CHAPTER III. ELECTRONIC DATA PROCESSING

FOSDIC

Equipment

One of the Census Bureau's objectives is to make census data available to the public as rapidly as possible. To accomplish this, research and planning, aimed at replacing manual card punching with high-speed equipment, was initiated early in 1951.

The Bureau's search for a mechanical solution to this problem was made in conjunction with the National Bureau of Standards. Late in 1953, engineers of the National Bureau of Standards delivered the first model of FOSDIC to the Bureau for testing.

FOSDIC can be described as a machine which is capable of "reading" information from a negative microfilm copy of an appropriately designed schedule and transferring it to magnetic tape for processing on electronic computers. This reading is done by a moving beam of light. The information read is transmitted onto the magnetic tape in the form of impulses. The tape containing the transmitted information is later fed into a computer. Basically, this is how FOSDIC operates: The beam of light scans the microfilm negative looking for an index mark (a black square printed on the FOSDIC schedule). Following the index mark is a series of circles printed on the schedule (see appendix I). The beam of light then scans each circle to determine the one filled in. A code indicating which circle was filled is then recorded by FOSDIC in the form of an impulse on magnetic tape.

FOSDIC actually consists of four separate units:

1. The tape unit holds the magnetic tape, drives it at a constant rate, records impulses on tape, and, at the end of a run, rewinds the tape for computer use.

2. The console contains all the operating switches, recording dials, cathode ray tube, light beam, film drive mechanism, and an oscilloscope for testing the strength of electrical impulses.

3. The scanning unit measures the distances the beam must travel, both horizontally and vertically, and "decides" where a mark has been made.

4. The program unit "tells" the machine what to do. This includes instructions on calibration of the document, on the order in which the questions will be scanned, and on the distances to travel to find the questions and the marking areas.

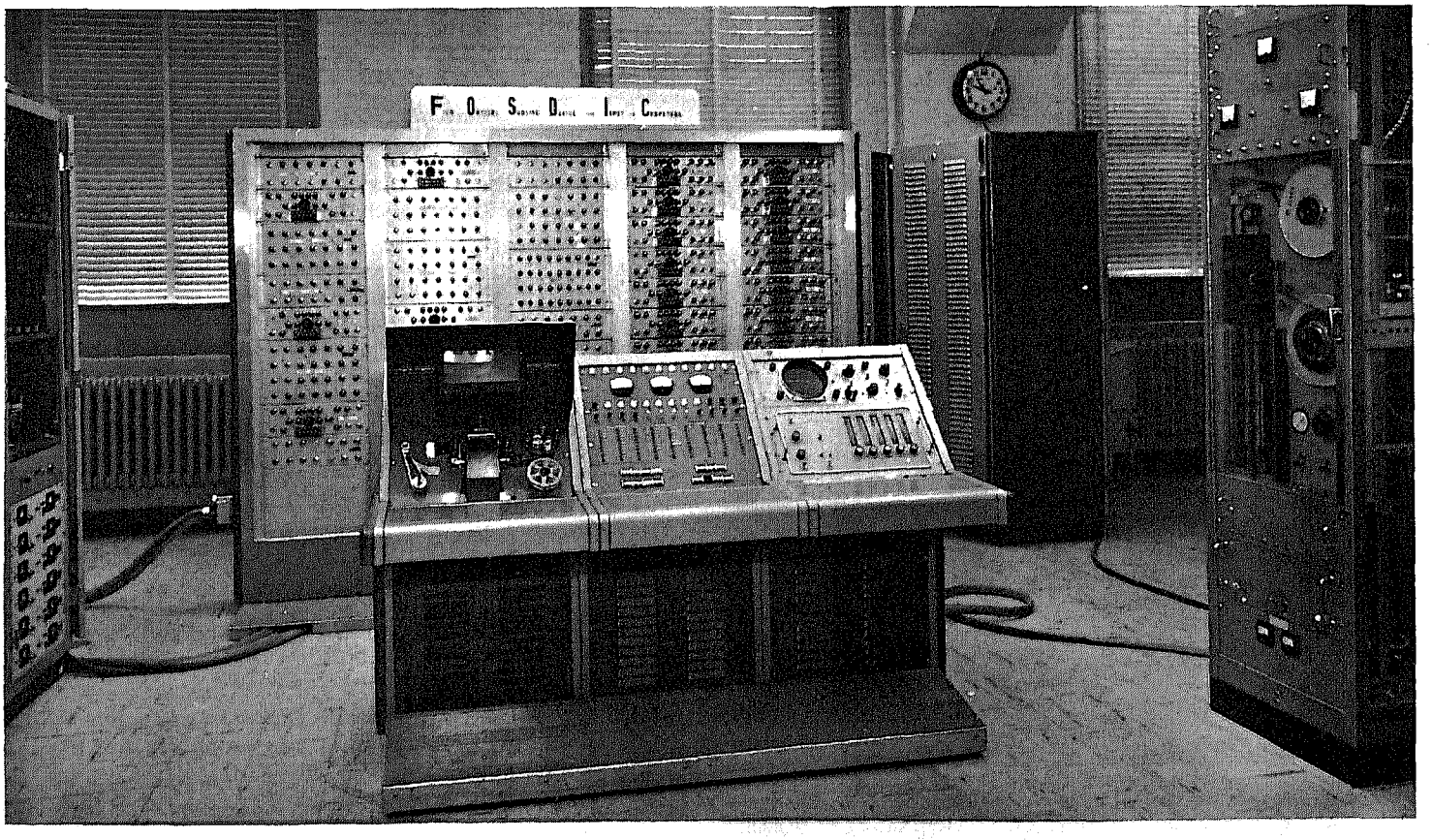


Figure 7. FOSDIC (Film Optical Sensing Device for Input to Computers). This equipment transfers data from the reels of microfilmed schedules to magnetic tape which is fed into the electronic computer. In the foreground is the operator's console; behind it is the scan unit; to the left is the program control unit; to the right are the power supply unit (under the clock) and the magnetic tape servo.

Much was learned about the abilities and weaknesses of the first model of FOSDIC when it was placed in operation on special censuses, a portion of the Census of Governments, and the National Housing Inventory. This knowledge led to the development of a new version, FOSDIC Model III. (Model II had been designed and built for the Weather Bureau in the interim.) The design of the new model was a cooperative project by engineers and technicians of the Census Bureau and of the National Bureau of Standards.

Test results on FOSDIC III were most encouraging and it was decided, in the spring of 1957, to build four production models for use during the 1960 censuses. Actual construction was begun about July of 1957 in the Census Bureau.

Liaison with the design engineers of the National Bureau of Standards continued throughout the entire construction program, because the design engineers were still improving certain functions. A tremendous number of changes had to be made during the assembly project to include the improvements.

Machine work for the mechanical assembly of FOSDIC included making the film drive mechanisms, the complete scan unit housing, the chassis for the scan unit, the printed circuit boards and their installation hardware, and indicating the light panels, and rebuilding the tape transport.

Electronic assembly started with the wiring of subassemblies and gradually worked up to major chassis assemblies and printed circuit boards and then to complex interwiring within a unit and finally between units.

