

# CONTENTS

## Chapter 6. Data Processing

---

	Page
Introduction .....	48
Precomputer Processing .....	48
General Information.....	48
Receipt and Check-In .....	48
Receipt and initial sort .....	48
Batch and check-in.....	48
Mechanical sort .....	49
Manual sort .....	50
Correspondence .....	50
General information .....	50
Reading subunit .....	50
Mailout subunit .....	50
Correspondence analysts .....	51
Interactive processing subunit (keyers) .....	51
Quality control .....	51
Suspense file .....	51
Special Case, "2+," and Multiunit Processing .....	51
Special cases .....	51
"2+" cases.....	52
Multiunits.....	52
Interactive Research Unit .....	52
General information .....	52
Name and address searching .....	53
CFN searching .....	53
Data Entry .....	53
General information .....	53
Data keying operations .....	53
Quality control .....	56
Computer Processing .....	57
General Information.....	57
Format and Simple Edit .....	57
Complex Edit .....	57
Failed Edit Correction and Data Merge .....	58
Failed edit correction.....	58
Data merge .....	58
Statistical Estimates .....	59
General information .....	59
Nonresponse estimation .....	59
Sample estimation .....	59
Tabulation and Data Review .....	60
General information .....	60
Analytical review and data correction .....	60
Tabulations for counties, States, divisions, regions, and the United States .....	60
Disclosure analysis .....	60

## INTRODUCTION

The Census Bureau carried out the 1987 economic and agriculture censuses concurrently, and processed the report forms from the data collection phase, except those for Puerto Rico and the outlying areas (see chs. 7 and 8), in Jeffersonville, IN, at its Data Preparation Division (DPD) facility. While many of the processing activities for the censuses were integrated (e.g., receipt and check-in, and correspondence), separate staffs keyed the data, and each of the various censuses used specialized computerized edit and tabulations programs.

The staff processed the agriculture census data in three main phases:

1. A precomputer operation at Jeffersonville
2. The computer edit and tabulation operation using the Census Bureau's mainframe computer system at the Suitland, MD headquarters
3. A failed edit correction operation, carried out at Jeffersonville concurrently with the computer edit

The Jeffersonville staff also conducted a post-tabulation data review of the census data tables.

The various operations used interactive systems and linked Suitland electronically with work stations and the processing staff in Jeffersonville. The Jeffersonville staff used approximately 70 terminals, each with a keyboard, screen, and access to one of several minicomputers located at the Suitland facility. These minicomputers had substantial data storage and processing capacity; they dispensed with many of the paper printouts previously used by displaying the data on the Jeffersonville terminals' cathode ray tube (CRT) screens.

The main computer facility at Suitland carried out the computer edit and tabulation of the census data, using the mainframe computer to format, edit, and tabulate the data received from Jeffersonville. Census records failing the computer edit were electronically referred back to Jeffersonville and displayed on terminal screens there. The processing staff reviewed the problems and entered corrections to the data file through the terminals. This eliminated the need to write corrections on paper batch edit listings and send them to the data keyers, as had been done for the 1982 census.

## PRECOMPUTER PROCESSING

### General Information

The DPD mailed approximately 4.1 million 1987 Census of Agriculture report forms in December 1987, and carried out a series of mail and telephone followups over the succeeding 7 months. Census responses began arriving at the DPD office in January 1988, with receipts totaling over 1.43 million by the end of that month. The DPD staff required for the agriculture and economic census data

processing activities varied during the processing cycle. Prekeying staff (i.e., for receipt, check-in, sorting, correspondence, and so on) peaked at over 300 employees in February 1988, while the keying staff reached its highest level in April, with approximately 203 employees. The precomputer processing phase involved—

- Receiving and checking in the report forms
- Sorting report forms and removing contents from envelopes
- Evaluating and responding to census-related correspondence
- Reviewing special cases (mostly nonagricultural), "2+" cases, and multiunits<sup>1</sup>
- Keying the data from the report forms to computer disk

The DPD staff was organized into sections and units based on the specific tasks each was to perform. Some units, such as mail receipts and check-in, and the correspondence, processed both agriculture and economic census reports, while other units were dedicated to either the agriculture or economic census processing operation.

### Receipt and Check-In

**Receipt and initial sort**—After identifying incoming mail by the different ZIP Codes assigned to the economic and agriculture censuses, the Postal Service presorted these materials for the DPD office into four categories:

1. Agriculture census receipts
2. Agriculture census postmaster returns (PMR's)
3. Economic census mail
4. Other mail

The clerical staff sorted the agricultural receipts by type, and removed those without visible barcodes. Packages without visible barcodes went to the remove-contents and sort unit; packages with visible barcodes went to the batching unit for check-in; and multiunit PMR's were sent to the multiunit processing unit.

**Batch and check-in**—The batching unit received mail receipts from the initial sorting operation, the opening operation, and the barcode equipment operations areas. Clerks collected receipts and grouped them into batches, by type (i.e., agriculture, economic, and so on); mail receipts still in envelopes were collected in batches of 300-400 pieces each and placed in mail trays, while correspondence receipts and single-unit report forms out

<sup>1</sup>"2+" cases were those for which two or more report forms were received in one envelope. "Multiunits" were generally companies or partnerships that had significant agricultural activities at more than one location, functioning as separate economic entities.

of their envelopes were grouped into batches of approximately 100 each and placed in plastic bags. The staff completed two documents for each batch before sending it along for check-in:

1. A Form EC-14, Batch Cover Sheet, with the batch number, date prepared, check-in status, estimated number of forms and destination after check-in
2. A Form BC-1476, Batch Log, for Check-in Control, for maintaining a record of each batch number.

The check-in unit clerks performed the bulk of the check-in operations using two laser barcode-reading machines, each with a six-pocket mechanical sorter. The barcode-reader terminal operator keyed the batch number and status codes from the form EC-14 for each batch to the laser machine's microprocessor. The computer program checked the codes, and if they were unacceptable, the terminal operator determined the correct codes, and entered those codes for the batch before submitting the materials for check-in.

Once a batch was accepted, the terminal operator placed the materials in the loading tray so that the barcodes faced the laser. The laser equipment "read" the barcodes on the address labels—which included a trade-area code identifying the type of activity (e.g., "agriculture"), and the census file number (CFN—an identification number assigned to each address in the census mail file) for the address—checking in each receipt to update the census control file. The laser machines checked in and sorted up to 400 receipts per minute each (real average production was 100-150 receipts per minute), using the six-pocket sorter to group them into unreadable barcodes (pocket 1), agriculture receipts (pockets 2 and 4), PMR's (pocket 3), all other trade area codes (pocket 5—nonagriculture census receipts inadvertently included in the agriculture census batches), and unable to sort (pocket 6). Initially, the operator resubmitted "unreadables" and "un-sortables" to the laser check-in three times; those still unread after three tries went back to the receipts unit for opening and rebatching. PMR unreadables went to wand/keyboard check-in. As processing progressed, this practice was changed to improve efficiency: In the revised procedure, the operator allowed pocket 1 and 6 materials to accumulate until all of each day's receipts had been run through the machine, then rebatched and reran the rejected receipts. If the sorter still was unable to "read" the barcode, the receipt was opened, batched, and checked in at the wand/keyboard station.

