

# CHAPTER 6.

## Data Processing

	PAGE
INTRODUCTION .....	35
PRECOMPUTER PROCESSING .....	35
GENERAL INFORMATION .....	35
RECEIPT AND CHECK-IN .....	35
CORRESPONDENCE .....	36
SPECIAL CASE, "2+," AND MULTIUNIT PROCESSING .....	37
INTERACTIVE RESEARCH .....	38
DATA ENTRY .....	38
COMPUTER PROCESSING .....	40
GENERAL INFORMATION .....	40
FORMAT AND SIMPLE EDIT .....	40
COMPLEX EDIT .....	41
STATISTICAL ADJUSTMENTS .....	42
TABULATION AND DATA REVIEW .....	42

### INTRODUCTION

The Census Bureau carried out the 1982 Census of Agriculture concurrently with the 1982 Economic Censuses and combined all phases of the processing operations that did not require specialized knowledge or expertise. Receipt, check-in, and much of the correspondence were handled by a combined staff at the Bureau's Jeffersonville, IN, processing center. Data entry was carried out by a separate staff, and the computerized processing of the data required specialized edit and tabulation programs for the various censuses.

The data collected in the agriculture census were processed in two phases—a "precomputer" operation carried out primarily at Jeffersonville and a "computer" operation using the main equipment at Bureau headquarters in Suitland, MD. The Jeffersonville phase of the processing activities included receipt and check-in of the report forms, correspondence for the census, keying the data (i.e., data entry) from the report forms to computer disks, then transferring it to tape for transmission to the main computer in Suitland, and review of problem cases from both the data-collection operation and the computer edit. Computerized processing at the Bureau's main facility in Suitland involved formatting, editing, and tabulating the data received from Jeffersonville. The Bureau expanded the use of automated and computerized equipment over that of previous censuses even in the "precomputer" phase of the processing. There, it employed automated barcode check-in equipment (introduced in the 1977-78 censuses) and mechanical sorters for much of the check-in sorting, and replaced the old microfilm research units with interactive computer terminals.

### PRECOMPUTER PROCESSING

#### General Information

The Bureau mailed 3.65 million agriculture census report forms from Jeffersonville in December 1982, and carried out a series of followup mailings over the next 7 months. Responses to the census began arriving at the Data Preparation Division (DPD) in Jeffersonville in January 1983. Receipts early in the census cycle reached as many as 312,000 pieces of mail per day. At the peak

of the processing cycle, approximately 950 employees worked on the combined processing operation at Jeffersonville.

Incoming mail, presorted by the Postal Service, was checked in and sorted using high-speed laser scanners and mechanical sorters. Various special units handled correspondence and research on problem cases.

After keying, the data were transferred from disk to magnetic tape for transmission to Suitland. The report forms were held until the Jeffersonville unit received disposition listings showing any problem cases. Clerks and/or analysts resolved any problems, sending corrections for keying.

#### Receipt and Check-In

**Receipt and initial sort**—The Postal Service presorted incoming mail for the Bureau's Jeffersonville facility into four major categories—agriculture census receipts, agriculture postmaster returns (PMR's), mail for the economic censuses, and other mail.<sup>1</sup>

The DPD clerical staff quickly fanned through agriculture receipts in the mail trays, faced receipts in the same direction, sorted them by type, and removed receipts without visible barcodes. Packages with visible barcodes, including PMR's, were sent to the batching unit. Receipts without visible barcodes were sent to a clerical unit where they were opened and the contents sorted. Report forms with and without attached correspondence, Bureau-originated correspondence with a CFN (census file number) present, "2+" cases,<sup>2</sup> agriculture multiunit report forms with adhesive barcoded labels, and out-of-scope recycles (primarily from special-case processing) all were referred to the batching unit. Congressional correspondence was referred to supervisors for disposition, and Bureau-originated correspondence without CFN's went to the suspense file, while other materials without CFN's were sent to the research unit.

**Batch and check-in**—The batching unit received materials from both the initial sorting area and the opening and sorting unit. Clerks collected receipts, still in their return envelopes, into batches of 600-700 pieces each, and grouped out-of-envelope materials into batches of approximately 100 pieces each. Each batch was placed in a plastic mail tray with a form EC-38 Batch Cover Sheet attached, on which the clerks listed the date, assigned destination of the batch, the estimated (batches going to laser check-in) or actual (batches referred for check-in or completeness/coverage keying) count and type of the pieces in the batch, a seven-digit batch number, and a two-digit check-in status code. The status codes, batch number<sup>3</sup> and date

<sup>1</sup>The Postal Service identified the Census Bureau's incoming mail by ZIP Code. The agricultural returns had a ZIP Code different from the economic returns. The Bureau estimated that presorting the incoming mail saved about \$250,000.

<sup>2</sup>Cases in which two or more single-unit report forms were enclosed in the same envelope, or in which additional CFN's were written in a designated area on the report form.

<sup>3</sup>Assigned sequentially within each status code group.

transmitted for each batch were entered in the Batch Log for Check-in Control, and the batch was referred for further processing.

Three laser barcode-reading machines, each with a six-pocket mechanical sorter, performed the bulk of check-in, since about 75 percent of all receipts were report forms in return envelopes with visible barcodes. The laser machine terminal operator entered the batch and status codes of each arriving batch for check-in in the laser machine's microprocessor. The computer checked the codes, and if they were unacceptable, rejected the batch. In that case, the terminal operator determined the proper codes and entered them in the file, using the keyboard terminal.