Each machine contains over 2,000 transistors, 2,500 crystal diodes, 1,000 electron tubes including special beam-switching tubes, 1,000 indicator lamps, 8 miles of wire, 81,000 connections, 50 large electron tube chassis, 65 large printed circuit boards, 15 individual direct-current power supplies, and a monitoring oscilloscope and electronic counter, plus about 20 commercial electronic packages for each machine.

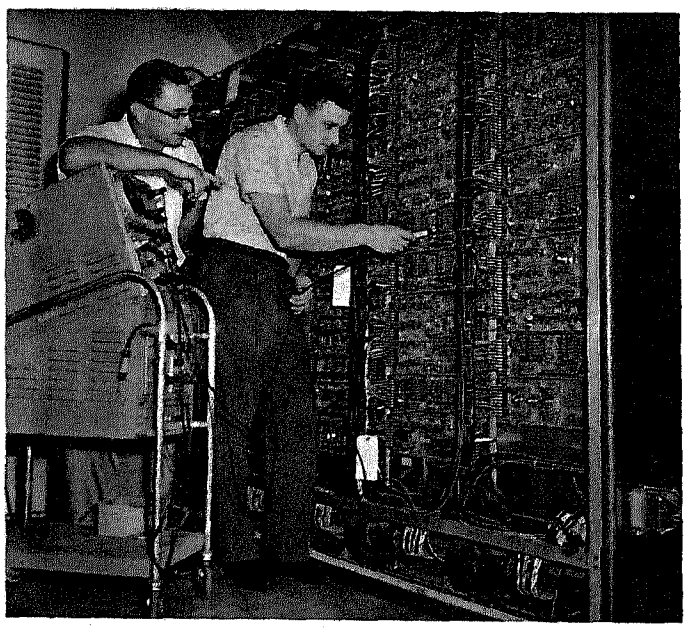


Figure 8. Making an Adjustment on the Back of a Fosdic Scan Unit. The set of machines composing the system known as FOSDIC was designed cooperatively by the Bureau of Standards and the Bureau of the Census. The four systems used in the 1960 processing were produced at the Bureau of the Census.

With the completion of construction, testing, and correction, the four FOSDIC's were ready in time for the data processing of the 1960 censuses. The decision to use FOSDIC in the data processing had various ramifications on other areas of the 1960 censuses; for example:

1. The enumeration schedule--which was to be microfilmed and then read on FOSDIC--had to be designed to meet the special requirements of FOSDIC. Some of these special requirements were (a) The maximum size of the reading area could not exceed 14" vertically and 21" horizontally and the minimum document size was 12" vertically by 4" horizontally, with no minimum restriction on the reading area; (b) FOSDIC would perform more reliably if only black ink were used in printing the schedules; (c) the index marks (black squares on the schedule which served as the "takeoff" points from which the FOSDIC beam moved) had to have at least a one-inch (center to center) minimum separation, vertically and horizontally. Other elements that had to be considered included spacing of circles, the size of the circles, and the thickness of the circle lines, each requiring precision to about 1/1000th of an inch.

2. The effectiveness of FOSDIC depended, in part, on the quality of printing of the FOSDIC schedules. The Government Printing Office printed all of the FOSDIC schedules in its own plant. Offset printing with copper plates was used because copper plates are noted for long wear. A quality control system was used during the printing operation. Two sample sheets were selected each time 5,000 sheets were printed by a press. The sample sheets were microfilmed and run through FOSDIC to measure the uniformity of answer circles and to decide whether size and intensity of index marks were acceptable. This procedure was in addition to the standard inspection and control procedures regularly used by the Government Printing Office.

3. While the FOSDIC schedule did not require use of any particular writing instrument, it was regarded as worthwhile to attempt to get reasonably uniform marks with a readily available writing instrument. Consequently, a series of tests of writing instruments was made. Since there would be more than 150,000 enumerators, a writing tool was needed which was low in cost, easily replaced, convenient to use, and consistent as to quality of work. Ballpoint pens and automatic and regular pencils of different degrees of hardness were tested. A first-quality, wood-encased No. 2-1/2 pencil with a suitable eraser was selected. Each enumerator received several pencils, and the same type of pencil was used when circles on the FOSDIC schedules were filled during coding operations. Use of these pencils was not required but was requested to the extent possible.

Operations

FOSDIC processing was an integral component of the electronic processing system. Operation and maintenance of the FOSDIC system included the following broad classes of operations:

1. Recruiting and training operators for the FOSDIC equipment
2. Maintenance of the FOSDIC equipment
3. Control of the microfilm after shipment of the exposed film from Jeffersonville, through development, receipt in Washington, issuance of micro-

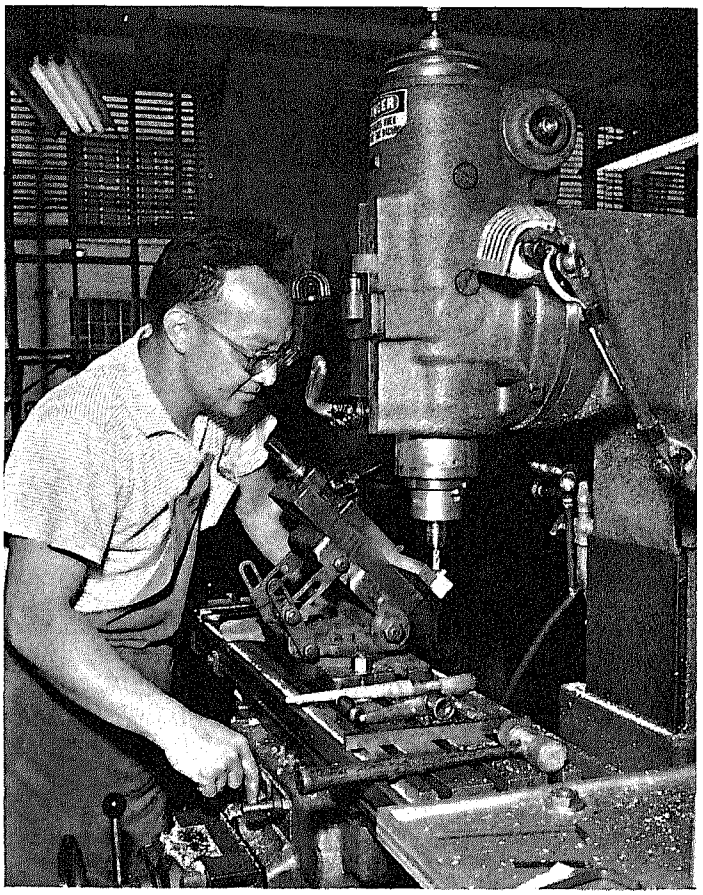


Figure 9. Making Parts for the Maintenance of FOSDIC.

film to FOSDIC, and retention of microfilm after FOSDIC processing

4. Transfer of the data from microfilm to magnetic tape, and sending the tapes, labeled as to content, to be put through the computer

5. Developing and testing FOSDIC programs for conversion of data on microfilm to FOSDIC output tapes

FOSDIC can read and record from microfilm very rapidly. One enumeration schedule for 100-percent data contained about 4,500 small circles arranged in groups, or "fields," between index marks (see example, appendix I). Under the directions of the program unit, the scanning section of FOSDIC found the first index mark, scanned each of the circles in a field to determine whether or not it had been marked by the enumerator, recorded the one that had been (or, if two had been marked, determined the denser marking and recorded that one only--a feature of FOSDIC which permitted erasures and corrections to be made on the FOSDIC schedules) then went on to the next index mark and repeated the operation. It examined each of the 4,500 circles in a frame of microfilm in less than two-thirds of a second. During the processing of the 100-percent data, FOSDIC scanned the microfilm at the rate of about 100 frames per minute (each frame was a microfilm of one 20-line 100-percent schedule) and recorded the data as magnetic impulses on computer tape.