Materials the laser barcode-reading machines could not check in were referred to laser wand/keyboard check-in. These included—

- Unreadable barcode receipts (including those with barcodes not visible through the envelope window)
- Report forms out of their envelopes

- Multiunit report forms when several forms were returned in a single envelope
- Respondent originated correspondence (ROC) and Census Bureau originated correspondence (BOC)
- Out-of-scope recycles

Clerks using laser wand equipment ran the wands over visible barcodes. When the laser wand equipment failed to read a barcode, or when no barcode was present (e.g., with letter correspondence), the clerks used conventional keyboard stations to key the trade-area code and CFN. The staff referred materials without CFN's to the research unit.

As the materials were checked in, clerks entered the check-in data from the laser barcode-reader machines, the laser wand stations, and the keyboard stations, on "pooler" tapes. When each tape reached capacity, or at the end of the daily check-in run, the clerks completed a Form DP-133, Pooler Lot Batch Number Control Record, with the processing project number (a four-digit identification number) and date, project title (i.e., 1987 Economic and Agriculture Censuses), phase (check-in barcode reader), and the pooler lot number. The check-in data were transmitted to the Suitland computer facility daily. The staff sent the paper documents to the pooler lot hold area, to await verification of successful check-in transmission and database update.

**Mechanical sort**—After check-in, the DPD staff used two laser barcode reading machines, each with 24-pocket mechanical sorters to sort report forms still in their envelopes. This required two sorting passes, the first by type of case (i.e., "must," sample, nonsample, and "short" form) and census geographic division (for a description of the census geographic divisions, see ch. 5), and the second by State. In the first pass the laser reader scanned the barcode on each address label and the sorter mechanically sorted the receipts into groups of forms as follows:

Pocket	Type	Geographic division
1	Machine rejects (unable to read barcode)	(X)
2	Must	1, 2, 3, and 6
3, 5, 7	Nonsample	1, 2, 3, and 6
4	Must	4 and 5
6	Must	7, 8, and 9
8, 10	Sample nonmust	1, 2, 3, 6
9, 11	Nonsample	4 and 5
12, 14	Sample nonmust	4 and 5
13, 15	Nonsample	7, 8, and 9
16	Sample nonmust	7, 8, and 9
17	Short form	1, 2, 3, and 6
18	Coverage evaluation	All divisions
19	Short form	4, 5
20	Abnormal farms	All divisions
21	Short form	7, 8, and 9

Pocket	Type	Geographic division
22	Multiunits	All divisions
23	Non-agriculture receipts	(X)
24	Machine failures (unable to sort)	(X)

(X) Not applicable.

The coverage evaluation (pocket 18), abnormal farms (20), and multiunits (22) receipts were not subject to the State sorting operation but were sent directly to the remove-contents and sort unit.

All other receipts already sorted by type of form (i.e., must, sample, nonsample, and short) went through a State sort in three waves by geographic division. Materials in divisions 1, 2, 3, and 6 were in the first wave; cases from divisions 4 and 5 in the second; and those from divisions 7, 8, and 9 in the third. The sorter automatically opened the envelopes during the second sort and grouped receipts, still in their envelopes, by State. Pockets 1 and 24 of the sorter were reserved for rejected materials and machine failures, which were submitted for resorting three times. If they remained unreadable, the staff referred them to the remove-contents and sort unit. As the sorting progressed, clerks collected the sorted materials, by State, and also referred them to that unit.

**Manual sort**—The remove-contents and sort unit received the bulk of materials for processing from the mechanical sorting equipment, on a flow basis, in State batches. The staff removed the contents of envelopes, maintaining the State groups, and sorted the receipts into groups of “2+” reports, agriculture special case reports,<sup>2</sup> and all others. For the “short” form 87-A0400, additional review was performed to identify obvious out-of-scope (O/S) cases. These were short forms with both “no” boxes checked in section 1 and no remarks, correspondence, or data entries anywhere on the form.

Coverage evaluation cases received additional processing; the staff photocopied each coverage evaluation report form, as well as any attached correspondence, and sent photocopies to the evaluation unit. The original report forms then were sorted into groups of “2+” cases, special cases, and all others. The “all others” group was sorted again by type of form (i.e., must, sample, nonsample, and short) and by State.

The remove-contents and sort unit referred the sorted receipts, in State batches, as follows:

Form type	Destination
Coverage evaluation (photocopies)	Agriculture evaluation unit
“2+” cases	Batch for check-in

<sup>2</sup>This occurred when (1) correspondence was received with the report form, (2) the front page of the report form was blank or no positive data were entered, (3) the respondent entered remarks on the front or back page of the form, or (4) acres were entered in section 1 of the report form, but no data appeared in the crops or livestock sections.

Form type	Destination
Special cases	Agriculture special case processing
Multiunit (non-“2+”)	Multiunit processing
Short form O/S recycles	Batch for check-in
All others	Batch for Data Systems Branch

## Correspondence

**General information**—The correspondence unit at Jeffersonville handled the bulk of the census-related correspondence receipts and documented requests referred from incoming telephone calls. The unit consisted of several subunits responsible for (1) reading and classifying correspondence and telephone referrals, (2) keying correspondence category codes to the computer file to take actions assigned by the readers, (3) interactive research (primarily to identify missing CFN’s and duplicate names) in the mail file, (4) mailing standard letters and report forms in response to correspondence or telephone requests, (5) adding addresses to the mail list by assigning CFN’s to newly identified operations, (6) updating the data base and mailing out related blank report forms and instruction sheets to newly identified agricultural operations, (7) handling Census Bureau-originated correspondence (BOC) that required a reply from the respondent and maintaining the suspense file for such cases, and (8) quality control.

**Reading subunit**—The reading subunit staff read and sorted incoming correspondence. The subunit referred BOC materials to the suspense file, sent documents without CFN’s to the research unit, and routed requests that seemed to require a tailored letter in reply to analysts. The readers evaluated all other respondent-originated correspondence (ROC), using a list of problem descriptions and recommended responses for the closest match to the correspondence subject to determine the appropriate corrective action. After reading and evaluating a case, readers annotated the correspondence with a two-digit unit code, a three-digit correspondence category (CORCAT) code, and a time extension date, if required. The unit code designated the processing unit to which the reader referred the case (e.g., “45” = correspondence analysts, “50” = mailout, and “C-” = correspondence category keying), while the CORCAT code identified the general type of problem or subject of correspondence (e.g., “118” = Title 13 quote request, “730” = quit farming, sold farm). The time-extension code showed the date after which the case, if still delinquent, would be included in further followup.

