



UNITED NATIONS
ECONOMIC
AND
SOCIAL COUNCIL



Distr.
GENERAL

E/CN.3/301
6 February 1962

ORIGINAL: ENGLISH

STATISTICAL COMMISSION
Twelfth session
Item 17 of the provisional agenda

THE USES BEING MADE BY THE STATISTICAL OFFICE OF ELECTRONIC
COMPUTERS FOR STATISTICAL COMPILATION

(Report by the Secretary-General)

1. In the report of its eleventh session (E/3375), the Statistical Commission suggested work on the application of electronic computers for statistical compilation as follows:

- (i) The Commission noted, as a task requiring additional attention in the future, work on the use of electronic equipment for purposes of statistical compilation and tabulation to meet modern needs (para. 16);
- (ii) The Commission requested the Secretariat to experiment with the use of electronic computers in facilitating the rapid processing, verifying and printing of trade-by-commodity-by-country statistics (resolution 6 (XI), para. 6);
- (iii) The Committee listed as a continuing project of high priority the study of the methodology of automatic data processing (para. 158 (4) m).

2. For some time prior to the eleventh session of the Commission, the Statistical Office of the United Nations had been considering the use of an electronic computer to process the data for its detailed quarterly publication, Commodity Trade Statistics (CTS) (Statistical Papers, Series D), consisting of about 800 pages an issue showing trade-by-commodity-by-country data according to the Standard International Trade Classification (SITC). Because of the bulk of the data and the amount of conversion and re-arrangement to which they had to be subjected, CTS appeared to be particularly suited to processing by computer. Feasibility studies were periodically undertaken but, until 1960, it would have been

so much more expensive to use a computer than to process by conventional punched-card methods that it was not practicable to use the computer. As computers were technically improved, the ratio of cost (per hour of computer time) to speed steadily decreased and at about the beginning of 1960, with the introduction of computers like the IBM 7090, computer costs for the CTS operation became comparable to conventional punched-card costs and, on the advice of the Commission, the change to processing by computer was made. Two issues of CTS have been processed by computer and, at the time of writing, a third is being processed.

3. By 1958, it had become apparent that it would pay to process the data for the annual World Energy Supplies (WES) (Statistical Papers, Series J) by computer. Here the input is very much smaller than for CTS (1,200 punched cards for the former as against about 300,000 for the latter) and, therefore, speed is of much less importance. In WES, a very large number of different aggregates is based on the input data with the result that the output amounts to the equivalent of approximately 10,000 cards. Because a computer can in one operation read a data item, convert it into standard units and add it to all the different aggregates to which it contributes, computers were well adapted to this operation. The only part of the work where speed is required is in sorting the output into the order in which it is to be printed; sorting by computers existing in 1959 proved to be too costly and it was, therefore, arranged that the output of the computer (an IBM 704) should be on punched cards which were then sorted by conventional punched-card equipment prior to tabulation. Now that the faster IBM 7090 is available, the operation has been modified so that it is carried out entirely by computer. Two issues of WES have been made by the original method and one by the improved method.

4. The Commission may be interested to learn what the Statistical Office, on the basis of the experience just described, has found to be the advantages and difficulties of data processing by computer. In the case of WES, the raw data had originally been processed by hand. Because of the many overlapping aggregates which were required, the hand work had been very slow and painful. At its completion, manuscript tables were reproduced by typewriter for photo-offset printing. Compilation by machine did away with this drudgery but, on the other hand, it put great emphasis on accuracy of detail; the input cards had to be exactly in the form and in the order expected by the machine. The first machine operation is a

/...

quick check of the input cards, which (as explained in para. 3) are few in number. In one of the actual runs, they were found to be perfect; in two, a few imperfections were discovered and had to be corrected by the Statistical Office. Once the data input is perfect, the machine begins to process it. This part of the operation has never gone without a hitch, usually because of magnetic-tape failure; incomplete data have been submitted to the Statistical Office by the service bureau doing the work and part of the machining had to be repeated with a resulting delay of a week or two in the production of the final tabulations. Despite this drawback, the use of the computer has led to a great saving of time and resources in the production of WES.