A total of 9,100 hours of time on the FOSDIC equipment were used in a 17-week period for processing of the 100-percent data. Of these hours, 3,700 were for production runs, and the remaining 5,400 hours were for scheduled and emergency maintenance of the machines, testing, and other nonproductive purposes. This was considerably less time than had been anticipated; it had been estimated that as many as 10,600 hours of FOSDIC time might be required, with as many as 5,300 hours of production. Only the time spent on maintenance, testing, and other nonproductive uses equaled, and slightly exceeded, the estimated time.

For the 1950 censuses, the card punching required about 200,000 man-days, with a peak of about 3,000 operators. The microfilm-FOSDIC complex, with a peak of about 100 camera and machine operators, accomplished virtually the same job for the 1960 censuses for both the sample and the 100-percent data, for an 18.5 percent larger population, in about 28,000 man-days.

FOSDIC Film Control

Developed microfilm was delivered daily to Bureau headquarters in Washington with a Microfilm Transmittal Form (see form 14, appendix J) which listed the reels of film in the shipment. The film received was recorded by States in a Microfilm Work Unit Control Register (see form 15, appendix J) which reflected the current status and history of each microfilm reel after its receipt in Washington.

Reels of microfilm were run through FOSDIC in accordance with established priority lists. A Film Issue Request Transmittal (see form 16, appendix J) was used to control the movement of microfilm to and from FOSDIC processing. Microfilm processed through FOSDIC was retained in the files; microfilm rejected by FOSDIC (see form "Reject Report--Microfilm, FOSDIC, and Computer," form 17, appendix J) was held for daily check of all rejected rolls for analysis of problems requiring remicrofilming.

UNIVAC 1105 COMPUTERS

Any general-purpose computer comprises the following parts: Control, input, memory, arithmetic, and output.

The control part of the computer is the electronic circuitry that interprets the manmade program of instructions and that directs the other parts of the computer to perform the desired functions.

The input part consists of devices which feed information into the computer.

The memory of a computer is the storage section which is used to store the instructions and data until they are needed. Storage is necessary because the computer can perform only one step at a time; e.g., select a number from storage, then select another number from storage, then compare the two numbers, then store the larger of the two numbers again, etc. Magnetic cores or magnetic drums are among modern computer memory elements. A core is a piece of metal which can be influenced by an electric current to maintain one of two distinct states of magnetization. A drum has a special surface which can be influenced similarly in small local areas. One memory stage represents one "bit" of information. Many of these bits in association form a computer "word." The size of the computer memory is the number of words it can store at one time.

Arithmetic is the part of the electronic circuitry that, under program control, adds, subtracts, divides, or multiplies. It can also perform logical operations such as comparing, extracting, etc. The computer can solve problems of higher mathematics, but the process is accomplished by a series of minute steps such as one addition repeated 5 times in order to multiply by 5.

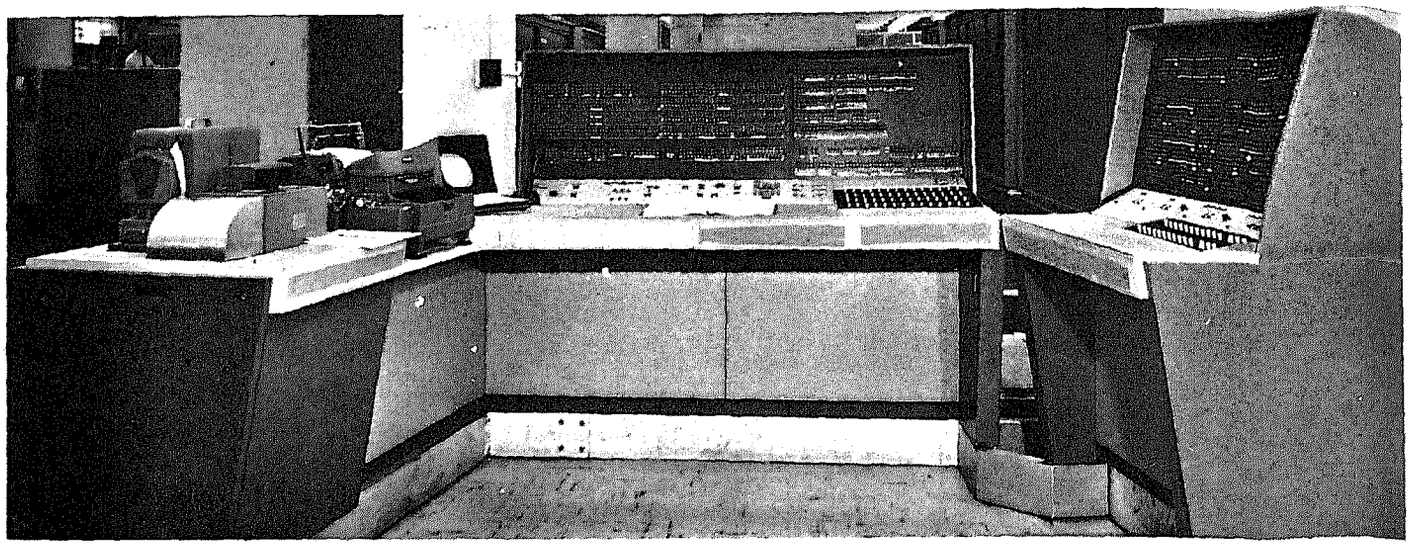


Figure 10. Univac 1105 Electronic Computer and Related Equipment. In the center is the operator's console; in the left foreground is a Ferranti paper tape reader (punched paper tape is a secondary medium of communication with the computer, used for special purposes). Between the paper tape reader and the main console is a flexowriter. To the right is an auxiliary console associated with the magnetic tape system.[11]

Output is the part of the computer that translates the computations into some form of presentation. This may be visual presentation which can be understood by the human eye and brain, such as statistical tables, or magnetic tape which can be used again by the computer or by other equipment, or some other type of presentation which is intelligible only to another machine.

For the 1960 censuses, the Bureau used Univac 1105 computers. The Univac 1105 is a high-speed, large-scale electronic computer. It is a binary machine capable of both scientific and data-processing applications. It can perform 41 different arithmetic and logical operations, and can execute an average of 20,000 instructions per second. Of course, some portions of the program may be repeated thousands of times in tabulating mass statistical data.

