

USES OF ELECTRONIC COMPUTERS IN PROCESSING CENSUSES

James L. McPherson, Bureau of the Census

Brazil, Canada, France, West Germany, India, Israel, Italy, Japan, Mexico, Norway, Sweden, the United Kingdom, the United States, Venezuela and Yugoslavia are countries which either are now, or soon will be, using electronic data processing equipment installed in their own territory for Census purposes. In addition, Finland and the United Arab Republics expect such equipment in neighbor countries to process some of their census materials. This we, at the U.S. Bureau of the Census, believe is an exhaustive list. We know there are installations of electronic computers in other nations. But we are reasonably sure that only the countries enumerated above are, at present, using or planning to use electronic equipment to process census data.

Russia

The U.S.S.R. is perhaps the major example of a country well equipped with powerful electronic computing equipment which has not assigned the processing of census data to such equipment. A group of United States computer engineers visited Russia in 1959. They visited several scientific laboratories and saw electronic computers in operation, new computers being built, and still newer computers being designed. It was their general impression that the state of the art in Russia has kept pace with western world development with the exception of the use of magnetic tape as an input-output medium. When a group of Russian engineers visited the United States several months before the U.S. engineers made a return visit, the Russians were quite impressed with our use of magnetic tape. There is evidence Russia has progressed significantly in this area. For example, we recently learned that at least one group is actively working on a mathematical model of the whole Russian economy a la Leontieff. This requires manipulation of matrices of high order for which magnetic tape is almost essential. We believe their problem has been one of producing tape with adequate magnetic quality rather than one of building transports to move the tape rapidly while it is being read from or written on.

However, despite the availability of electronic data processing equipment, the Russian population Census of 1959 was tabulated with punched card equipment. A short undated report entitled "Employment of Computing Machines for the Processing of Statistical Data in the U.S.S.R." describes the processing of the "1959 All-Union Population Census--an operation of great importance and involving a large volume of work."

This report indicates that a decentralized organization accomplished the basic processing. Evidently 209 million "perfocards" were punched and tabulated at 57 statistical boards throughout the union. These computing stations prepared summary cards which were forwarded to the Central Census Computing Machine Station of the Central Statistical Board. There were about 10 million of these summary cards which were run through sorters and tabulators to prepare dis-

trict, city, regional, territorial and republican totals. Photo offset reproduction of tabulator output was used for publication of results.

The report claims that "modernization of computing equipment made for considerable reduction in labour expenditure in processing census data (by approximately one third as compared to the preceding census)." However, the kinds of equipment modification described in the report relate to "relays" and "sorting of perfocards" which suggest that the 1959 Russian Census was processed on multi-column sorter-tabulators similar to IBM 101 equipment rather than with electronic data processing machinery.

The speed with which the work was done is impressive particularly in view of our impression that only manual and punched card techniques were employed. Preliminary results based on apparently manual "sorting of data according to territorial units and packaging of census sheets in lots" were published on May 10, three months after the enumeration. A brief report on the first part of the final results based on punched card tabulations was published February 4, 1960, less than one year after enumeration. And "A third report, giving population distribution according to means of subsistence, social groups, economic branches and professions, and educational levels of mental and manual works, was published in December 1960."

Canada and the United States

The United States processed a comparatively small part of the 1950 Censuses of Population and Housing with Univac I. Almost all of the tabulating for the 1954 and 1958 Economic Censuses was accomplished with electronic computers. Tabulation of the 1960 Census of Population and Housing is being performed entirely with electronic data processors.

Electronic equipment contributed significantly to an important "first" in the U.S. history of Population censuses. For every one of the seventeen censuses of the United States preceding the 1960 Census, the counts for the apportionment of seats in the House of Representatives among the several States were "hand counts." The 1960 counts for apportionment purposes were produced by the electronic data processing equipment at the U.S. Bureau of the Census.