The laser machines checked in and sorted up to 400 receipts per minute each, using the six-pocket sorter to group them into unreadable barcode cases (pocket 1), agriculture census receipts (pockets 2, 4, and 6), PMR's (pocket 3—these were inadvertently included in batches of agriculture census receipts), and all other trade-area codes (pocket 5—these were nonagriculture census receipts inadvertently included in the agriculture census batches). Batches of PMR's were run through the laser machines and checked in separately from census receipts. Clerks at 15 laser-wand stations checked receipts referred from the opening and sorting area—such as 2+ reports, report forms with attached correspondence, and agriculture multiunits with adhesive barcoded labels, as well as cases rejected by the laser machines. The laser equipment "read" the barcodes on the address labels—which included a trade-area code (i.e., a code identifying the type of activity—agriculture, retail trade, manufactures, and so on), CFN for the address, and the processing sort code—checking in each receipt, and capturing and storing the required check-in data for use in updating the census nonrespondent file. The check-in unit held the agriculture census receipts until it received control listings and resolved any problems with the transmitted and checked-in data. Receipts then were referred for further sorting.

**Mechanical sort**—After bar-code check-in, the Bureau used laser barcode readers and two 24-pocket mechanical sorters to sort report forms still in their envelopes. The sort was carried out in two "passes"—a primary sort by type of case and census geographic division<sup>4</sup> and a secondary sort by State. In the primary sort, the laser reader scanned the barcode on each address label, and the sorter mechanically sorted the receipts as:

Pocket	Case and geographic division
1	Unable to read barcode
2	Master processing sample, all divisions
4	Abnormal, all divisions
6	Must, divisions 1, 2, 3, 6
8	Must, divisions 4, 5
10	Must, divisions 7, 8, 9
12, 14	Nonmust sample, divisions 1, 2, 3, 6
16, 18	Nonmust sample, divisions 4, 5
20, 22	Nonmust sample, divisions 7, 8, 9
3, 5, 7	Nonsample, divisions 7, 8, 9
9, 11, 13	Nonsample, divisions 4, 5
15, 17, 19, 21	Nonsample, divisions 1, 2, 3, 6
23	Incorrect trade area code
24	Machine failure, unable to sort

Cases with unreadable bar codes were resubmitted for laser sorting three times. If they were not successfully identified and sorted they were collected and referred to the Remove Contents and Sort Unit, which also received, after the primary sort, abnormal cases, and those with incorrect trade-area codes.

<sup>4</sup>See ch. 2, p. 9 for a map.

The remaining sorted cases were submitted for the secondary, or State-level, sort by type of case (must, nonmust sample, non-sample) and geographic division in three waves: cases from divisions 1, 2, 3, and 6; cases from divisions 4 and 5; and cases from divisions 7, 8, and 9. The sorters automatically slit open the envelopes during the secondary sort, and the sorted receipts, still in their now-opened envelopes, were collected and referred to the Remove Contents and Sort Unit.

**Manual sort**—Report forms that had been removed from their envelopes and checked in at the wand/keyboard stations were referred to the Remove Contents and Sort Unit and manually sorted into the same categories described above in the mechanical sort. The primary function of the Remove Contents and Sort Unit involved removing report forms from their envelopes after barcode check-in, scanning the forms and designating them as either special cases (i.e., report forms with attached correspondence, no positive data on the front page, or remarks on the front or back pages), or keyable report forms. Special-case report forms were referred to the Special Case Processing Unit; report forms in the master processing sample, "abnormal" report forms, and report forms from Hawaii and Alaska were referred to Agriculture Division; and keyable report forms were sent to be batched for data keying.

## Correspondence

**General information**—The Correspondence Unit at Jeffersonville handled the bulk of census-related correspondence received. The unit consisted of five major subunits responsible for (1) reading and classifying correspondence; (2) analysis of special and problem cases; (3) interactive processing (keying); (4) the suspense file; and (5) typing labels or filling special requests.

**Reading subunit**—The Reading Subunit read and sorted all incoming correspondence: Bureau-originated correspondence (BOC) was referred to the suspense file, requests that seemed to require a tailored letter in reply were routed to an analyst and documents without CFN's to the interactive search unit, and other respondent-originated correspondence (ROC) was evaluated to determine the appropriate action. For ROC cases, readers referred to a list of problem descriptions and actions for the closest match to the subject of the correspondence. The readers annotated the correspondence with the two-digit unit code, a three-digit category or other referral code, and a time extension date if required. The two-digit unit code designated the processing unit to which the case was to be referred (e.g., "45"= interactive research, "47"= analysts, "50"= mailout), while the three-digit category code identified the general type of problem or subject of the correspondence (e.g., "118"= Title 13 request and agriculture question on confidentiality, "200"= better physical location information needed). The time extension date code showed the date after which the case, if still delinquent, would be referred for further followup. (Normally, these extensions were set for the response cutoff date immediately following receipt of the request, generally 4-6 weeks.)

Quality control in the reading unit involved verification on a 100-percent basis of all batches of 60 or fewer pieces and sample verification of all other batches. The sampling rate for batches with over 60 pieces varied from 1 in 3 pieces (from a random start) to 1 in 10, depending on the size of the batch, with a minimum sample size of 15 items from each batch. Clerks

reviewed each piece selected for verification and corrected every error detected. Batches verified 100-percent were accepted if the error rate was 3 percent or less, while sample-verified batches were acceptable if the number of errors was equal to or 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 of correspondence with 2 errors was accepted; a sample of 47-66 pieces was accepted if only 4 errors were detected). All batches verified on a 100-percent basis were released for further processing along with sample-verified batches accepted by the quality-control clerks. Rejected batches were recycled through the correspondence unit process and the reading quality-control procedures. The reading clerks' performance was evaluated on the basis of the rejection rate for batches processed, using blocks of 10 consecutive batches as a "decision line" for each clerk's performance. To maintain their work status, clerks had to have two or fewer batches rejected in a decision line; a clerk whose rejection rate exceeded this was retrained.

The reading subunit also checked single-unit PMR cases with name and address corrections to make certain that the information needed for keying—the CFN and the name and address change—were clearly readable. Readers referred first-time single-unit PMR cases to the interactive processing subunit for keying.