Each 1105 at the Bureau has the following major components:

1. 18 Uniservo tape handlers with reading and writing rates of 20,000 characters per second. (The Uniservos are both input and output.) For each nine Uniservo tape handlers there is an input-output magnetic core buffer
2. A paper-tape reader (input) with a speed of 200 characters per second
3. A paper-tape punch (output) with a speed of 60 characters per second
4. Two banks of magnetic core memory for a total of 8,192 words of high-speed memory
5. A 16,384-word magnetic drum memory

Magnetic tape is used for computer input and output. (The preparation of data tapes for input to the computer for the 1960 population and housing censuses was done by FOSDIC, and the translation of output tapes into printing was done by the high-speed printer.) Information is recorded on the tape by means of binary code, a combination of only two different conditions, i.e., zero and one, or magnetic spot and no magnetic spot, or on and off, etc. For example, the Univac code for A is 010100, or, no-yes-no-yes-no-no, representing the two states of magnetization in six locations on the tape, and the code for 9 is 001100. A character is recorded by placing magnetic spots across the tape in the appropriate places. As many as 200 characters can be recorded on 1 inch of tape, and the 1105 computer can read 100 inches of tape per second. Thus data can be transferred from magnetic tape to the computer memory at the rate of 20,000 characters per second.

For the 1960 censuses, the Bureau used two Univac 1105 computers in its Washington headquarters supplemented by two additional 1105 computers at the University of North Carolina, Chapel Hill, and at the Armour Research Foundation of the Illinois Institute of Technology, Chicago. The Bureau shared in the cost of the latter two installations and received a proportionate share of the available computer time. In addition, for a short time the 1105 computer at Griffiss Air Force Base at Rome, N.Y., was used on a contract basis.

HIGH-SPEED PRINTERS

The two electronic high-speed printers at the Bureau are 600-line-a-minute machines manufactured as auxiliary units for the Univac computers. Identical printers at each of the two university installations were also used for the 1960 census work.

Input to the printer consists of magnetic tape, primarily tape which was the output of computer processing. The pattern on the magnetic tape is sensed by a tape unit similar to those serving the computer. As information is read from the tape, it is transferred to a small magnetic-core buffer memory. In the printer itself there is a revolving drum with 130 bands of fixed type on its surface, each band containing 51 different letters, numbers, and special characters. Univac codes for enough characters to print one line are transferred from the buffer memory to a vacuum-tube printer memory. The contents of the printer memory are interpreted, then as each band of type reaches the position in which the proper character for the line is in place behind the ribbon and paper, a hammer is activated which presses the paper and ribbon against the character on the wheel. When all characters for a line have been printed, the paper advances one line and the process is repeated—all at the rate of 10 lines per second.

PREPARING COMPUTER PROGRAMS

A computer "program" is a set of instructions to the computer for processing the data. "Programing" is the preparation of a set of such instructions. In the course of processing the 1960 census data more than a thousand programs, varying in length from 1,000 to 8,000 instructions per program, were required.

Programing has two phases. The first is planning, the second is coding. Coding is the process of translating detailed specifications for an operation into a set of codes which can be interpreted by the computer as instructions. This phase requires an intimate knowledge of the computer system. Exact layouts of the formats of the input data file record, the output data file record, and the working and storage memory areas of the computer must be set up. Then the operation specifications must be broken down to simple step-by-step machine operations such as "add field x to field y" and written as an instruction in a code recognized by the computer.

The planning phase of programing is the process of shaping the requirements and desires for certain published results from given data to the capacities and limitations of the computer system and to the limitations of resources and time made available for the operation. The desired results were usually expressed in the form of statistical table outlines, editing requirements, descriptive text, and other specifications furnished to the programers by subject matter specialists. The planning phase requires considerable knowledge of the subject matter by at least the supervisory programing staff in order for them to understand not only the specifications but also the intent of the specifications. This phase requires a considerable translation and reworking of the specifications into an order and mode of presentation which will conform to the computer order and mode of operation and, thus, facilitate a precise and complete translation to computer instructions.

Once a program, or set of computer instructions, for an operation has been prepared, a period of testing is required. Output from a test or from actual data input is examined closely in order to ascertain whether the output conforms to expectations, and, if not, the program is altered until the output does conform. Many programs are so complex that a rather lengthy period of testing and adjustment is required; the process is referred to as proving in a program. Once the program is proven in, it is considered ready for use in production. However, even during production, unanticipated characteristics of the input data may require adjustments and rerunning of the program to insure satisfactory output.

WEIGHTING SAMPLE DATA

The computer carried out a ratio estimation process for inflating the sample, by dividing the population and housing units in an area into groups by various characteristics for which both 100-percent and sample data were available, then obtaining, for each group, the ratio of the 100-percent number to the sample number. Then each person (or housing unit) was assigned an integral sample weight approximately equal to the ratio for his group.

More specifically, the population of each ED was divided into 44 groups, each of which was referred to as a ratio estimate cell. Heads of households were classified by age, sex, color, and as owner or renter of the housing unit occupied, and other members of the household were classified by age, sex, and color, as shown below.

RATIO ESTIMATE CELLS FOR POPULATION

Item	Male		Female	
	White	Nonwhite	White	Nonwhite
Under 5 years	x	x	x	x
5 to 14 years	x	x	x	x
14 years and over:				
Heads of households:				
Owners:				
14 to 24 years	x	x	x	x
25 to 44 years	x	x	x	x
45 years and over	x	x	x	x
Renters:				
14 to 24 years	x	x	x	x
25 to 44 years	x	x	x	x
45 years and over	x	x	x	x
Other household members:				
14 to 24 years	x	x	x	x
25 to 44 years	x	x	x	x
45 years and over	x	x	x	x

Occupied housing units in the ED were classified into four groups by color of the head of the household—white and nonwhite—and by whether he was owner or renter. Vacant housing units were classified into three groups by type of vacancy, i.e., whether the housing unit was for sale, for rent, or "other."

To obtain sample weights for persons and housing units, in an area for which data were to be published, all the numbers in one type of ratio estimate cell for all the ED's in the area were added together. Each resulting ratio estimate cell for the area was then examined to determine if the 100-percent count in that cell was less than 50 or if the weight was over 16. If either condition prevailed, the number of cells among which the total population had been distributed was reduced by successively dropping distinctions between groups—e.g., between age groups 25-44 and 45+, or between white and nonwhite—in a specified order, until the condition no longer existed. As each such distinction was dropped in the main estimation process, the sample count for each cell was adjusted by a preliminary ratio estimate made in the two overall classes corresponding to the dropped characteristic (provided these two classes met the size and weight criteria mentioned above). If no such combinations had to be made, the target number for each cell (the number to which the sample cases in that cell would be inflated) was the 100-percent count. If one or more combinations had to be made in the course of the estimation process, the 100-percent total for each set of combined cells was allocated among the original cells in proportion to the (adjusted) sample counts for those cells to provide the target numbers.

Because the basic blocks of data on the tapes were for ED's, the target numbers for the area were then converted into target numbers for each ED in the area by apportioning the area target number in proportion to the corresponding sample counts in each ED.

After the target numbers for the ED were obtained, each sample person and housing unit in the ED was assigned an integral weight such that, for each ratio estimate cell, the sum of the weights equaled the target number, and such that each of the weights differed from the theoretical (fractional) weight for the cell by less than one. The assignment of integral weights was done to avoid the problems of inconsistency which can arise from successive tabulating and rounding fractional weights in varying cross-classifications. Occasionally, an alternate weighting was adopted because of problems arising out of differences in the 100-percent counts, found after the control tapes were established.