The Dominion Bureau of Statistics of Canada has installed electronic equipment to process the 1961 Census of Population, Housing and Agriculture. Effectively all of the tabulations of these censuses will be accomplished with this installation. (The session on "The Methods of the 1961 Census of Canada" at these meetings is scheduled to include a paper on Canadian tabulation plans.)

Both Canada and the United States aspire to reduce processing time through the use of electronic equipment. This is but one of several objectives but it has been considered (by Canada and U.S.) to be an extremely important one. Processing time can be defined in many ways. However, from the point of view of the

general user of census results, the elapsed time between collection of the data and final publication of the results is the time statisticians spent processing the information and is, therefore, in his mind at least, the "processing time."

A part of the processing time involved in any census is the time spent in input preparation. Here, "input preparation" means conversion of the information recorded on a document by an enumerator or a respondent to a medium of some kind--punched cards, punched paper tape, magnetic tape--amenable to processing through mechanical, electro-mechanical, or electronic equipment. In many types of censuses, some clerical operations such as coding and editing of the information on the enumeration document precedes the preparation of the tabulation medium. But even when the only work which precedes input preparation is of a routine housekeeping nature, the use of mechanized tabulation facilities must await the creation of a suitable input medium.

The Canadian census officials were, to the best of our knowledge, the first to tackle the problem of speeding up the preparation of input. Manually operated card punching machines introduces significant delays in a census processing operation. As early as the 1951 Census of Canada, the DBS statisticians pioneered the use of document sensing equipment. This was an extension of the IBM mark sense machine to accommodate a form over twice the size of the conventional punch card. The Canadian enumerator recorded the answers to the census questions on these mark sense documents and machines, rather than key punch operators, created the punched cards which were then processed through sorters, tabulators and other conventional punched card processing machinery. This was a successful and satisfactory solution to the input preparation problem and the use of document sensing schedules was repeated in the 1956 Census of Canada.

In the mid-1950's we at the U.S. Bureau of the Census sought the advice and assistance of our National Bureau of Standards in connection with the input preparation problem. We described the Canadian use of document sensing to NBS engineers. We indicated that, in our opinion, there were at least two restrictions the document sensing approach imposed that we hoped might be eliminated. One of these was the rigidly specified size of the document which we believed was too small to accommodate the amount of information we wanted to collect per unit of enumeration. The other was the requirement that a specific writing instrument--either an electrographic pencil or a special fountain pen with electrographic ink--had to be used.

Our FOSDIC equipment resulted from our collaboration with NBS. This equipment permits a wide range of document sizes so we can make the dimensions of the form suit the content of the census. It does not require a special writing instrument, although we recommend a medium soft pencil. The input to Fosdic is microfilm images of census schedules. The output is magnetic tape ready for processing by our electronic data processing equipment.

For their 1961 census, the Canadians have similar equipment which they call electronic document reading equipment. The output is magnetic tape ready for their computers. The input is the document itself rather than a microfilm picture.

In both countries the objective has been rapidly to create an input medium for the electronic data processors with a view to speeding up the census processing time. We Censusites in Canada and the U.S. attach great importance to early publication of results and try to act accordingly. There are perhaps one or more other nations which do not share our concern with early publication and there certainly are other countries behaving in a manner which holds little or no promise of significantly reducing the time between enumeration and publication regardless of how rapidly electronic computers may accomplish the tabulation phase of a census activity.

In addition to facilitating a speed-up of the whole census process, electronic equipment can contribute in other important ways to census work. More complex cross classifications of the data than could be economically managed with punched card equipment become comparatively easy with electronic computers. The uniformity with which internal inconsistencies in the basic information can be detected--and in many census type investigations automatically corrected--is several orders of magnitude better with electronic computers than it was with earlier combination clerical-punched card processors. Definitions of derived descriptors of the data can be quite complex thus making them more useful when electronic machines are used to compute them. For example, the U.S. 1960 Census reports on family income will include, for hundreds of geographic areas, the median income of families consisting of man, wife and two children under 18 years of age. In earlier censuses we were able to show only over-all median family income with no control on family composition. For many purposes area to area comparisons will be much more meaningful when the median income relates to a "standard" family.