The readers also prepared Form A301, Mail File Update Document, for cases with name and address corrections and sent the A301’s to name and address keying.

**Mailout subunit**—This subunit prepared and typed special request letters and address labels. It also prepared and affixed computer generated or typed address labels for

mailing packages based on CORCAT codes or special instructions. (The computer generated labels were printed in the analysts' office daily, derived from data keyed the previous day.) The unit also received BOC cases, matched them to the suspense file, and referred them to the appropriate analyst for resolution.

**Correspondence analysts**—Analysts received unresolved cases (e.g., agriculture census cases, economic census cases, multiunits, and so on) and cases requiring a tailored reply for evaluation and resolution. Usually, this involved routing the case to the appropriate processing unit or preparing a letter responding to ROC cases. The Jeffersonville staff referred all congressional correspondence to Suitland for response, while Agriculture Division analysts at Jeffersonville handled all other ROC cases. The Agriculture Division staff prepared “standard paragraphs” addressing frequent questions or objections concerning the census and analysts used these to “assemble” letters, modifying the specific wording of the paragraphs to respond to specific points, or drafted entire letters for particular cases.

**Interactive processing subunit (keyers)**—The interactive processing subunit used computer terminals and the interactive processing system to update the computerized census mail list. For the majority of cases, the keyers began by entering the date, then the 11-digit CFN for each case, the CORCAT code, and the time extension code. Up to 30 CFN's and their associated codes could be keyed at a time, after which the data were subjected to quality control checks. A keyer (usually the lead clerk) then was able to write the data directly to the file by pressing two keys on the terminal keyboard. For name and address corrections, clerks had prepared a Form A301, Mail File Update Document, and sent it to the data keying unit, while holding the source document until the computer added the corrections to the file so that corrected address labels would be generated for mail followup. (In some cases, because of particular time constraints, address labels were typed and the letters prepared and mailed directly by the mailout subunit.)

**Quality control**—Quality control (QC) of the correspondence reading operation involved clerical review of materials from each batch submitted to the correspondence reading and keying subunits. The QC clerks verified batches of 60 or fewer pieces on a 100-percent basis, and larger batches on a sample basis varying from 1 in 3 pieces to 1 in 25 (from a random start) with a minimum sample size of 15 items.<sup>3</sup> Clerks reviewed each piece selected for verification and corrected any error identified; batches verified 100 percent were accepted if the error rate was 3 percent or less, while sample-verified batches were

accepted if the number of errors was less than a specific number—ranging from 2 to 5—depending on the size of the sample (e.g., a sample of 15-27 pieces with 2 errors was acceptable).

The QC clerks reviewed each piece of correspondence selected for verification to confirm that (1) it had been processed according to established procedures; (2) the date processed was stamped on the case; (3) each case was correctly annotated with the unit number, category code, and check-in status codes; and (4) all information to be keyed was clearly marked. The clerks corrected any errors, and listed errors identified on the Form DP-697, Clerical Quality Control Record. All batches verified 100 percent were released for further processing, together with acceptable sample verified batches. The sample-verified batches rejected because of too many errors were referred for recycling through the correspondence unit and the QC verification procedures.

**Suspense file**—The suspense file held all ROC cases requiring a reply, regardless of the unit of origin, as well as all BOC cases. Cases remained in the suspense file for a maximum of 35 days; if at the end of that time no additional responses had been received, the cases were referred directly to an analyst who determined what additional action, if any, should be taken.

**Special Case, “2+,” and Multiunit Processing**

**Special cases**—Reports with attached correspondence, remarks entered on the front or back page, blank front page with no positive data, or acres reported in section 1 but no crops or livestock on the report, were designated “special cases” by the remove-contents and sort unit, and were referred to the special case unit. The unit staff (1) determined whether referred cases represented agricultural operations meeting the census farm definition and assigned inscope or out-of-scope codes as required, (2) identified possible “2+” cases, (3) assigned CORCAT codes to cases requiring written replies or more data on the report form, and (4) referred cases to other units or analysts as required. Most special cases proved to be out of scope (i.e., not farms) and were sorted by reason for being out of scope and then sent to the check-in unit for updating the census data base. Once the status had been updated, the check-in unit sent the forms to central files. Report forms determined to be in scope and without other referral problems were sent for data keying. The staff referred the remaining cases as follows:

Type	Referred to—
“2+”	“2+” processing
Problem cases	Agriculture analysts
Form letter assigned	Correspondence category keying
Claims filed	Interactive search

<sup>3</sup>The actual rates of verification by batch size were as follows: 61-99 pieces, 1 in 3; 100-150, 1 in 5; 151-199, 1 in 6; 200-600, 1 in 10; and 601-1,000, 1 in 25. The typical batch contained between 75 and 150 pieces of correspondence.

Type	Referred to—
Report forms, requiring correspondence other than a form letter	Correspondence analysts

Quality control procedures in the special case unit subjected the cases completed by each clerk to sample verification before the work was released from the unit.

**“2+” cases—**Most “2+” cases were the result of mailing two or more report forms to (1) an individual, who may or may not have operated more than one farm; (2) different individuals involved in the same operation (e.g., husband and wife, two or more partners, several heirs to an estate, and so on); or (3) an accountant or trust manager of a bank who returned the report forms for several operations in a single envelope. The “2+” case folders, each containing all the reports forms connected with each case, were routed from check-in to the “2+” unit on a flow basis.

Clerks in the “2+” unit reviewed all referred cases to determine (1) whether the report forms represented one or more separate operations that met the census farm definition, (2) the scope of any operations reported, and (3) whether the report forms needed to be linked for reference during further processing. Clerks linked report forms by CFN if different CFN's were being used for the same operation, or the respondent owned or operated more than one agricultural operation. The reviewing clerk linked operations by first determining which CFN was to be assigned the one-digit primary code, and then assigning a one-digit secondary code to the other reports or CFN's. The linkage codes of the primary CFN characterized the scope of the primary case (e.g., in scope [“1”] or out of scope [“2”]). The linkage codes of the secondary CFN's characterized the status each individual secondary report form/CFN related to the primary CFN (e.g., “5” indicated an out-of-scope CFN linked to an in scope CFN). If three or fewer reports had to be linked, and one was in scope, the clerk involved listed the CFN's and linkage codes on the inscope report for keying as part of the data entry operation. If more than three report forms had to be linked, the clerk involved completed a Form 87-A306, Linkage Document, for the situation. The clerk entered the CFN's and linkage codes for all the report forms on the A306 and added this document to the case folder. Identical linked CFN cases—i.e., the secondary or other duplicate, linked report forms—were annotated “void duplicates” and sent for disposal. The quality control staff verified the “2+” unit's work using procedures similar to those employed for the reading and other precomputer processing units. After verification, the disposition of cases for further processing was as follows:

Type	Disposition
A306 documents and inscope reports	Batch for data keying
Abnormals, Hawaii, Alaska	Agriculture analysts (Suitland)
Multinits	Multinit processing unit

Type	Disposition
Referrals	Agriculture coverage analyst
Correspondence cases	Correspondence unit
Out of scope (except linked secondaries)	Batch for check-in
Out of scope (linked secondaries)	Central files
Void duplicates	Burn box

**Multinits—**The Agriculture Division established multiunit company folders for agriculture multinits identified prior to the 1987 census, and mailed the report forms for these cases as part of the regular census mailout. Analysts accumulated report forms from a multiunit in its company folder until they were able to account for all of its identified agricultural operations, then reviewed the contents of the folder before forwarding them for data keying. The analysts checked the report forms of each company for completeness; reviewed and edited data entries; compared data between sections of the report forms to ensure consistency; and carried out historical data comparisons for land in farms, value of sales, and major commodities for specified operations.<sup>4</sup> The staff made telephone calls to problem cases and respondents who had returned incomplete report forms and held the company folder involved out of the processing cycle until followup was completed. All report forms for a given company were reviewed and corrected before they were released for data keying. Analysts determined what action should be taken to correct the most frequently encountered problems (e.g., bracketed entries (see p. 55), reporting in units other than specified fractional entries, and so on), decided whether specific report forms should be referred to Suitland for resolution of special problems, and completed Mail File Update Documents (Forms A301 and A301A) when needed. After reviewing and correcting any problems, and verifying suspect data, the analysts sent in scope report forms and correction documents for data keying. Jeffersonville analysts annotated out-of-scope (O/S) report forms, made O/S check-in action updates, returned the forms to the company folder, and refiled the folders.<sup>5</sup> The A301 and A301A documents went to the batching unit for name and address keying to update the mail file.

## Interactive Research Unit

**General information—**The research unit resolved coverage problems for selected cases, usually by searching the 1987 census mail file to match names and addresses for the following kinds of cases:

<sup>4</sup>Analysts made historical data comparisons for a particular operation if (1) the reported acreage was 1,000 acres or more, or if the change in acreage from 1982-87 was 1,000-5,000 acres or more; (2) the reported total value of products sold (TVP) was \$500,000 or more; or (3) if a multiunit showed a significant difference in operation from 1982.

<sup>5</sup>The multiunit folders remained on file for possible referral to the economic census processing operation; operations out of scope for the agriculture census could represent other economic operations of a company that were in scope for another of the economic censuses.

- *Successors.* Successors were the current operators of farms listed in the census file under different operators' names. Successors' names for must cases not matched to the census mail list became "adds" to the census file. (Only must cases were subject to successor search.)
- *Partners.* Partners cases were those in which a respondent supplied the name or names of other persons who had filed or would file a report form for the same operation.
- *Claims filed.* There were three types of "claims filed" cases:
  - a. A respondent provided another name but no additional census file number (CFN).
  - b. A respondent did not indicate any other name or CFN under which he or she might have reported.
  - c. A respondent claimed to have reported under another CFN.

Approximately 96,000 cases were sent to the research unit during the processing of the 1987 agriculture census report forms.

Research clerks used interactive routines on computer terminals to search the census data base to try to match cases to report forms already checked in. Analysts in the various processing subunits annotated report forms or correspondence referred to the research unit identified in the upper margins of the first page of the form by symbols indicating the type of referral: "S" for successor, "CF" for claims filed, or "P" for partner.

**Name and address searching**—Research clerks entered into the search routine the last name and ZIP Code, if known, for each case being researched. If the respondent did not provide an address, the clerk entered the original label ZIP Code. The computerized search routine used the SOUNDEX principle (see ch. 3 for information on SOUNDEX) to search the census data base using the reported last name of the operator involved and the ZIP Code of the farm in question.<sup>6</sup> The search program carried out searches at three successive geographic levels—five-digit ZIP Code, three-digit ZIP Code, and State. The system displayed possible matches for the clerks, who then annotated the report forms with identifying CFN's and check-in status codes, and with match codes "M" (matched), "PM" (possible match), "NM" (nonmatch), or "NA" (nonacceptable name (i.e., illegible, obviously fictitious, or government agency)). The clerks referred annotated report forms for further processing, and routed matched ("M") and possible match ("PM") successor cases to agriculture analysts.

<sup>6</sup>When no ZIP Code was provided by the respondent, the research unit used the outgoing address label ZIP Code. When the one from the respondent differed from the label code, the case was referred for ZIP Code research for confirmation of the reported address's correct code.

**CFN searching**—The census file number (CFN) was the primary numerical identifier for each report form or case received and/or processed, and report forms or correspondence received with incomplete or missing CFN's were submitted to the correspondence research unit for searching along with other cases. These cases were matched to the mail file to try to determine the original CFN. When successfully matched, the CFN from the mail list was assigned to the receipt and the case was checked in and routed for further processing. Unsearchable cases were sent to central files, while other problem cases were referred to analysts for resolution.

## Data Entry

**General information**—Data entry (or keying) required transcribing data responses from the census report forms to a machine-readable data file. The DPD staff used a key-to-disk system with interactive edit programs that selected the next program based on a question asked on the previous program, performed preliminary edits, and displayed questions to identify various situations in each record as the data were keyed. The questions helped lead the keyer through the interactive routine.

The data entry staff used key stations, each consisting of a keyboard with a cathode ray tube (CRT) viewing screen, which allowed the operator to monitor and edit keyed data as well as receive messages or questions displayed by the input program. Quality control procedures included reviewing samples of each keyer's work and, when necessary, correcting keying errors.

After data were keyed and verified, a lead operator transferred the data from the disks to magnetic "pooler" tapes for transmission to the Suitland computer complex by telephone datalink. The keying unit received work units (WU's) of census questionnaires batched by State, so each of these "pooler" tapes contained data for only one State. Once the Economic Programming Division (EPD) programmer "ran" the data from a given pooler tape, verifying that the data had been received at Suitland and was acceptable, the DPD unit erased the tape for reuse.

**Data keying operations**—The batching and control subunit weighed report forms using electronic scales (instead of hand counting into batches), batched the report forms by type (must, sample, nonsample, and short), placed each batch of forms in a plastic envelope, and attached a Form 87A405, Batch Cover Sheet, with a WU number assigned by the computerized data entry production control system. The batched reports then were placed in a rolling bin and sent to the data keying staff. The data keying staff in DPD's Data Systems Branch received report forms, linkage documents, and mail file update documents in WU's of 50-100 documents each. Supervisors assigned these materials to the keying staff according to State priorities, or closeout schedules. Agriculture Division requested that Data Systems Branch key and transmit 5,000-10,000 cases per State closeout.