This estimation procedure reduced the component of sampling error arising from the variation in size of household and also achieved some of the gains of stratified sampling, with the strata being the groups for which separate ratio estimates were computed.[1]

COMPUTER OPERATIONS

In computer processing language, computer operation is the process of producing an "output data file" from an "input data file" according to a program or set of instructions stored in and followed by the computer. The operation required (1) mounting a program tape, input data tapes, and blank tapes to receive output, (2) "loading" the program tape into the computer and putting the computer under the control of the stored program by means of which the input data were read into the computer and processed, resulting in output data on tape, and (3) demounting, labeling, and distributing the output tapes to the high-speed printer for printing or to storage for later use as input in further computer operations.

The computer operation involved the attendance of (1) a console operator who manipulated controls to read programs into the computer, in addition to his more general duties of attending to and controlling the computer; (2) an engineer available in case of breakdown of the equipment; and (3) three or four "monitors," or "computer production specialists," who directed and controlled the flow of tapes and programs to and from the 18 computer Uniservos. A control group prepared detailed computer operating instructions for each computer program for use by monitors at the computer console. The instructions specified input and output files, order of monitoring and manner of operation, and labeling and disposition of output tapes. The control group also received original input data tape from FOSDIC, "staged" the tape for use in specified programs in specified computer production shifts, controlled the flow of tape through the high-speed printer, distributed printouts, maintained control records, prepared reports based thereon, and filed and cataloged tape. The magnitude of the operation is indicated by the fact that more than 24,000 tapes were controlled through thousands of different operations consuming as many as 350 hours of computer time a week on five computers in four locations.

In developing the computer time schedule it was originally estimated (early in 1958) that 6,400 hours of operating time on the Univac 1105 computers would be required for processing the 100-percent data. The actual operating time requirement proved to be 3,200 hours.

Although the actual computer time requirements for processing the 100-percent data fell considerably below the original estimates, it appeared that the opposite would prove true in processing the sample data. The original February 1958 projections assumed that about 13,300 computer hours would be sufficient to process the sample data, while the June 1961[2] work schedules indicated that substantially more hours would be necessary. This increase reflected a more extensive program of editing and of tabulation than had been envisioned in early 1958 when the original calculations were made.

EDITING AND EVALUATION OF EDITING

Much is known about the extent of the editing in the 1960 censuses because the computers prepared a record of the editing, and the figures on the amount of allocation and imputation are published along with the results of the censuses. Comparable information for earlier censuses on the number of errors and omissions and the amount of office editing is limited because the time and cost involved in keeping such records would have been prohibitive when the data were processed with card-tabulating equipment.

Some innovations in 1960 reduced errors in the processing and others produced a more consistent quality of editing. The use of FOSDIC introduced fewer errors than the equivalent manual card-punching operation of 1950. Moreover, the extensive use of electronic equipment in the editing process insured more uniform processing of the occasional faults in the data than could have been accomplished by clerical work. On the other hand, the inability of the electronic equipment to read persons' names, place names, and other forms of script, and its inability to perform certain other operations that can be done readily by clerks, introduced a measure of inflexibility at certain points in the processing operations.

The amount of editing performed on census data is generally small, but it varies from subject to subject and from one subject item to another and among geographic areas.

In the processing of the 1960 censuses, allocations (or assignments or imputations) were made when information on personal or housing characteristics was lacking on the schedules or when certain information reported was inconsistent. In addition, remedial action was required for certain other imperfections in the data. The most common situation which required remedial action arose when a given characteristic was not reported or was inconsistent with other characteristics reported for that person. In this census as in earlier censuses, the general procedure used under these circumstances was to draw a value of the characteristic from a distribution of the characteristic for the appropriate subgroup in the population. Thus, a person who was reported as a male relative of the household head, but for whom marital status was not reported, was assigned a marital status from a marital status distribution for male relatives of household heads. Less frequent situations requiring remedial action included, for example, occasional failures in the microfilming process that caused an entire schedule page to be unreadable by FOSDIC. Moreover, some of the marks entered on the schedules could not be read by FOSDIC. When an occasional entire page in an enumeration book could not be read by FOSDIC, the remedial action consisted

[1] For a more detailed description, see p. xxix ff. of any State report in the series of PC (1)-C publications of the 1960 Census of Population.

[2] The date this report was written.

in substituting the persons on the preceding page. When a housing unit was recorded as occupied but the computer was unable to read any marks for the occupants, the data for persons in a previously enumerated housing unit were substituted.

In earlier censuses, the distributions from which assignments were made were derived from previous censuses or surveys. The use of the electronic computer improved upon this procedure by making feasible the use of distribution implicit in the 1960 data being tabulated. In addition, the superior flexibility of the computer permitted the use of a greater number of homogeneous subgroups and thus improved the accuracy of the assignments.

The technique used for the 1960 censuses may be illustrated by the procedure used in the assignment of unknown ages. The allocation of unknown ages was carried out in the following steps:

1. The computer stored reported ages of persons by sex, color or race, household relationship, and marital status.

2. Each stored age was retained in the computer only until a succeeding person having the same characteristics and having age reported was processed through the computer.

3. This stored age was assigned to the next person whose age was unknown and who otherwise had the same characteristics. (To prevent clustered allocations, several ages were stored for each subgroup, to be used in rotation if necessary.)

This procedure ensured that the distribution of ages assigned by the computer for persons of a given set of characteristics would correspond closely to the reported age distribution of such persons in the current census.

Whether the schedule really contained information for a person, rather than a spurious mark or two, was decided on the basis of the configuration of marks on a line provided for a person on the schedule. If the line contained marks for at least two of the general characteristics--relationship to head of household, sex, color, age, marital status--and at least one of these entries was in relationship to head of household, sex, or color, the inference was made that the line contained entries for a person. Names were not used as a criterion of the presence of a person because the electronic computer was unable to distinguish between a name and any other entry in the name space, and the schedules were not ordinarily edited by hand. If a person met the minimum requirements for being counted, however, any missing characteristics were supplied by assignment in the computer.

Similarly, when the item on tenure of a housing unit was left blank and other entries on the schedule did not indicate the type of tenure, then the type of tenure of the preceding housing unit was assigned. There were numerous other situations in which a number was assigned from the data for a preceding housing unit.

Specific limits were established for the number of substitutions and other allocations that would be permitted to be made automatically without further investigation. If the number of corrections for an ED was beyond the limits, the enumeration book in which the errors or omissions occurred was examined visually, and sometimes corrective action was necessary. For example, it was sometimes found that the enumerator had failed to make the marks on the schedule page sufficiently dark to be read by FOSDIC, or had made some other easily corrected error. In these situations, the pages were corrected, remicrofilmed, and rerun on FOSDIC.

Examples of other types of reponse assignment (or allocation for missing entries) included the following: Allocations of color were made for household members who were relatives of the head of the household, by assigning the color of the household head; allocations of marital status and sex were automatically made for persons identified as wives of household heads or as heads of households with wife present; and allocations were made at random for missing information on quarter of year of birth. If the number of bedrooms was not reported or was inconsistent with other items (e.g., the number of bedrooms was reported as larger than the total number of rooms, or no bedrooms were reported for a unit containing four or more rooms), the number of bedrooms was made to be one less than the total number of rooms in units having one or two rooms, and two less in units having three or more rooms.