These are but a few illustrations of ways in which electronic computers can be and are being used to improve the quality of census data. All countries which are using or expect to be using electronic equipment to process their censuses expect to realize this kind of benefit.

Continental Europe

Every one of the six countries on the continent using or planning to use electronic computers for census work contemplates the use of IBM computers. Norway, where a British computer--the Deuce will perform some of the tabulating for the Census of Fisheries, is the only country using any equipment not supplied by IBM.

A convenient classification of electronic computers has three categories--small, medium, large. Since we are statisticians at this meeting, none of you will have difficulty pairing high, low and medium costs with their appropriate mates.

Small electronic computers generally have small, slow memories and very slow input-output

facilities. We know of no country where small computers are used for census work.

Medium computers usually have small but reasonably fast memories; they almost certainly have punched card input-output facilities and may have magnetic tape transports associated with them. The IBM 650's and the 1401 in Italy and the 1401 in Norway are medium computers.

Large computers have large fast memories, magnetic tape input-output facilities and may or may not have direct punched card input-output equipment on line. The IBM 705's for France and Yugoslavia and the 7070's for West Germany and Sweden are large computers. Both Sweden and West Germany use an IBM 1401 as a peripheral device. Typically, in a 1401-7070 installation, the 1401 is used to prepare input for a 7070 by converting data from punched cards to magnetic tape and to accept magnetic tape output from the 7070 for printing on hard copy.

Italy is the only country not only on the continent but anywhere in the world that plans to emulate Canada and the U.S. by using electronic equipment to prepare magnetic tape without the intermediate medium of punched cards. Plans for the Italian population census contemplate use of the same kind of document to magnetic tape machinery as the Canadians are using. Norway, Sweden, France, West Germany and Yugoslavia are not underdeveloped countries in the general sense of this term and they certainly have on their census staffs personnel sophisticated in statistical work particularly insofar as the tools of the census trade are concerned. Without arguing the merits of either side of the case, we can point out that Italy, Canada and U.S. consider it important to employ electronics to accelerate input preparation while Norway, Sweden, France, Germany and Yugoslavia apparently place less emphasis on exploiting electronics for this purpose.

The United Kingdom

The U.K. is using an IBM 705 to process their 1961 Population census. A British machine the Deuce IIA is used for their annual census of agricultural holdings. Another British machine --the National-Elleott 405-- is used for their censuses of production and distribution.

The U.K. census personnel are allied with the continentals (Italy excepted) in the matter of input preparation. That is, they are evidently satisfied to prepare input by manually punching cards. In fact, the plans for processing the 1961 Population Census of the U.K. indicate our British colleagues are rather relaxed about processing in general. One gathers the impression that the philosophy is to use electronic equipment because it is more reliable than alternative machinery and will therefore contribute to producing a higher quality product. There does not, however, appear to be any particular desire in the U.K. to exploit to speeds of electronic equipment for the sake of getting the job done faster.

Latin America

Mexico and Venezuela have medium size computer--Remington Rand SS90's and Brazil has a large computer--the Remington Rand 1105 (which is the same equipment as the United States Census uses).

We understand the Mexicans have just about completed tabulation of their 1960 Census of Population with their SS90 and are currently tabulating the 1960 Census of Agriculture on this equipment. According to the manufacturer, the Mexican computer is not equipped with magnetic tape input-output equipment. The input has been punched cards and output is either punched summary cards or printed copy.