**Correspondence analysts**—Unresolved cases (e.g., multiunits, economic census cases, and so on) and cases requiring a tailored reply from the Bureau were referred to a correspondence analyst for evaluation and resolution. Generally, this involved either routing the case to the appropriate processing unit or preparing and mailing a letter responding to the incoming correspondence. The analysts made maximum use of "standard paragraphs" prepared for use in written responses to frequent questions or objections concerning the census. Analysts could "assemble" letters from these paragraphs, modify the paragraphs as required to address a point more specifically, or draft an entire letter to respond to a particular case. Analysts in Jeffersonville handled agriculture census correspondence, while economic censuses correspondence was referred to the Economic Census Staff in Suitland.

**Interactive processing subunit (keyers)**—The interactive processing subunit (not to be confused with the Interactive Research Unit [see p. 38]) used computer terminals with access to the Bureau's computerized mail list to update that list by keying the identification data and action codes—i.e., CFN, correspondence category code, time extension code, name/address corrections, and so on—to the census mail file. The computer programs automatically incorporated name and address corrections so that corrected address labels would be generated for mail followup, held addresses that had received time extensions *out* of the followup file until the expiration of the extension, and so on.

The interactive processing subunit's work also was subjected to quality control. Batches were verified on the same general system used for the reader verification (i.e., 100-percent for batches of 60 or fewer cases, sample verification of all others). Verifiers used an automated quality-control system to carry out the actual review of cases; the verification clerk identified each case to be reviewed, and the computer program identified and displayed each error detected on a printout. A batch was accepted if it had a critical-error rate of 3 percent or less (a "critical" error was one that would lead to misidentification of

the specific case [e.g., CFN or SIC code error]), or an overall error rate of 5 percent or less. Rejected batches were recycled through the interactive processing subunit until they met the quality control standards.

**Suspense file**—The suspense file held all ROC cases requiring a reply, regardless of the unit of origin for the specific case.

## Special Case, "2+," and Multiunit Processing

**Special cases**—The Special Case Unit reviewed report forms and other documents referred to it by the sorting and check-in keying staffs. The clerical staff (1) reviewed report forms and other materials to determine whether or not they represented farms, and assigned in-scope or out-of-scope codes as required; (2) assigned correspondence category codes for cases requiring reply or to avoid double handling by correspondence readers; and (3) referred cases as necessary. In-scope cases not requiring replies were sent for data keying; problem cases went to the agriculture analysts; successors, partnerships, and claims-filed cases were referred to the Interactive Research Unit; and cases that seemed to require a tailored letter were referred to correspondence analysts. Nonfarm cases were sorted by reason for being classified as out-of-scope, and referred either to the staff analyst (further processing for these cases depended on the results of research by the analyst), or to check-in.

**"2 + " cases**—The "2+" Processing Unit received material on a flow basis from both the check-in and correspondence staffs. Most 2+ reports resulted from (1) two reports returned by the same individual who may or may not have operated two farms (e.g., one report mailed by J. S. Jones, and a second by John Jones, who was the same person); (2) two or more report forms mailed to two individuals involved in the same operations (e.g., husband and wife, two or more partners, several heirs of an estate, and so forth); or (3) two or more report forms mailed to an accountant or trust manager who returned multiple forms in a single envelope. Clerks in the 2+ Processing Unit reviewed each case to determine whether (1) the report forms represented a single operation or multiple operations, (2) each report form represented a farm, and (3) linkage of the individual report forms was needed for further processing. Linkage was required if (1) different CFN's had been assigned to the same operation, (2) one of the report forms showed an in-scope tenant or successor accounting for an operation listed as in scope in 1978 but reporting as out of scope for 1982, or (3) one respondent was shown to have ownership or operator interest in more than one operation. Report forms were linked by the reviewing clerk, who assigned a single-digit linkage code to each form and entered it in the "extra CFN" space on the form. If more than three report forms had to be linked, a form 82-A306 Linkage Document was completed with the CFN's and linkage codes for all the report forms involved and was added to the case folder. In each case, the clerks assigned a "primary linkage" code to one CFN (i.e., a report form with the selected CFN) and "secondary linkage" codes to the other report forms/CFN's in the case. The primary code served to characterize the entire case, while the secondary codes linked each individual report form/CFN back to the original CFN.

The 2+ Processing Unit's work was verified using procedures similar to those employed for the reading and other precomputer processing units. After verification, the clerks referred cases for further processing using a set of tables describing actions to be taken in specified circumstances.

**Multiunits**—The Agriculture Division established 1978 and 1982 multiunit company folders for agriculture multiunits identified prior to the census, but report forms for these cases were included in the multiunit packages mailed out as part of the general mailing for the 1982 Economic Censuses, and followup was carried out by the Economic Census Staff as part of the regular followup to economic multiunit cases. Agriculture report forms returned by multiunits were identified during the completeness and coverage keying phase of the economic census processing operation. Completeness and coverage keying involved verification that (1) all the establishments of a multiunit company had been covered in the census, (2) essential data had been reported for each establishment, and (3) the data were accurate (determined by comparison with historical data). Data keyers used interactive computer terminals and a screening program to check individual report forms. Since the agriculture census report forms used for multiunit establishments did not request the same kinds of payroll and employment data as did the economic census forms (the latter asked for the same kind and detail of information as required by the IRS form 941 tax returns), these items were left blank on the computerized record for each agriculture case.

The edit programs used in the economic censuses to check the completeness of the records rejected the agriculture records as incomplete and displayed both the record rejected and historical data. Thus the edit programs automatically identified the agriculture cases. All agriculture-related multiunit cases identified during completeness and coverage keying were referred to the agriculture staff for processing and tabulation as part of the agriculture census. Agricultural multiunits identified during the agriculture census itself, rather than before, were enumerated as part of the agriculture census telephone followup.

