blue-green"	82-AO101 through 82-AO111 (as appropriate) "Grey"	82-AO1(I)	82-A8	82-AO1(B)	82-AO1(L1)
Sample[1]	82-A7A(1st	82-AO201 through 82-AO212 (as appropriate) "Yellow"	82-AO201 through 82-AO212 (as appropriate) "Grey"	82-AO1(I)	82-A8	82-AO1(B)	82-AO1(L1)
Must	82-A7A (1st class)	82-AO301 through 82-AO311 (as appropriate) "Green"	82-AO301 through 82-AO311 (as appropriate) "Grey"	82-AO1(I)	82-A8	82-AO1(B)	82-AO1(L1)

[1]Hawaii comprised a "region" by itself, and employed only a sample form, hence there were only 11 regionalized versions of the nonsample and "must" forms.