The Venezuelans have used, and are now using, consultants from the United States. After a visit in the fall of 1959 your author recommended early acquisition of tabulating facilities by the Venezuelans for their (then) forthcoming census. In his view it was more important to decide what machinery would be used than to deliberate about the relative merits of alternative equipments. Not until more than a year later was a decision to install an SS90 reached. We understand, from a recently returned U. S. consultant, that the equipment is now installed and ready to begin processing the 1961 Census of Venezuela. Our informant tells us, however, that card punching has not yet started. Evidently the Venezuelans have not yet been persuaded to overlap the various operations involved in processing a census. Thus they concentrated first on getting a hand count of population. Almost all personnel resources were devoted to this task which was completed only a month or two ago. The next step will be the selection of a sample (designed by a consultant from the U.S.) of the household schedules. Presumably not until this sample has been selected and any necessary coding has been completed will the preparation of the input medium begin. Whether and how much electronic computers will contribute to processing the Venezuelan Census remains to be seen.

Brazil embarked on an ambitious attempt to emulate U.S. techniques and procedures for processing their census. An early action in this program was the acquisition of a Remington Rand 1105 computer. We are familiar with many of the problems which attended delivery and installation of this machinery in Brazil. The recent political upheaval in Brazil has slowed information about the use of this equipment to a trickle and what little information we do get relates to technical and engineering matters concerning the equipment rather than plans for its use in connection with censuses.

Other Countries

We know there is, or soon will be, an IBM 705 in Japan for Census use and a Russian built "Ural" computer in India. This is about all we know about these two countries. We must be candid and admit we haven't endeavored to learn more.

Israel originally planned to use punched card handling equipment for all their census tabulations. Recently they decided to use an IBM 1401 at the Mechanization Center in Jerusalem to do some of the work connected with inflating some sample questions and to compute (from data tabulated using electro-mechanical equipment) certain percentages, ratios, averages and medians for each of 800 settlements for which they plan to publish these derived data.

EDITING

Editing of census information involves not only the identification of missing or internally contradictory information but also action of some kind to correct the deficiency. Electronic computers can be depended upon consistently, thoroughly and reliably to apply whatever rules the user specifies to identify those observations which require corrective action of some kind before the data are good enough to tabulate.

Views with respect to just how the requisite corrective action should be accomplished seem to vary widely. At one extreme we would place the United Kingdom. As we understand them, the U.K. plans for the population census contemplates using the electronic computers only to identify cases where information is missing or self contradictory and refer these to humans for corrective action.

At the opposite extreme we would place the Norwegians and Swedes who hold the view that not only can the "untabulatable" observations be recognized by the computer but that the computer can be directed to impute satisfactory and consistent answers by reference to complex mathematical models which can be thought of as multi-dimensional function generators. These function generators would be unusual since the several terms of the function would sometimes be independent variables--the satisfactory answers to census questions--and sometimes dependent variables--the missing or inconsistent answers to census questions for which satisfactory answers are to be imputed.

Our strong view at the U.S. Census on this question is a two-fold one. First we use our computers to both identify need for corrective action and to impute "tabulatable" answers where necessary. Second, and of major importance, we believe any workable method for imputation can be used and consequently it need not be particularly complex.

The important consideration which leads to our position on this question relates to the basic quality of the information we have collected. If we fail to get answers to a significantly large number of questions or if we get internally inconsistent answers for a large proportion of the population enumerated we cannot expect to make "good" information out of "bad" regardless of how fancy an imputation model we might invent. We can keep track in the computer of the number of times we impute answers to questions. Our philosophy is that when it is only infrequently necessary to impute answers, then any imputation technique can do little violence to the subsequent tabulations. When imputation is necessary for a large proportion of the observations we face a serious problem. It is obvious that we have been unsuccessful in our attempt to collect the information. Whether we should (a) suppress publication, (b) publish what we were able to obtain, (c) embark on a re-enumeration activity, or (d) take some other kind of action are decisions that must be made in this kind of a situation. One thing we certainly would not plan to do in such an unhappy event would be to expect our electronic computer to choose an appropriate course of action.