Data entry combined clerical screening and data entry into a single operation. Keyers identified problems on the report forms and made decisions as to whether a given problem should be keyed, flagged, ignored, or handled in some other manner. Keyers opened the plastic envelopes containing the report forms/documents, wrote his/her "keyer ID" on the form 87-A405 cover sheet, and checked the report forms for problems as data were entered. Report forms were pulled from the batch and rejected at data entry for the following reasons:

- *Report form was not keyable.* The majority of data values and their location could not be determined.
- *Linkage code missing.* The linkage code for the extra CFN (i.e., a different CFN than the one in the label area) was missing.
- *Remark requiring reply.* There were remarks on the form, or attached, requiring a reply.
- *Blank report.* No data were reported in sections 1-29. (For telephone followup report forms, no data were reported in sections 1-29 and "Census Use Only" box 037 was not equal to "9"—i.e., the computer replicated selected 1982 data from the census data base.)
- *Maximum values were exceeded.* Data field(s) exceeded the maximum value allowed.
- *Geographic area code (GAC) validation.* The State reported for the principal county location of agriculture operations did not match the State reported in item 5 on the Form 87-A405, Batch Cover Sheet.
- *Check digit failure.* The input edit program rejected the CFN, extra CFN, or "Census Use Only" box 036 check digit after three attempts to key the field.

The keyers pulled rejected report forms from the batch, circling check digit rejects, maximum value failures, and extra CFN's with missing linkage codes in red ink; and wrote other reasons for rejecting the report in the label area (also in red ink). They held rejects aside from the rest of the batch until all report forms in the batch had been keyed, then counted and posted the number of rejects to the A405 Batch Cover Sheet. The keying staff referred rejects to supervisors, who sent them daily to the batching and control subunit for rerouting to the appropriate processing subunit (e.g., correspondence, agriculture evaluation).

Keyers employed a series of input programs to key data from the Batch Cover Sheet and report forms, usually in the following order:

*Batch header.* Assigned to key batch header information from the Batch Cover Sheet.

*Identification.* Key the CFN, extra CFN's, and "Census Use Only" boxes 035-042 from the report forms.

*Name/address correction.* Key corrections made to the name and/or address by the respondent.

*Geographic area code validation 1.* Key answers to questions about the principal county located in section 1, item 8 (Location of Agricultural Activity for "This Place").

*Geographic area code validation 2.* A continuation of GAC validation 1, this program checked county/State location.

*Reported data (29 programs in all).* Assigned to key section identifiers and keycodes with reported data.

*Telephone number.* Key the area code and telephone number located in section 29 on regular report forms, or above section 1 of the telephone enumeration report forms.

*Rejects.* Key the number of rejected report forms.

*Help.* Display the last 15 records keyed for reference.

The header and identification programs identified the batch number, State, number of report forms, CFN, and other relevant information about the batch/report forms. The name/address correction and GAC validation 1 and 2 programs were "interactive," that is, the routines guided the keyers as they keyed the information required. The first of these programs enabled the keyers to make name and address corrections, if necessary; the second was used to confirm county location; and the third to confirm State location if the county reported by the respondent did not match the four-letter county code on the address label, and the State reported did not match the State listed on the Batch Cover Sheet. In each case, the interactive program displayed questions identifying conditions that required action by the keyer, and provided guidance for keying corrections to the names, addresses, and geographic locations.

To enter reported data from sections 1-28 and the telephone number (if any) from section 29, the keyer first keyed the three-digit section identifier for each section containing data,<sup>7</sup> followed by the yes/no response to the question for data. The keyer next entered a three-digit main keycode for each cell containing data (including write-in cells) or a three-digit subkeycode (001-005) for certain data cells with coded crops or animal specialties, followed by the reported data within the data cell (up to nine digits). The section-identifier code also "called" the input program for that section (e.g., the input program for

<sup>7</sup>The keyers did not key section identifiers for blank sections except for sections 1 (acreage in 1987), 11 (land irrigated in 1987), 24 (fertilizers, phosphates, and lime used in 1987), 25 (insecticides, herbicides, fungicides, nematocides, and other pesticides, or other chemicals used in 1987), and 29 (person completing this report and the date) on must and sample report forms; and sections 1, 11, and 29 on nonsample report forms.

section 7 was activated by keying “-07” on the previous program). For example, in section 7 (other crops), a respondent might report that 75,000 pounds of shelled popcorn was produced on 5 acres of land on “this place” in 1987. The keyer entered

1. The section code “-07” (“-” was used for “S” as a section code digit), indicating the section,
2. “1,” indicating the yes box was marked (keyed for data)
3. Keycode “662” to identify the crop as popcorn (pounds, shelled),

4. “5,” reporting acres for data,
5. “001,” the subkeycode for quantity harvested,
6. “75000”—the quantity harvested for data.

Keyers proceeded through the report form, entering the various codes as needed. They were expected to decide whether to ignore, key, or flag data for any problem item, as shown in table 6-1.

Table 6-1. **Keyer Problem Instructions**

Problem	Description	Keyer action
Alpha entries	Respondent used an alphabetic equivalent for a numeric value (i.e., “ten” acres, instead of “10” acres).	Keyer interpreted the value and keyed it in numerics.
Dollars/cents	Respondent reported dollars and cents instead of dollar value only.	Keyer entered only dollar value.
Altered stub (The “stub” was the list of items or descriptions usually running down the left side of a section.)	An altered stub involved a change or addition to the preprinted items.	Keyer flagged this problem by keying a “+” for the data item; no data were keyed.
Bracketed entries	A single entry reported for multiple data cells.	Keyer keyed the reported data followed by a “-” flag.
Data field exceeded nine digits	-	Keyer rejected report form after circling the data cell.
Data reported outside of a data cell	Data were reported but were written outside corresponding data cell.	Keyer keyed the data for the nearest data cell or handled the item as a bracketed entry if the response appeared to be bracketed.
Double entries	More than one entry in a single data cell.	Keyer repeatedly keyed the key code and data until all data values were keyed.
Fractions and decimals	Fractions and/or decimals reported when not requested, or fractions and/or decimals reported when “tenths” requested (i.e., “1/2,” “1/3”).	Keyer decided how to handle based on rules given in the keying instructions.
Range entries	Data reported with an upper and lower limit instead of a specific number.	Keyer keyed upper limit only.
Reference to other data	Symbols such as arrows, ditto marks (“), or remarks (“all”), used to indicate “the same as” reference to other data.	Keyer decided to key or flag, based on instructions and examples in the keying instructions.
Wrong units	Units used in reporting data were inconsistent with those listed in a data cell or preprinted to the right of crop names below write-in sections.	Keyer compared the reported units to the preprinted units and keyed the data if the units were the same, or flagged the data by keying the reported data followed by a “+” if the units differed.
Negative entries	Negative values reported for acres or dollars.	Keyer keyed a “-” flag for the section.