REVIEWING COMPUTER DIARIES

One of the outputs in the initial computer processing of raw data was a "diary" tape which provided, for each ED, information which indicated the quality of the data. Since more than 270,000 ED's were involved, the information on each was generally confined to 120 digits or less, or one high-speed-printer line. Primarily, two general types of information were supplied. One related to the identification and completeness of the data for each ED, the other to the quality of the data for the ED as indicated by the extent of allocation for missing information. Some other checks of the quality of the enumeration were also established; for example, the computer indicated the number of persons in an enumeration book who were not assigned to any household.

A diary review operation was set up to inspect the diaries printed from the diary tapes, to determine the acceptability of the data and to take corrective action where indicated. The first step was usually a comparison of ED numbers listed on an Advance Transmittal Listing with the ED numbers shown in the diary, to find cases where the ED number was garbled by the microfilm-FOSDIC system. At the same time, the computer counts of persons and housing units were compared with the counts shown on the breaker sheet at the front of the enumeration book. Marked differences between the counts might indicate that the wrong breaker sheet was attached to the enumeration book, or that not all data pages of the enumeration book were microfilmed, or that adjacent ED's were combined when a breaker sheet was missing for the second ED, etc.

The second step in the operation concerned the evaluation of the quality of the data. The microfilms of ED's with high "NA" rates ("not available," or missing entry rates) were inspected on a microfilm viewer in order to ascertain whether the enumerator failed to make the entries or whether FOSDIC failed to read them. In the former case, the ED was referred to a subject specialist who decided whether hand editing and reprocessing would improve the data for the ED. In the latter case, the ED was put through the microfilm-FOSDIC-computer process again.

The result of the diary review operation was that ED's defective with respect either to identification or to quality

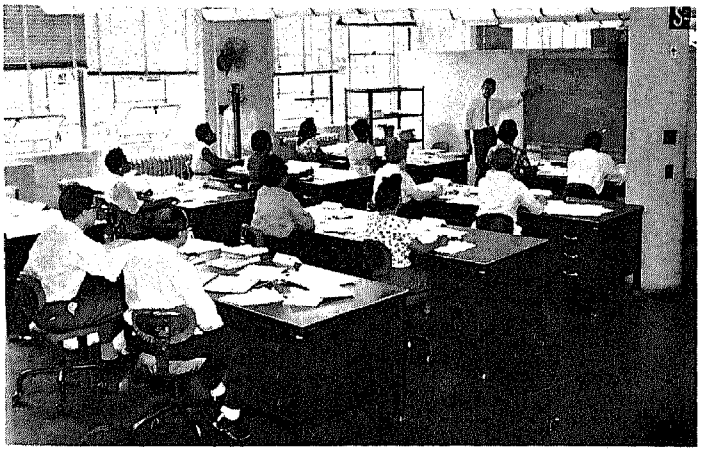


Figure 11. Training Class for Review of Computer Diaries.

and completeness were screened out, corrected or entries edited in, and reprocessed. In the stage-I operation, roughly 10 percent of the ED's were reprocessed for one reason or another. For the sample, the reject rate was expected to be higher, chiefly because of the additional checks to discover sampling biases and because many more characteristics were covered.

The diary review functions divided into two major groups: Those which required examination and correction of the original enumeration books maintained in Jeffersonville, and those which could be pursued in Washington by reference to microfilm and control records. To the maximum extent practicable, the review of the diaries for both work units (complete reels of tape) and individual ED's was conducted in Washington. The correction of the errors located during the review and the remicrofilming of the corrected enumeration books were carried out in Jeffersonville.

BASIC PLAN OF TABULATION

All the data to be tallied from the 100-percent schedules were obtained in the first run of the FOSDIC tapes through the computer. The basic processing of the 100-percent data on the computers involved the following steps:

1. The tape produced by FOSDIC from the information on the microfilmed schedules was used as computer input tape, and was fed into the electronic computer along with a program tape containing instructions for editing and tallying and an ED control tape. The ED control tape listed all ED's which were to be processed and information for those which had been processed to date. The information included total population and housing counts for each ED.

2. This first computer tape produced three outputs: (1) a tally output tape containing counts for each ED and each city block of the population and housing characteristics tallied on a 100-percent basis, (2) a diary review tape which showed how many allocations of each type had been made in each ED as well as other information from which a decision was to be made as to the adequacy of the processing system as well as the quality of data, and (3) an updated ED control tape.

3. After the diaries were reviewed, some of the ED's had to be reprocessed, and the control tape again updated. As soon as it was determined that the tally output was acceptable, the original input record tape was erased, as the tallies contained all the 100-percent information that was to be produced from the censuses.

4. The ED tallies on the tally output tape were summarized for every geographic area for which the 100-percent data were to be published, on a summary tape.

5. The summary tapes were put through the computers, with the instructions supplied by "table preparation program," to produce the 100-percent tabulations. Separate sets of table preparation programs were required for the different series of publications containing the 100-percent data.

The ED control tapes in combination with master identification tapes established the geographic areas for which data were to be published.

For the sample data, the FOSDIC tape went through two different edit runs and produced edited records and diary review tapes. Editing the sample differed from editing the 100-percent data in that the record of the individual in the sample was modified in accordance with the editing rules, because the sample records were to be used for a series of tally operations. For example, in the editing of the sample data, if the mark for sex had been omitted on the schedule for the wife of the head of the household, the tape was changed to indicate that she was female, whereas in the case of the 100-percent data she would have been tallied as a female but her individual record on the 100-percent tape would not have been altered. This was one of the numerous respects in which the treatment of the sample data was more complicated and more time consuming.

Next, the persons and housing units in the sample were weighted, weights being assigned to each individual record through the ratio estimate procedure.