Agriculture census analysts accumulated report forms from a multiunit in its company folder until they had accounted for each of its identified agriculture operations, and then reviewed the materials in the folder before sending them for data keying. The analysts checked the completeness of individual report forms for each company, reviewed and edited data entries, and identified and corrected any duplication among various agricultural operations of each company. Incomplete report forms were referred for telephone followup and the company folder involved was held out of the processing cycle until followup was completed. All the report forms for a given company were reviewed before release for data keying. Report forms from operations determined to be nonfarms were identified as such by the analyst with a final disposition code of "5X" written on the address label. Company folders in which all the report forms were out of scope—that is, none of the report forms were from operations qualifying as farms under the Bureau's definition—were identified by a strip of black tape applied along the top edge after review. Folders that included at least one in-scope report form had a strip of green tape applied. All in-scope report forms and correction documents were pulled from the folders and referred for data keying; the folders themselves were refilled at Jeffersonville.

## Interactive Research

**General information**—The Interactive Research Unit was responsible for resolving coverage problems for selected cases. The bulk of its workload consisted of searching the 1982 census mail file for matches for the following kinds of cases:

- *Successors*—A successor was the current operator of a farm listed in the census file under a different operator's name.

Successors' names not matched to the census mail list became "adds" to the census file.

- *Partners*—A partner case was one 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*—A respondent claimed to have returned a completed report form.

Research clerks used interactive computer terminals to search the census data base to try to match these cases to report forms already checked in. The interactive terminals and the specialized search programs employed replaced the manual microfilm search operations used in previous censuses.

**Name and address searching**—Research clerks entered the name, State, and ZIP Code if known, of the case being searched into the computer. The name and address search routine programmed for the research unit used the SOUNDEX principle to search the census data base with only the reported last name of the operator and the ZIP Code of the farm in question.<sup>5</sup> The SOUNDEX system provided for variant spellings or misspellings of a name. The routine carried out searches at three successive geographic levels—first at the five-digit ZIP Code level, then at three-digit ZIP Code level, and finally, if no match had been made, at the State level.

Each possible match was displayed to a clerk for review and confirmation of status. The clerks annotated the report forms with identifying CFN's and 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]), and referred the cases for further processing. Cases identified as out of scope after review of the status codes were referred to batch for check-in. All in-scope successor cases were routed to the agriculture analysts.

**CFN searching**—The census file number (CFN) was the principal numeric identifier for each report form or case received and/or processed. Report forms or correspondence received with an incomplete or wholly or partially obliterated CFN, or lacking one altogether, were submitted to the Interactive Research Unit and were processed using the same search disposition routines as were used for the name and address search.

The processing operation referred 78,329 cases to the Interactive Research Unit. Research clerks resolved 59,159 of these, or over 75.5 percent; the remaining 19,170 cases were referred to analysts for review and determination of status.

## Data Entry

**General information**—The 1982 data entry system, together with the format and simple edit operation (see below), replaced most functions of the 1978 clerical screening operation as well as parts of the technical and "must" case reviews. Data were keyed on a key-to-disk system with an edit program that displayed questions to identify various conditions on the form as data were keyed. The interactive input programs were automatically linked, based on responses to the questions. This helped lead the keyer through the interactive routine. The keyers

<sup>5</sup>For cases with no address information provided by the respondent, the ZIP Code on the outgoing address label was used. If the name and address agreed with the label, but the ZIP Code was missing, the label code was used. If the respondent provided a ZIP Code that differed from the label code, the case was referred for ZIP Code research for confirmation of the correct ZIP Code for the reported address.

corrected and/or flagged many of the problems previously corrected in clerical screening.

The agriculture census data entry staff used 9 data entry systems, each with 16 key stations. Each station consisted of a keyboard and a cathode-ray tube (CRT) viewing screen, enabling the operator to monitor and edit keyed data and receive messages or queries displayed by the input program.

Each keyer's work was subject to quality-control verification, which involved review of a sample of the report forms in each work unit and, when necessary, correction of any keying errors.

After verification and correction, the data were moved from the disks to a magnetic "pooler" tape containing data for only one State (report forms were batched into work units for keying by State), and transmitted to the Suitland headquarters computer facility by telephone data links. As soon as the data were "read" at Suitland, the pooler tapes were erased and reused.

**Data keying operations**—Report forms, linkage documents, and correction documents referred for data keying were batched into work units of 50-100 cases each by type of form (i.e., must, sample, and nonsample) and State. Each work unit was placed in a plastic envelope with a form 82-A405 Batch Cover Sheet attached, and stored in a rolling bin until ready for keying. The batching subunit released materials to the keying/screening operation according to State priorities. In general, work units for a single State were released for keying in groups of approximately 5,000-10,000 report forms each.

Data entry involved keying and general screening of the individual report forms in a single operation. The keyers opened each batch and checked the report forms prior to attempting to enter the data. Report forms were pulled from the work unit and "rejected" from data entry for the following reasons:

- *State code*—State code (first two digits of the CFN) on the label did not match the State code on the cover sheet or within the report form itself.
- *Photocopied report*—An unattached photocopied report preceded or followed an original report, and the CFN's matched.
- *Linkage code*—Linkage code (numeric code linking different establishments of a 2+ or multiunit) for 2+ or multiunit report was missing.
- *Check digit failure*—During keying, the CFN, extra CFN, or geographic area code check digit failed the input edit program three times.

Keyers pulled each rejected report form from the batch, circled in red the linkage code, check digit, and/or CFN for State rejects, and wrote "REJECT" in the top margin of the report forms identified as nonkeyable-report rejects, or "DUPLICATE," for photocopied rejects. Rejects were held out of the batch until the remaining report forms were keyed. The batch then was checked to make certain all components were accounted for; the numbers of report forms keyed and rejected were entered on the form 82-A405 Batch Cover Sheet, and the rejected forms were placed on the top of the batch.