Table 6-1. **Keyer Problem Instructions**—Con.

Problem	Description	Keyer action
Remarks	Comments or reporting errors that (1) required a change to reported data, (2) contained data, (3) related to the manner in which data were reported, or (4) required a reply. Other categories of remarks were (a) illegible entries (not interpreted by supervisor), (b) nonkeyable sections, (c) sections that were full and data were reported below, (d) the section was not distinguishable or there was a question as to how the data were to be keyed.	Keyer keyed the section identifier and the yes/no response, followed by a "+" for data.

**Quality control**—There were quality control (QC) procedures for the data keying operation to ensure that the information on the report forms was accurately recorded for editing and tabulation. For the 1987 census, verifiers checked keyers' work to identify any errors made by the keyers. All errors identified during the verifiers' review of keyed work were corrected and reverified before the data were transmitted to Suitland for computer processing. The procedures defined errors as either keyer errors—essentially miskeying that resulted in such problems as miskeyed fields, field or document omission or duplication, field keyed unnecessarily, and so on—or nonkeyer errors, which included mechanical or supervisor errors, verifier corrections, etc.

Data keyers progressed through four stages of verification of their work. These began with 100-percent review or verification of two batches, followed by sample verification, and then qualification for the final stage of post-proficiency status. After initial training, including instruction on the specific agriculture data entry procedures, keyers entered "stage 1" of the verification program.

In stage 1, which was part of the keyer training program, all keyers' work was verified on a 100-percent basis with all detected errors corrected. Two complete WU's (a WU consisted of approximately 50 sample or nonsample, or approximately 100 short (A400) forms) had to be keyed before the keyers could proceed to stage 2.

In stage 2, keyers' work was sample verified<sup>8</sup> and the quality control reviewers used decision tables that established the acceptable number of errors (again, all errors, keyer and nonkeyer, were counted with respect to acceptability) within specified numbers of data fields verified from the sample questionnaires. For example, a WU with 1,057 to 1,116 data fields verified was acceptable if the total number of defective fields was less than 22, or about

2 percent. (Rejected WU's were sent through a rectification process, which involved 100-percent verification and correction.) Keyers remained on stage 2 until completion of the proficiency training period, at which time they progressed to stage 3.

In stage 3, keyers tried to qualify for the final stage. WU's were sample verified at the same rates as in stage 2, but the allowable error rate was reduced (in a WU with 1,108-1,187 fields, more than 17 defective fields, or less than 1.5 percent, required rejection of the unit). Only keyer errors were counted to determine qualification. If a keyer received 5 consecutive "accept" work unit decisions in a sequence of 10 or fewer decisions, the keyer advanced to stage 4. Each decision represented one keyed batch. (Keyers failing to qualify were further instructed and then allowed to reenter stage 3.)

In stage 4, each keyers' work was sample verified (except for small WU's) and all errors, keyer and nonkeyer, were counted in accepting a keyed WU. Overall error rates for keyers at stage 4 were not to exceed 1.2 percent. Once qualified for stage 4 verification, keyers remained at that stage for the duration of the census processing.

After data keying and verification, the processing staff moved the batches of keyed report forms to a holding area, keeping them there until disposition listings were received showing which records had failed and which had passed the computer edit. The processing staff pulled the report forms for cases that failed the computer edit from the keyed batches and regrouped them into edit review WU's for the interactive edit review and correction process. Thereafter, the forms went to central files for sorting, boxing, and storage.<sup>9</sup>

<sup>8</sup>The WU's were sampled using a sliding scale: small WU's (i.e., 9 or fewer sample or nonsample forms, or 19 or fewer short forms) were verified 100 percent; the sampling rates for larger units varied from 20 percent for those with fewer than 19 questionnaires (39 short forms), to 5 percent for units with 40 (80 short forms) or more.

<sup>9</sup>The 1987 agriculture census report forms remained at Jeffersonville until Sept. 1989, when the DPD staff pulled the report forms for 270,000 "large" farms (the definition of a large farm varied from State to State) and retained these for the Agriculture Division's large farm reference file (used in compiling the 1992 census farm list). The staff sent the remainder to the Federal Records Center, Dayton, OH, where they were to be stored for 11 years after the date of the census. After that, the individual report forms were, by law, to be burned. (Unlike the decennial census of population and housing procedure, the Bureau did not microfilm the agriculture returns.)

## COMPUTER PROCESSING

### General Information

The data, once clerically processed and keyed to computer tape (as described above), were transmitted to the mainframe computer facilities at Suitland, MD, which carried out the bulk of the data formatting, sorting, editing (validating, cross-checking, and refining the data file), and tabulation, while the Jeffersonville staff did most of the data review and correction, using the interactive minicomputer system.

This computer work cycle began as soon as the first data from report forms were keyed and transmitted to Suitland in mid-January 1988, and continued until the final tabulations were completed in March 1989. Approximately 2.4 million individual census records were edited, of which some 2.088 million met the agriculture census's farm definition and were included in the agriculture census file. The computer processing cycle consisted of three major operations:

- Formatting and simple editing
- Complex editing and edit correction
- Data tabulation

### Format and Simple Edit

The format and simple edit operation (1) converted individual data records into binary records that could be manipulated by using the data processing programs, (2) carried the informational flags set during keying operations to the computerized record, (3) added historical individual-cell data to the file for comparison purposes, and (4) "flagged" problems identified during the formatting process.

The data records created for each census report form during data entry contained "fixed" record layouts that, for computerized editing and tabulation, had to be converted to "variable" output records with binary coding for numeric values. The data entry format program converted the data records into a series of fixed and variable portions; the fixed ones contained each record's identification information—State and county codes, CFN, SIC code, and so on—while the variable portions included a field for each data item reported, imputed, or changed, but nothing for items left blank in the original record. The computer recognized the individual data items from keycodes at the beginning of each segment, and ignored blank segments.

The format and simple edit program carried the flags set during the data entry to the computer records, but also established new flags for problems identified during the formatting cycle, such as (1) illegal geographic or report-form codes, (2) cases with no reported sales or livestock inventory, and (3) cases with individual items flagged (i.e., illegal keycodes, invalid crop codes, etc.).

The computer also added historical data for individual items to the file at this time. Later, the computer program checked the information in the record against these historical data for completeness and reasonableness. Cells or records identified through a comparison program as incomplete, or that exceeded established limits, were flagged and displayed for analyst's review.