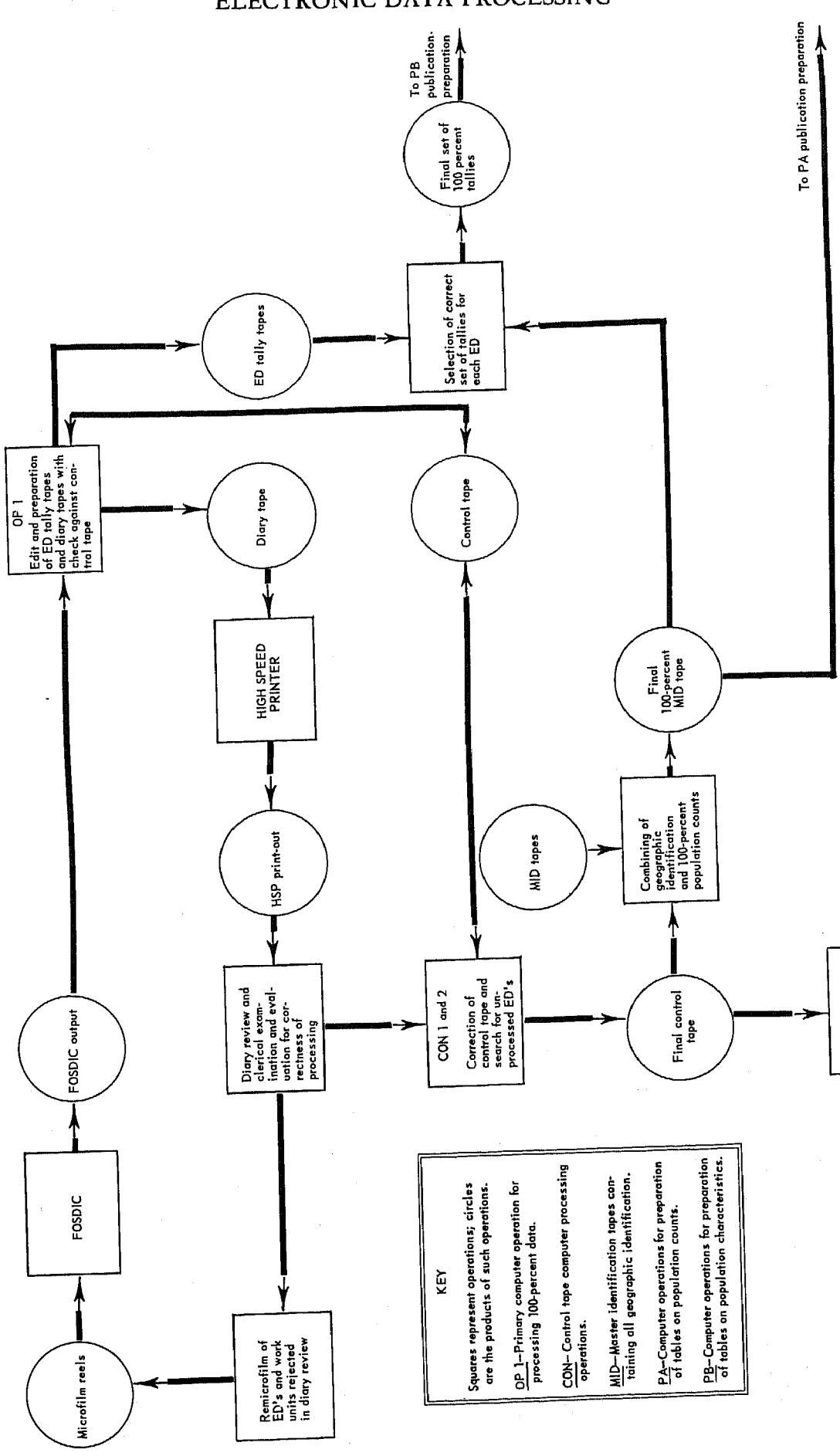
The individual records were then tallied, and summarized for the tables, and when necessary the 100-percent data were picked up from the 100-percent summary tapes for those tables which were to include both sample and 100-percent data.

Although the electronic computers prepared the tabulations, the output of the census computers (like the input) existed only in the form of magnetic signals on reels of computer tape. The final computer operation, therefore, was the translation of the magnetic tape impulses to printed data. This was accomplished on the high-speed printers which produced publication copy at the rate of 400 lines of print per minute.

Special techniques were developed to enable the output of the high-speed printers to be used directly as copy for the subsequent printing operations. Normally, only the insertion of page headings and numbers, table titles, and ruling of column and heading lines was required after the tables were printed out by the high-speed printer.

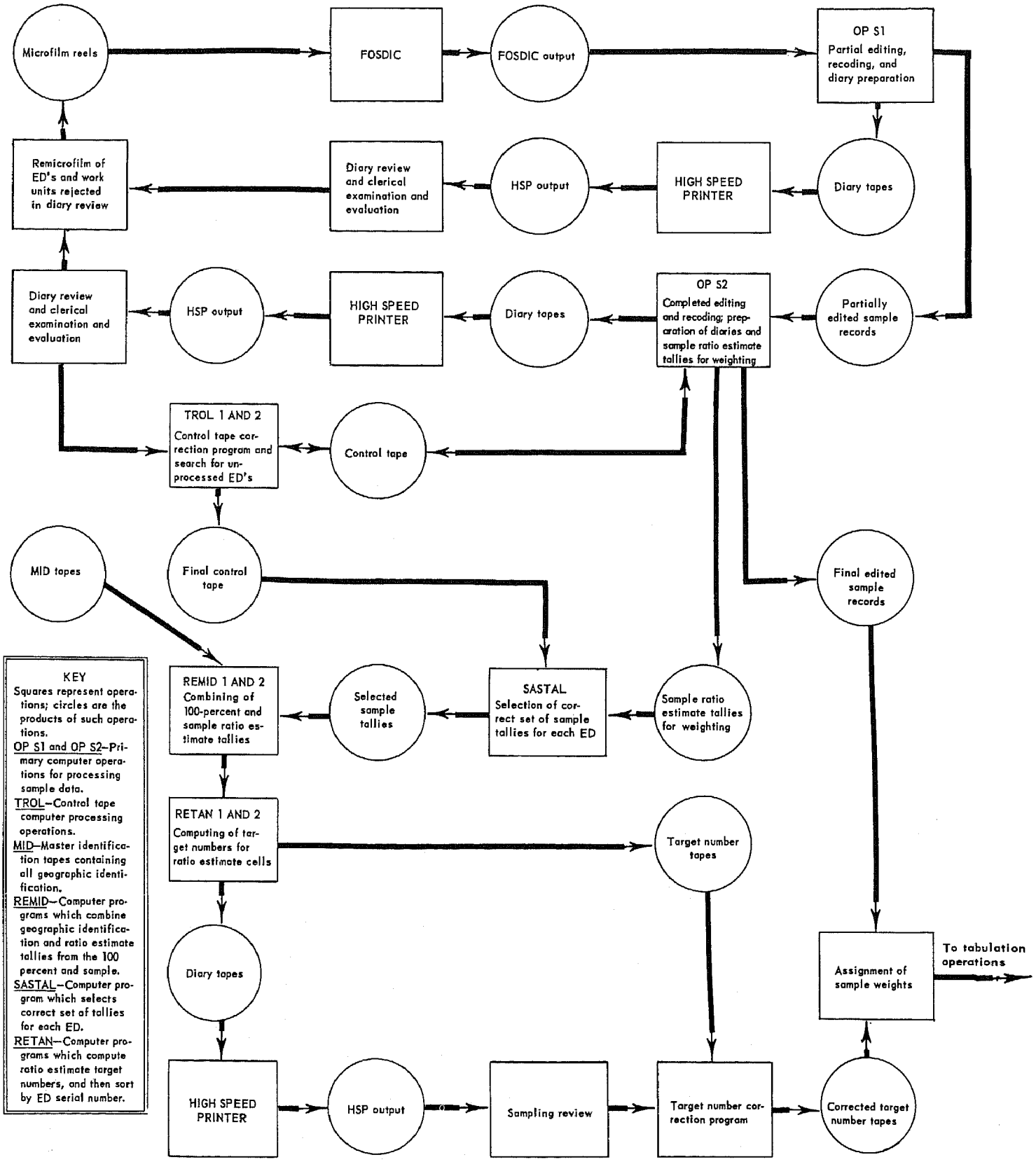
In the process of producing the tabulations for publication, a number of statistics were obtained on tape which could not be published because the publication program was necessarily limited but which will be kept available for possible future print-out and release. In particular, the detailed cross-classifications published only for large areas were often obtained by tallying for small areas and then summarizing to large area total, thus the tallies for small areas are available on tape and can be printed as desired.

PROCESSING 100-PERCENT DATA (Microfilm Through Certified Counts)



*CON 3 was planned but never required.