The keying/screening system introduced for the 1982 census was a significant change from previous censuses in that it transferred much of the responsibility for screening individual report forms to the data keyers themselves, eliminating a costly clerical operation. The new system employed two basic kinds of process routines for keying the contents of the report forms: In the first three programs, or "formats," were used successively in keying fixed data items from each report form, the first for

name/address corrections (if any) for each case, the second to confirm county location, and the third to confirm State location if the county location reported by the respondent did not match the four-letter county code on the address label. In each case, the program displayed questions that identified conditions requiring action. In this fashion, the program provided guidance for keying corrections to the names and addresses, and geographic locations. The second used two types of input formats—one for the section identifier and the other for reported data. The processing routine automatically linked the section identifiers and section 1 of each report form,<sup>8</sup> since virtually all forms received included some data in that section. Reported data in sections 2-29 were keyed in the following manner:

- *Data for section identifier.* Keyers selected this format if data and/or remarks were reported within a specific section. The format number for the format to be used was identical to the section number (e.g., format 16 was selected for keying section 16).
- *Reported data.* Once the first format was selected, the format for reported data for the specified section linked automatically with the data file for that section, and the operator could key the data cell keycodes and reported data.

The data entry system used a fixed-length input record for reported data; this consisted of a single keycode/data field for each data cell or item. Only data cells containing data within each section were keyed. After selecting and entering the appropriate format number for a section, the keyer proceeded through the section, entering first the main keycode (three-digit code located in the upper left corner of the data cell and/or below the write-in crop, fruit, livestock, or poultry sections of the report form).

With the keying/screening system, keyers were expected to resolve many of the problems previously identified and corrected in a separate clerical operation. These included:

- *Nonnumeric entries.* Respondent reported an alphabetic equivalent of a numeric value (e.g., "five hundred ten" instead of "510").
- *Data outside of a cell.*
- *Altered stub.* The "stub" is the list of subject headings or titles usually running down one side of a statistical table or report form. The change or addition differed from the preprinted items or descriptions on the report form.
- *Dollars/cents.* Respondent reported specific value instead of dollar value only.
- *Extra CFN.* One or more extra CFN's—identical to the CFN on the address label or otherwise—reported.
- *Double entries.* Multiple entries for a single data cell.
- *Fractions/decimals.* Fractions/decimals reported where not requested, or fractions other than "tenths" reported when "tenths" requested, e.g., "1/2," "1/4."
- *Range entries.* Upper and lower range reported instead of specific number.

---

<sup>8</sup>Section 1 requested data on acreage, including all land owned, leased, or rented, and all land leased or rented to others; the names and addresses of landlords and tenants; whether the respondent had Federal or State grazing permits; and locations of agricultural activities for this place and for agricultural operations in other counties.

- *Reference to other data.* E.g., "same," "ditto," "all," and symbols such as arrows pointing to other data.
- *Summation entries.* A single entry reported for several cells.
- *Incorrect units reported.* Units reported inconsistent with those requested.

The keyers checked any respondent's remarks sections of the form and determined whether the case (1) should be referred to a supervisor (if it required a reply of some sort), (2) contained data, (3) indicated the respondent may not have operated a farm or was a possible landlord, or (4) required a change in the reported data indicated by a remark "flag." The remark flag alerted technical reviewers to reporting errors. Report forms with one or more illegible entries (but not so many as to render the entire form a reject case) were referred to the Agriculture "contact coder"<sup>7</sup> for correction, then keyed.

Batch edit referrals (i.e., lists of item codes for items rejected by the computer edit program), generated from rejected report forms, were referred to Agriculture Division subject-matter analysts after each batch had passed through keying, while cases that seemed to require some sort of written response were routed to the correspondence unit.

**Verification**—As with other major clerical operations, data entry for the census was subject to quality control to ensure completeness and accuracy in keying information from the report forms. Each data keyer's work was reviewed to identify and correct any errors. An "error" was defined as (1) a keystroke error in keying an item code or data item, (2) an omission of a field or document, (3) a procedural error (i.e., the keyer failed to follow keying procedures for an item or document), or (4) duplication of a field or document. All errors were to be corrected before the data involved were released for output to computer tape.

Data keyers progressed through three major phases of verification of their work: training, proficiency qualification, and production. These three phases were divided into six stages of verification—stage 0 for training, stages 1-3 for proficiency qualification, and stages 4-5 for production work. During training, the work of keyers in stages 0 and 1 was verified on a 100-percent basis, and stage 2 on a sample basis.<sup>8</sup> Stages 0-2 lasted 5 working days. To advance from one stage to the next, a keyer could have a field error rate (i.e., the rate of data fields identified as defective because of data entry problems) no larger than 3.5 percent for stages 0-1, and 2.5 percent for stage 2. Keyers failing to qualify were retrained, while qualified keyers moved on to stage 3. Keyers remained in stage 3 until they attained an error rate of no more than 1.5 percent, at which time they were moved to stage 4.

During stage 4, keyers' work was verified on a sample basis, and maintenance of an error rate of 1.5 percent or less was required. A keyer receiving 5 or more consecutive "accept" decisions in a sequence of 10 or fewer keyer decisions, qualified for advancement to stage 5. A "keyer decision" was made on the basis of the number of keyer errors relative to the total number of data fields keyed, and the number of decisions determined whether a work unit was accepted or rejected.

<sup>7</sup>An Agriculture Division subject-matter specialist.

<sup>8</sup>The sample selected for verification depended on the size of each batch. Batches with up to 9 documents were verified 100 percent, while larger batches were randomly sampled at rates ranging from 1 in 5 (for batches of 10-19 cases) to 1 in 40 (for batches of 80 or more cases).