5. The use of the computer for CTS replaces processing on conventional punched-card equipment and permits comparison of the two methods as applied to a large-scale operation. In this case, the size of the input (300,000 cards) makes a preliminary check of the data cards by computer impossible and the computer must deal with aberrant data cards as it goes along, subject to correction at a later stage. Most of the cards are supplied directly by Governments and, so, are in the national format. Only a cursory examination of the cards is made by the Statistical Office before they are machined; the machine itself makes the detailed check. The data cards for a given reporting country may be presented to the computer in any order. Because the machine has to be instructed as to the card forms, the country and commodity codes and the quantity and value units in which national data are expressed, about 10,000 instructions cards have to be provided each quarter by the Statistical Office; many of these can be used without change for successive issues of CTS. As the machine will tolerate no mistake in its instructions, the preparation and placing of such a large number of cards with complete accuracy puts a heavy strain on the staff responsible and it is greatly to the credit of their intelligence and devotion that very few mistakes have occurred. When a mistake occurs, the machine must skip the entire reporting country to which the faulty instruction refers and that country has to be added later. Another cause of errors of this kind is misplacement or omission of cards in transferring them to tape, an operation requiring about eighteen hours. Though the cards are presented to the service bureau in linear order, machine halts caused by bent or off-centre cards and machine and tape failures require a certain amount

/...

of manipulation of the cards by service-bureau staff beyond the mere feeding of them in linear order to the machine. This has occasionally resulted in the transposition or omission of instruction cards.

6. For the last two issues of CTS, the United States Bureau of the Census has kindly arranged to provide data on magnetic tape so that the expenditure of time and the risk of error associated with the card-to-tape machine are avoided. This represents a considerable effort on the part of the Bureau; since its output is on UNIVAC tape, the Bureau must arrange for conversion to IBM tape by a tape-to-tape machine at the Navy Department.

7. Once the data are on tape, computer failures (mostly connected with magnetic tape or the magnetic-tape drives) have occurred two or three times during the running of each issue. These sometimes require temporary suspension of operations, in which case there may be a serious delay before computer time can again be scheduled.

8. When CTS data were processed by conventional punched-card equipment, a "proof tabulation" for each reporting country was made after conversion into standard units and codes. Though various proof totals were incorporated in the tabulation, a good deal of checking had to be done by eye and this work was very time-consuming and much disliked by the staff. The computer now checks, verifies proof totals and, in addition, compares the quantity, value and unit value of every entry with the corresponding entry for the previous period. The staff then has only to examine what the computer finds to be out of line. The computer also makes many regional totals for publication which it was not possible to make on conventional punched-card equipment. These two changes represent a great advance over conventional methods.

9. The time required to process an issue has been much decreased, but this may not be entirely fair to conventional methods, since the work was done internally and often had to be interrupted for considerable periods because of urgent accounting tasks. The arrivals of national data have so far been spaced well enough to make it possible to begin machining a day or two after the arrival of the last data. Because of machine errors and the difficulties of scheduling machine time, it usually takes about two weeks to carry out the twenty-two hours of machine time (eighteen hours card-to-tape; four hours on computer) required under ideal

/...

conditions to complete all work through the checking of the data. Under ordinary conditions, preparation to correct the data errors discovered by the machine takes the Statistical Office another week; complications due to the change to the SITC, Revised, are expected to double this time in the case of two issues. The sixteen hours of machine time (four hours on computer, twelve hours on tape-to-printer) ideally required for correcting errors, making aggregates and printing the tables usually have in fact consumed another week or more. A few days are devoted to scrutiny of the final tables, adding notes, etc. and then the issue is ready for photo-offset printing. If there were no errors or scheduling delays, two weeks (from the receipt of the last data) should suffice to prepare 800 pages for photo-offset. In practice, this time is usually more than doubled. The Statistical Office is investigating ways of reducing the major delay, which is due to the difficulty of scheduling time on the computer. The photo-offset reproduction is done internally and is subject to interruption for work of higher internal priority.

10. The cost now quoted, on the basis of past experience, by the service bureau for an issue of CTS of about 800 pages of tables is \$10,000. This sum covers work starting from data and instruction cards supplied by the Statistical Office and resulting in pages essentially ready for photo-offset. It includes the services of specialized personnel and an allowance to cover the cost of errors made by the service-bureau staff or the machines and minor errors made by the Statistical Office; it excludes programming costs which were essentially all met when, in 1960, a consultant wrote the programme in co-operation with the Statistical Office. The figure of \$10,000 should be compared with the price of setting 800 pages of tables in type from good manuscript copy; this is about \$20,000 in New York. It is difficult to make a comparison with the costs on conventional punched-card equipment because that work was done internally on a low-priority basis.

11. It has, because of lack of resources, so far been impossible to take advantage of the manifold opportunities for statistical analysis available when trade-by-commodity-by-country data are on magnetic tape in standard form. The Statistical Office has only been able to make the aggregates and arrangements appearing in CTS. The benefits of the early and relatively difficult stages of preparing and checking the standard tape have, therefore, hardly begun to be realized at all. Another paper (E/CN.3/300) before the Commission is devoted to the discussion of ways of exploiting these opportunities.