PROCESSING SAMPLE DATA
(Microfilm Through Sample Weighting)



CRITICAL REVIEW OF COMPLETED TABULATIONS

Ordinarily, the task of reviewing completed tabulations for reasonableness is an analytical function not included in the duties of a data-processing staff. It was recognized, however, that because of the unusual capabilities of the computer, very substantial savings could be made and a more efficient job of analysis could be done if, at the same time that basic publication tables were being produced, a corresponding set of analytical tallies was also produced by the computer. These consisted chiefly of high-speed-printer printouts showing the results of a series of analytical checks. The following ratios, which were prepared by the computer and reviewed to see if they were within limits previously determined to be acceptable, are typical examples of such checks:

Ratio of persons of all ages not in the labor force to persons 14 years old and over in the labor force
Ratio of white foreign-born to native-born of foreign and mixed parentage
Ratio of specified mother-tongue groups to total foreign-born
Ratio of persons living in a different house in the U.S. from the one lived in 5 years earlier, to population 5 years old and over
Ratio of World War II veterans to total males 14 years old and over
Ratio of private wage and salary workers in agriculture (i.e., nonowners not government employed) to total employed in agriculture
Ratio of median income of unrelated individuals to that of families
Ratio of median income of females to that of males
Ratio of dilapidated housing units to total housing units
Ratio of housing units in one-unit structures to total housing units
Ratio of housing units with public sewage disposal to total housing units
Ratio of housing units with more than one person per room to total occupied housing units
Ratio of heads of households who moved between 1955 and 1960 to total occupied housing units

An extensive program of such checks was developed and included in all computer programs dealing with population sample data and with both 100-percent and sample housing data. Time did not permit similar development and programing for 100-percent population data, which were the first to be published, but a modest critical review program that relied on clerical postings and calculations made with adding and calculating machines was developed and carried out for 100-percent data on general population characteristics. For each of the general population characteristics for which tables were prepared for publication, a series of checks was made to see if the data fell within acceptable tolerance bands. The tolerance bands varied with each item and area. Typical examples of such critical review checks were the following:

Nonwhite population as percent of total population
Comparison of percent nonwhite in 1960 to percent nonwhite in 1950
Comparison of 1960 percent distribution of the single, married, separated, widowed, and divorced population per household to the corresponding 1950 percent distribution
Wives of heads of households as percent of total population
Ratio of males to females
Married males as percent of all males 14 years old and over

CONTROL RECORDS AND PROGRESS REPORTS (WASHINGTON)

A comprehensive system of control records was established to cover all aspects of the FOSDIC, computer, and diary review activities. Detailed control records were maintained showing the receipt, location, and contents of microfilm reels after their processing by the film development center. Other records documented the release of microfilm to FOSDIC for processing and its return. Daily reports were prepared and summarized to indicate the progress of work through FOSDIC, the reject rates, and the like. Control records were maintained to show the location, flow, and progress of the FOSDIC output tapes and other computer tapes through the various computer edit and tabulation programs on the equipment in the Bureau and at Chiago, Chapel Hill, and Rome. Control records were kept on the receipt and movement of computer diaries through the Washington diary review operation, on the diary review work in Jeffersonville, and on receipt of diary review materials by and from the Jeffersonville office. At the end of each week a summary report was prepared. The work progress indicated by that report was periodically evaluated in terms of its relationship to the work schedules for stage-I processing of 100-percent data and stage-II processing of sample data. Weekly reports were also developed showing the rate and nature of rejects (see form 17, appendix J).

As work moved through the edit and tabulation phases and reached the point where tabulation runs were made in final table format, control records and reports were developed and maintained showing the location and status, by State, of work on each publication series--population counts, population characteristics, housing counts, block statistics, etc.

A weekly Summary of Status of Computer Programing Activities, showing key dates for completion of major programs and the current status (in terms of percentage completed) of the various computer programs, was prepared.

In addition to the general control records and periodic progress reports described above, the responsible operating units maintained more detailed, supplemental work records used as operational tools and for day-by-day work control. Examples of such records were the State Listing Worksheet, used to record completion of the preparatory steps involved in running the tabulations by State for the PC (1)-B series of publications, and the Distribution of Tables Status Report used in recording the dates various tabulation runs were completed by State.

QUALITY CHECKS

Diary review of the computer output constituted the basic quality check during the tabulation stages of processing the population and housing data. As indicated above, diary procedures required an intensive review, in terms of preestablished tolerances, of (a) the accuracy of identification and the completeness of the data for each ED, and (b) the nature and extent of computer allocations for missing information. Further, they required a determination of corrective action when the tabulated data fell outside the tolerance limits.

In addition to the diary review checks during tabulation, the quality of the data was subject to review and rejection at various other stages during the microfilm-FOSDIC-computer processing. As the last step, displays of tabu-

lated data and of the tables themselves, as well as the critical review output, were reviewed by population and housing subject specialists prior to publication.

On the basis of prior experience in data processing as well as lessons learned in the census pretests, it was recognized that a substantial amount of reprocessing activity was an inherent part of the data-processing system for the 1960 censuses. For the operations on the 100-percent data, reruns through the computer for one reason or another required about 20 percent of the total production time on the computer. Approximately 8 percent represented reruns that were made in attempts to correct defects in the original enumeration or to make adjustments for enumeration errors. The remainder resulted from failures occurring within the data-processing system itself, such as FOSDIC output that could not be read by the computer or that proved to be faulty, etc.

The extent to which work units (reels of microfilm or computer tape) were rejected at various stages during processing of the 100-percent information is indicated below.

Operation	Number of work units processed[1]	Rejected		Remicrofilming required	
		Number	Percent of total processed	Number	Percent of total processed
Pre-FOSDIC review (density check)...	15,021	556	3.7	556	3.7
FOSDIC.....	14,305	767	5.4	767	5.4
Computer.....	12,794	966	7.6	87	.7
Diary review.....	12,604	[2]480	3.8	[2]87	[2].7

[1]The total of work units processed through all operations was 12,604, but reprocessing of work units failing earlier operations resulted in increases in the total numbers processed through such operations.

[2]Excludes 21,034 ED's (7.3 percent of all ED's) which passed diary review at work unit level but failed tolerances at ED level. Of these, 16,636 ED's were remicrofilmed in new work units.

The extent to which substitutions and allocations were made by the computer for omitted persons or for missing characteristics, when an entry was missing, poorly marked, inconsistent, or could not be read by the electronic equipment, as well as the principal reasons for making the allocations and substitutions, and the variations in rates among urban and rural areas, are shown in the following table. The number of persons for whom allocations were made in each category is expressed as a percent of the total persons in that category.

PERCENT OF ALLOCATION BY SIZE OF PLACE

Area	Persons substituted[1] because of--		Persons with 1 or more allocations
	Noninterview	Mechanical failure	
U.S. total.....	0.4	0.1	3.0
URBAN			
Urbanized areas:			
Central cities.....	.6	.1	3.9
Urban fringe.....	.3	.1	2.6
Other urban:			
Places of 10,000 or more..	.3	.1	2.7
Places of 2,500 to 10,000.	.4	.1	2.5
RURAL			
Places of 1,000 to 2,500....	.3	.1	2.4
Other rural.....	.4	.1	2.4

[1]This type of allocation consists of cases where persons and all their characteristics were substituted for an estimated number of omitted persons. Omissions because of noninterview consisted of persons in households for which the enumerator obtained no population data. Omissions because of mechanical failure consisted of full schedule pages which were not properly recorded in the electronic processing system.