### Complex Edit

The format and simple edit program converted the raw data records into binary codes and flagged selected problem cases. The edit programs used were capable of carrying out several thousand individual operations in all, although usually only a fraction of this number were required for editing any particular record. Agriculture subject matter specialists wrote and transmitted the computer edit specifications to the computer programmers using "decision logic tables (DLT's)." Each DLT was a tabular display of the elements comprising a specific edit operation from its inception to the solution.

The computer processing staff carried out the complex edit by State, in batches consisting of formatted records sorted within State, by county, and CFN. The edit checked each record in the data file and

1. Determined whether it represented an agricultural operation meeting the agriculture census farm definition and deleted out-of-scope operations from the data file.
2. Supplied missing entries based on similar farms within the same county.
3. Assigned farm classification codes needed for tabulating the data, including acreage, tenure of operator, value of agricultural products sold, type of organization, and standard industrial classification (SIC) code (by type of farm).
4. Reconciled acres reported for individual items with the total acreage reported.
5. Checked consistency between and within sections of each record.
6. Checked values for products sold, using average prices in each State for each production item, and substituted calculated values if the report values exceeded acceptable limits.
7. Identified nonsample records representing farms that met the "certainty" criteria established for each State, and converted these records to sample records. (Certainty criteria varied by State from minimum acreages of 1,000 to 10,000, or minimum sales of \$40,000 to \$100,000. Institutional and other special cases were included in the certainty classification—even if they failed to meet the other criteria, as were

all farms in counties that had fewer than 100 farms in the 1982 census.)<sup>10</sup>

8. Identified and "flagged" cases with substantial computer generated data changes for clerical review and verification.

The edit also identified and retained in the data file records for agricultural operations that *normally* would be expected to have sufficient agricultural sales to qualify as farms, but failed to do so, for whatever reason, during the census reference year. The edit tested the records for such places against criteria developed to identify agricultural operations that normally would meet the farm definition.

## Failed Edit Correction and Data Merge

**Failed edit correction**—The Jeffersonville processing staff used the interactive edit referral/data correction system to carry corrections to the data file. The mainframe computer at the Suitland, MD, facility began editing agriculture census records in January 1988, reviewing each data record, comparing recorded data item responses to established specifications, and identifying any problems. The edit established a failed edit file, and listed there each record that failed the edit, together with the particular item(s) that had failed. By mid-February, the Agriculture Division had begun reviewing and correcting these cases.

The Economic Programming Division (EPD) staff in Suitland moved data for records that failed the edit to the minicomputer system, using a software system that electronically linked the mainframe and minicomputer systems at the Suitland facility. Two data files were established in the system, one for interactive processing, and a second external to the interactive programs. The EPD staff programmed the mainframe computer to organize new WU's of failed edit cases, which then were moved to the minicomputer system. The maximum edit review WU size was 99 cases and the WU's were established as follows:

1. Cases reporting \$1 million or more in value of sales of agricultural products and/or 30,000 acres or more in place.<sup>11</sup>
2. Must cases (both preidentified, i.e., with estimated sales qualifying them as must cases in the initial mailout, and nonmust cases with reported sales qualifying them for inclusion in the must category).
3. Sample and nonsample cases.
4. Short form cases.

<sup>10</sup>The Census Bureau obtained the data required for these conversions by correspondence with the addressees involved, or by imputation based on responses from farms of similar size in the same geographical area.

<sup>11</sup>At the beginning of the edit correction work, Agriculture Division staff in Suitland reviewed and corrected these "\$1 million" cases, but once the interactive systems staff had begun working on the corrections, they were referred to Jeffersonville for processing with other failed-edit cases.

The EPD staff printed out disposition listings of the CFN's and other identifying information for all cases in each WU. The Jeffersonville clerical staff located and removed the report forms for the failed records from the original keying work unit, placing all the report forms for each WU in a single batch so that the questionnaires involved would be available to the edit review processing clerks.

To begin the review and correction of the unacceptable records, processing clerks in Jeffersonville retrieved assigned WU's from the failed-edit file using interactive terminals linked to a minicomputer in Suitland, and called up the first record in each for display on their terminal screens for review, comparison with the original report form when necessary, and correction. The clerks worked through each record, making any corrections identified on the screen display and reviewing each record after correction before proceeding to the next.

After all the records in a WU had been reviewed and any necessary corrections had been made, the clerk released the WU for quality control review. Each edit correction clerk's work was subjected to quality control review. During the first 2 weeks of activity, quality control staff reviewed a sample of each clerk's work. There were no quality control requirements during this period, as the clerk was considered to be undergoing training. After 2 weeks, quality control staff began sample verification of each clerk's work for production. During the decision period, samples from 10 consecutive WU's were checked, and 8 had to be of acceptable quality (samples checked depended on the size of the particular WU, varying from 1-in-6 for WU's of maximum size (99 cases) down to 1-in-3 for units of 27-53 cases; WU's with 26 or fewer cases were verified 100 percent). After quality control procedures were completed, the WU was released for further processing by EPD.

Each day, EPD staff transferred the corrected files from the interactive file to the mainframe's failed edit file. The corrected records were matched to the original failed records and the latter were deleted from the data file. The "new" records then were reedited. This cycle of editing and correction continued until each record passed the computer edit and could be incorporated into the data file. The failed-edit correction program began in mid-February 1988, and continued for over 8 months, until the end of October 1988. During that period approximately 738,500 individual census records, including 45,000 "repeat edit failures," failed the computer edit and were referred for correction. No record was added to the passed-data file until all corrections had been incorporated and the computer edit record accepted.

**Data merge**—After editing and failed-edit correction, the data files for each State were merged into a single file, in sequential order by State, county, and identification number. Using a "merge" program, the computer tallied agricultural operations by size (i.e., acreage, head of livestock, etc.), value of products sold, and type (used to aid in adding data for nonrespondent cases), and identified and

displayed problem cases for review and correction before tabulation. The processing staff unduplicated the computerized main data file, using a census file number (CFN) matching program to identify and display duplicate records for review. Usually the first of any duplicate records identified was retained, while the rest were deleted from the record.

Statistical Estimates

**General information**—About 13.8 percent of the addressees on the 1987 agriculture census mailing list never responded, and selected data were collected from only a sample of all farms on the list. Nevertheless, the data published from the 1987 census represent all farms in the United States because the Census Bureau used statistical estimation procedures to inflate the respondent data to compensate for nonresponse and the use of sampling.

**Nonresponse estimation**—The Bureau carried out the nonresponse weighting operation editing the data files. Addresses on the census mail list were classified as representing “large” (i.e., with \$100,000 or more in expected annual sales or with 1,000 or more acres (the acreage requirement) varied by State) or “other” agricultural operations. There was an intense followup of nonrespondent large farms, including a telephone followup beginning in February 1988. In April 1988, the agency began selecting a sample of approximately 27,000 of the “small” nonrespondent addresses for inclusion in the Nonresponse Survey, carried out in the following 3 months. The survey data were used to compute stratified State-level estimates of the number of nonrespondent cases that actually represented farms. (The telephone followup and Nonresponse Survey are covered in ch. 5.)