A keyer in stage 5 had an acceptable error rate of only 1.0 percent, and in sample verification had to receive 8 "accept" decisions in a sequence of 10 keyer decisions. Failure to maintain these standards resulted in reversion to stage 4.

Every error detected was to be corrected and reverified before data were accepted for transmittal to Suitland for computer processing.

After data keying and verification, the keyed report forms were kept in a holding area until computer processing of the data was satisfactorily completed. Thereafter, the processed forms went to central files for sorting, boxing, and storage.

## COMPUTER PROCESSING

### General Information

While the 1982 Census of Agriculture required a considerable amount of manual sorting, reviewing, and checking (described above), it also made extensive use of programmable mechanical sorting equipment and interactive computer terminals for data entry. Much of the data assembly, editing (the mechanical process that validated, cross-checked, and refined a data file), and tabulation was carried out on the mainframe computers at Bureau headquarters in Suitland. Computer processing began as soon as the first group of report forms arriving at Jeffersonville were keyed and the data cleared for transmission to Suitland, and continued until the final tabulations were completed in July 1984. Approximately 3 million individual census records were edited; of these, some 2.24 million represented agricultural operations meeting the Bureau's definition of a farm and were incorporated into the agriculture census file.

The computer processing cycle was divided into three major phases—formatting and simple editing, complex edit and edit correction, and data tabulation. These operations are described below.

### Format and Simple Edit

The format and simple edit operation (1) converted the individual data records into binary records that could be manipulated using the processing program, (2) carried the problem flags set during keying operation to the computerized record, and (3) established "flags" for problems identified during the formatting process.

The individual data records created for each census report form during data entry had to be converted from the "fixed" record layout used for the input data into a "variable" output record that had binary coding for numeric values for use in all subsequent computerized editing and/or tabulation. The formatted record consisted of a fixed-length section containing all of the record's identifying information—such as State and county codes, CFN, SIC code, and so on—and a variable-length section containing keycodes, original data, current data (i.e., data inserted by any editing), flags, and the like. The variable-length section included a computer "word," or record segment, for each item reported, imputed, or changed, but nothing for sections or items left blank in the original record. The computer processing programs identified the data items from keycodes and ignored blank segments of any record.

The format and simple-edit program also ensured that flags set during data entry were carried to the computer records and established flags for problems identified during formatting, including the following:

- *Form reject flags.* Cases with illegal State, county, or report form codes (i.e., codes outside acceptable ranges); non-numeric State or county codes, or CFN's; invalid check-digit; or 10 or more individual data-item rejects.
- *Simple-edit out-of-scope flags.* Cases with no reported sales or livestock inventory.
- *Item rejects.* Cases with individual items flagged, such as illegal keycodes, invalid data, invalid crop coded for a particular State (e.g., pineapples for Maine), keycode out of section or sequence, invalid keycode for report-form type, or invalid enumeration date.
- *Simple-edit referrals.* Cases with altered stubs; double, illegible, or summarized entries; "other" livestock/crop coding; or invalid units for a particular item. (Most of these problems would have been flagged during data entry, but those not previously identified were flagged by the format/simple edit program.)

### Complex Edit

**General information**—While the simple edit and format procedure converted the data records into binary codes and set certain problem flags, the complex edit program (1) checked key items in the data files; (2) matched selected ratios and reported data against set limits based on experience in previous censuses; and (3) corrected errors by rounding individual data items, substituting sums of detailed items for a reported total, or imputed on the basis of one of a variety of ratios that included the questionable component.

The computer programs used to perform these tasks comprised several thousand individual operations in total, although usually only a fraction of this number was required in editing the data for any particular report form. Agriculture subject-matter specialists transmitted the edit program requirements—or computer edit specifications—to the computer programmers by means of decision logic tables (DLT's), i.e., tabular displays of the elements of various edit operations from the conception to the solution. The actual editing was done by State, in batches consisting of formatted records sorted by State, county, and CFN, assembled by setting specific cutoff dates.

**Computer edit**—Using these edit programs, the computer checked each data record and—

- Supplied missing entries. (Specific items were imputed based on reported information for operations of a similar type in the same geographic area.)
- Reconciled acres reported for individual items with the total acreage reported.
- Imputed production for crops when the reported production was outside acceptable limits.
- Edited to assure consistency between and within various sections of each record.
- Calculated and checked values for products sold, using average prices in each State for each production item, and substituted these values for reported values if the latter were outside acceptable limits.
- Determined whether each record represented agricultural operations meeting the Bureau's definition of a farm.

- Classified (and coded) individual records by acreage, tenure of operator, value of agricultural products sold, type of organization, and SIC code (type of farm).

Records for operations determined to be out of scope of the census were deleted from the data file and transferred to the out-of-scope file. Printouts were made listing cases identified as out of scope during computer edit, and the clerical staff at the Jeffersonville facility reviewed the report forms for these cases to ensure accurate keying and correct classification.

The edit programs also enabled the computer to identify and retain in the data file records for agricultural operations that lacked in the census year, but normally would be expected to have, total annual agricultural products sales value of \$1,000 or more. The records for these places were tested against criteria developed to identify agricultural operations that normally would qualify as farms under the Bureau's definition. Thirteen criteria codes were used for these "farms" to identify the general type of product (cash grains, livestock, and so forth) where minimum acreage or inventory had been specified.

In addition to identifying records for places that "normally" would qualify as farms, the computer-edit program also identified nonsample records representing farms meeting the certainty criteria established for each State, and converted these nonsample certainty records to sample records. (Certainty criteria varied between States from minimum sales of \$100,000 to \$500,000, or minimum acreage of 1,000 to 10,000 acres.) Institutional and other special cases also were included in the certainty classification even if they failed to meet the minimum sales or acreage criteria, as were all records for addresses in counties that had fewer than 100 farms in 1978. The Bureau obtained the additional detailed data needed for conversion by correspondence with the addresses involved or by imputation based on responses from farms of similar size in the same geographical area. Any conversions based on reported sales or acreage were coded as certainty cases.

**Failed-edit correction**—The last step in the computer edit was the referral split, in which the computer identified cases that failed the complex edit process (i.e., cases that required correction and re-editing because of one or more failed items), and generated batch edit listings (BEL's) that included a printout for each individual data report with "nonsuppressible" flags.<sup>9</sup> The listings showed the item or items for each report that (1) had failed the edit, (2) had been changed by the edit, or (3) had a referral flag. No listing page contained items for more than one farm record, although printouts for single cases sometimes ran to more than a single page.

The batch edit listings were referred to Jeffersonville, where they were matched to the report form file, and the listing sheets and corresponding report forms reviewed clerically. The clerks checked each case and determined the general action to be taken, assigning the appropriate disposition codes, as:

Code	Action
1	Make corrections, re-edit record.
2	Make corrections, bypass specified sections of edit.
3	Make corrections, bypass edit except coding, SIC coding, and summing.
4	No corrections; change failed edit to passed edit.
5	Delete record from file.
6	Make corrections, change referral disposition code to 3.

<sup>9</sup>Flags denoting geographic code errors, "other" crops entries, remarks on the report form, altered stubs, and wrong units reported.

Reports requiring numerous corrections were recycled through the entire processing operations—keying, formatting, and editing. For most cases, corrections were marked on the batch-edit listing sheets and the *corrections* were keyed to tape, verified 100 percent, and transmitted to Suitland, where they were matched to the data file. The corrected files then were reedited to ensure the required changes had been made and to determine whether more were needed.

**Data merge**—After computer edit and failed-edit corrections were completed, the corrected data files for each State were merged into a single file in sequential order by State, county, and identification number. The merge program tallied farms by size, value of products sold, and type (used to help add data for nonrespondent cases), and identified and displayed problem cases for review and correction before tabulation. The resulting computerized data file then was unduplicated using a census file number (CFN) matching program to identify and print out duplicate records for review. Usually, the first of the duplicate records displayed was retained, while the rest were deleted from the file.

## Statistical Adjustments

**General information**—Approximately 14 percent of the addressees on the census mail list, generally smaller farms, never responded to the census, while certain data were collected only from about one out of every five farms. The Bureau of the Census employed two kinds of statistical adjustment to compensate for (1) nonresponse and (2) the use of sampling to collect certain data: Imputation for nonresponse was carried out using a statistical procedure and data collected in a sample survey of addresses still nonrespondent to the census in April 1983, while stratification and sample weighting was a statistical procedure used to provide data estimates by extrapolating the characteristics of sample farms to estimate totals for selected data for all farms.

**Imputation for nonresponse**—The Bureau imputed census data for nonrespondents after the data files were corrected, merged, and unduplicated. Nonrespondent addresses were classified as “large” or “small” farms based on expected annual sales of agricultural products (a “large” farm had to have expected sales of \$100,000 or more). The Bureau carried out a telephone followup of “large” nonrespondent cases and, in April 1983, selected a sample of approximately 13,500 small nonrespondents for a special mail followup to collect the data necessary to develop estimates of the number of nonresponse cases that were farms. (See ch. 5 for details of general telephone operations and for information on the nonresponse survey.)

The Bureau used the State-level estimates of the percentage of nonrespondents to estimate the number of nonrespondents for each county. Respondent farms from similar groups in each county then were randomly sampled to represent the nonrespondent farms. This meant simply that selected farms were counted twice in the tabulation, while all other farms were counted only once. This nonresponse adjustment procedure assumed that the respondent and nonrespondent farms had similar statistical characteristics. The procedures assumed that, for example, a respondent dairy farm and a nonrespondent dairy farm, in the same county and with similar value of sales (one reported, the other estimated), had approximately similar characteristics in terms of acreage, size of herd, and so on.

**Stratification and sample weighting**—The 1982 Census of Agriculture, like the 1978 enumeration, included data collected from a sample of all farms. Report forms containing sample items were mailed to all addresses in Alaska and Hawaii, and to a sample of all other addresses on the census mail list. The sample selected included all “certainty” cases, i.e., abnormal farms, farms with a complex organizational structure (e.g., multiunit operations), and all addresses identified as “large” farms. The definition of a large operation for sampling purposes varied by State, from a low of 1,000 acres or \$100,000 in sales in New England to as high as 10,000 acres or \$500,000 in sales in some Western States. Large farms identified during processing were sent a supplemental report form containing only the sample items. In addition to the abnormal, complex, and large operations, all farms in counties with fewer than 100 farms in 1978, and all farms in SIC code groups with very few farms at the State level in 1978, were included as well. The resulting file constituted an approximate 20-percent sample of all farms.

Sampling introduced into the census data several factors that could cause substantial variation and/or bias. In the initial, pre-enumeration selection, fully half the addresses from which the sample data were to be collected did not represent farms; thus tabulated data came only from part of this sample—those addresses representing farms. Further, the sample was stratified using sources of variable quality, and the response rate for addresses in the sample may have been different than for nonsample cases.