The strata were defined by form type, expected value of sales, and previous census status. The estimated survey proportion of farms was multiplied by the number of census nonrespondents to estimate the number of census farms among the census nonrespondents within each stratum. The number of nonrespondent farms was proportionately distributed to each county within each stratum; and a systematic sample of respondent farms was selected to represent the nonrespondents. This was done by assigning a “nonresponse weight” of “2” to the selected records (i.e., the data responses for each selected record were doubled). The “large” nonresponse cases and records not selected were assigned nonresponse weights of “1.” This procedure assumed that the respondent and nonrespondent farms had similar characteristics, such as value of sales, acreage, and so on.

**Sample estimation**—The 1987 agriculture census collected selected data (items 23-28 on the sample report forms) only from an approximate 20-percent sample of the agriculture census mail universe. The following types of addresses received the sample form:

1. All addresses in Alaska and Hawaii.
2. All “must” cases.
3. All addresses expected to represent “large” farms (the definition of a large farm varied by State, from a minimum of 1,000 acres or \$40,000 in sales in New England, to as high as 10,000 acres or \$200,000 in sales in some Western States).
4. All farms in counties with less than 100 farms identified in the 1982 census.
5. A systematic sample of 1-in-2 addresses for counties that contained 100 to 199 farms in the 1982 census.
6. A systematic sample of 1-in-6 addresses of counties that contained over 200 farms in the 1982 census.

The sample data estimates were designed to estimate the totals that would have resulted had all census respondents been asked for the data requested in items 23-28 of the report forms. The staff used a ratio-estimation procedure to assign a sample weight to each record. This inflated the sample data to represent all farms in the subject population.

The sample records were classified into “certainty” farms (i.e., mostly large) and “noncertainty” (mostly small) farms. The certainty farms were defined as the first four types of addresses listed above, and were assigned a sample weight equal to “1.” To calculate estimates for the noncertainty sample addresses in the fifth and sixth listings above, the addresses were partitioned into 32 mutually exclusive strata, formed by 8 sales groups, in turn divided by 2 standard industrial classification (SIC) code groups, then by 2 acreage classifications, as follows:

Value of sales	SIC	Acres
\$1 to \$999	01 all crops	0 to 69
\$1,000 to \$2,499	02 all livestock	70 or more
\$2,500 to \$4,999		
\$5,000 to \$9,999		
\$10,000 to \$24,999		
\$25,000 to \$49,999		
\$50,000 to \$99,999		
\$100,000 or more		

Each farm record was assigned an initial weight equal to the ratio of the total farm count to the sample farm count for the stratum containing the sample farm. Where necessary, the procedures combined strata to increase the reliability of the final estimates. (The staff used a specific “collapsing” pattern to combine strata that (1) contained less than 10 sample farms, or (2) had a ratio of total farms to sample farms that was more than twice the mail sample rate—1 in 2 or 1 in 6.) The resulting total and sample farm counts were used to compute the sample weights to assign to the record.

The final weight assigned to a sample record was the product of the nonresponse weight and the sample weight. The totals for the sample data were calculated by multiplying the reported sample data values by the final weight.

## Tabulation and Data Review

**General information**—After editing, correction, and data merge, the data records were ready for tabulation. The computer tabulation programs produced a series of detailed data matrices, each consisting of several thousand different items, that would provide the basic material for most of the data tables drawn from the census file. The staff used the data matrices to extract analytical data for analysis and correction in a detailed county-level format.

Once the analytical tabulations had been reviewed and any problem records corrected, a listing was produced showing any change expected. Once the change listings were reviewed, the data matrices were retabulated to serve as the data source for the *Advance Reports* and the Volume 1, *Geographic Area Series*, reports.

**Analytical review and data correction**—The Agriculture Division staff at the Suitland headquarters used county-level analytical tabulations to interactively review the aggregated data on the minicomputer system. All the data items reported on individual report forms were tabulated for each county and State, for all farms and for farms with \$10,000 or more in reported value of agricultural product sales. The analysts used historical data from the 1982 agriculture census to review the 1987 data for reasonableness and accuracy. They employed analytical tables developed for review purposes as their basic review reference documents, but also used related data, drawn mostly from USDA estimates. The analysts used the interactive system on the minicomputers to electronically search the data file for records containing the questionable data and recommended corrective action.

Representatives of the USDA's National Agricultural Statistics Service (NASS) reviewed the analytical tables produced during the review, as well as the analysts' criticisms and recommendations. The NASS reviewers identified any additional problems they found in the tabulations, and suggested additional corrections or alternative solutions to problems previously noted.

The Jeffersonville staff carried corrections to the data records and all deletions from the data file using the interactive minicomputer systems. Agriculture Division staff reviewed all corrections for accuracy and to ensure that the data criticisms cited for changes were satisfied. The data then were released for tabulation.

**Tabulations for counties, States, divisions, regions, and the United States**—The Bureau drew the county and State tables directly from the data matrices, and State cross tabulations from the detailed data file itself. Using the tabulating programs, it summed the State totals to produce data for census geographic divisions, regions, and the United States. The published tables included historical data from the 1982 and 1978 censuses. The tabulation program for the 1987 census had "advance" tabulations of selected data from each State and county file for the *Advance Reports*, followed by the general tabulations for the Volume 1, *Geographic Area Series*.

**Disclosure analysis**—Title 13, United States Code—Census, prohibits the publication by the Census Bureau of data that could be used to identify any individual respondent to any of its censuses. The agency used a procedure called "disclosure analysis" to maintain the confidentiality of the data by reviewing all the data tables before releasing them for publication. This procedure identified and suppressed data items, the publication of which (1) would result in the direct disclosure of data reported by a particular respondent individual or company, or (2) would reveal information about an individual by derivation—that is, by the user adding or subtracting a published subtotal from a published total to reveal individual data. However, the number of farms associated with a particular data item was not considered a disclosure of confidential information in itself; only the associated data values were. While most of the disclosure analysis was carried out by computer, the automated equipment and programs could not perform the entire analytical function. Agriculture Division staff had to check many tabulations and cross tabulations manually before the data tables could be sent for publication.

As a general rule, the Bureau did not publish any agriculture census data for counties with 10 or fewer farms in the census reference year. The disclosure procedures set minimum numerical limits for publishing data values for farms reporting a particular item. Since the tables included identical information arranged under several different classifications, the identification and suppression of a figure in one table required reviewing all related tables and the suppression of the relevant datum in each of them.

Disclosure analysis and suppression for the 1987 agriculture census tabulations for the 50 States, and the 3,079 counties or county equivalents, was completed in July 1989, and for the U.S. summary volume in November 1989.