To improve the accuracy of the estimates drawn from the sample, the Bureau carried out a “poststratification” of the noncertainty sample operations by creating 128 new strata, formed by 8 sales groups, in turn divided by 8 SIC code groups, then by 2 acreage classifications. These were:

Sales	SIC code groups	Acres
\$1 to \$999	011 Cash grains	0 to 69
\$1,000 to \$2,499	013 Field crops, except cash grains	70 or more
\$2,500 to \$4,999	016 and 018 Vegetables and melons, and horticultural specialties	
\$5,000 to \$9,999	017 and 019 Fruits and tree nuts, and general farms, primarily crops	
\$10,000 to \$19,999	024 and 021 (except industry 0212) Dairy farms; and livestock, except dairy, poultry, and animal specialties (except beef cattle [except feed lots])	
\$20,000 to \$39,999	0212 and 027 Beef cattle, except feed lots, and animal specialties	
\$40,000 to \$99,999	025 Poultry and eggs	
\$100,000 or more	029 General farms, primarily livestock	

The Bureau multiplied the data from sample farms by the ratio of total farms to sample farms in each stratum to develop the estimates for the sample items. The weight of a certainty farm remained “1.”

## Tabulation and Data Review

**General information**—After edit, correction, and merge, the data records were ready for tabulation. The individual records were tabulated by computer into a series of detailed data matrices, each comprising several thousand different items that would provide the basic inputs for most of the data tables drawn from the census file. The Bureau also employed these data matrices to



prepare matrix tabulations, after analytical corrections, in a detailed county-level format, for use in reviewing and identifying problems in the preliminary data.

After the tabulations were reviewed and problem records corrected, the Bureau made a special tabulation of the corrected records as they were before and after correction. The uncorrected data then were subtracted from the corrected data, and the net totals merged into the data matrices. The corrected matrices formed the detail data file and served as the data source for the preliminary reports and the review tabulation for final Volume 1, *Geographic Area Series*, reports. After completion and correction of the review of the volume 1 tables, the final volume 1 tabulations were run from the master matrices. State-level cross-tabulations were run in a separate computer pass after the final correction of the volume 1 tabulations.

**Analytical tabulations**—County-level analytical tabulations were used to review the aggregated data. All of the items reported on the individual report forms were tabulated for each county and State, for all farms (that is, for all operations meeting the Bureau's definition of a farm), and for farms with sales of \$10,000 or more. Historical data from the 1978 census final reports<sup>19</sup> were included for use in reviewing the 1982 data for completeness and accuracy. Agriculture Division analysts used the analytical tables as their basic review documents, but also employed substantial amounts of related data, drawn mostly from USDA estimates. The analysts wrote detailed descriptions and criticisms of data problems on form 82-A302 Table Review Criticism Sheets. These provided for subsequent use records of what was questioned, what action was suggested, and the net change for each item questioned.

Representatives of USDA's Statistical Reporting Service's (SRS) State offices also reviewed the analytical tables and the criticism sheets. The SRS reviewers identified any additional problems they found in the tabulations and offered additional or alternative solutions to problems noted previously.

The detailed criticisms of the data, together with suggested corrective action, were transmitted to Jeffersonville.

**Data corrections**—Corrections to the data record and all deletions from the data file were made by the Jeffersonville staff using the form 82-A210 Individual Form Correction Transcription Record. An A210, completed for each individual record requiring correction, was filled out with the identification data (State and county codes, CFN, and so forth), and with the keycode for each item to be corrected or changed, together with the correct datum.

After completing the A210's, the staff—

- Reviewed the critique and corrective action for problem data suggested by the reviewers.
- Validated the data in the submitted criticisms, or made the necessary corrections.
- Obtained reports from farm operators for places that had not been included in the tabulation.

<sup>19</sup>The 1978 State-level data incorporated estimates based on the results of the 1978 Area Sample Survey. Since no comparable survey was part of the 1982 census program, the 1978 data used for comparison purposes were produced by subtracting the area-sample data from the 1978 data, leaving only data reported by farms on the mail list.

- Corrected data-keying, reporting, and processing errors.
- Identified and deleted from the data file duplicate records not previously detected.
- Assigned correct State and county codes for large operations to ensure these operations were tabulated in the proper geographic location.

All the corrections made were reviewed by the Agriculture Division staff for accuracy and to ensure that the data criticism were satisfied. The preliminary reports then were tabulated and reviewed. Any additional corrections needed were made by computer, or hand corrections were made to the tabulation printout. The data file was corrected as often as necessary to ensure its accuracy.

**Tabulations for counties, States, divisions, regions, and the United States**—County and State tables were drawn from the data matrices, and State cross-tabulations directly from the data file. Data for census geographic divisions and regions, and for the United States, were prepared by summing data from the States. The historical data for the 1982 State tables were taken from the 1978 tabulations less estimates from the 1978 area sample survey.

**Final disclosure analysis**—The Bureau of the Census, prohibited by law from publishing data that could be used to identify any individual respondent to any of its censuses or surveys, employed a procedure called "disclosure analysis" to maintain the confidentiality of the data. This involved a review of all data tables before they were released for publication. This review identified and suppressed data items whose publication (1) would result in direct disclosure of the data for a particular respondent, or (2) could be used to reveal information about an individual by derivation—that is, adding or subtracting a published subtotal from a published total would reveal individual data. Publication of the number of farms reporting an item was not in itself considered a disclosure; only related information about an item was suppressed.

As a rule, the Bureau did not publish any data for counties with 10 or fewer farms in 1982. The disclosure procedures established lower limits for the number of farms reporting a particular item before the data could be published. When the minimum acceptable number of farms reported an item, the Bureau published the the data unless comparison of different tables could result in disclosing that one or two farms accounted for over half of the total published. There were exceptions to these general practices, usually when very large specialty operations were involved. For example, there were numerous cases of data involving poultry producers, feedlots, greenhouses, and so forth, in any of which the publication of the data might result in the identification of a specific operation, but whose absence from the tabulations would grossly distort the totals.

Various tables included identical information arranged under several different classifications, so that the identification and suppression of a figure in one table required review of all related tables and the suppression of the datum in them as well.

The disclosure analysis of the 1982 census tabulations was carried out partly by computer and partly by Agriculture Division statisticians, as the computer programs were incapable of completing the entire analysis.