STATISTICAL COMMISSION
Twelfth session
Item 18 of the provisional agenda

MASTER PROGRAMME FOR STATISTICAL COMPILATION BY COMPUTER
(Report by the Secretary-General)

1. At its eleventh session, the Commission listed as a continuing project of high priority the study of the methodology of automatic processing of statistical data (E/3375, para. 158 (4) m). In order that processing by computer be as automatic and, therefore, as fast and as economical as possible, human intervention during processing must be reduced to a minimum. This means that complete instructions in minute detail for processing, coded in a way intelligible to the computer, must be supplied to the computer before the processing begins. These instructions are known as the programme and their preparation is known as programming.

2. Even the largest digital computers can perform only a relatively small number of different operations. Computers achieve their results by being able to perform the operations extremely fast. The programme must present the process as an ordered sequence of the basic operations and even the simplest statistical compilation may thus require hundreds of these, each specified by an appropriate instruction. This set of detailed instructions, coded as required for the computer in question, is called the "object programme", and is said to be in "machine language". An object programme, designed for one make and model of computer, cannot, in general, be used on another.

3. To reduce the labour of preparing programmes and to increase the range of computers to which a programme will apply, means have been developed by which a computer itself can take a programme written in a more general and summary "language" and translate it into an object programme for its own use. The translation is done by a programme called a compiler, designed and programmed for a specific make and model of computer, which will translate the summary language into an object programme. Among common summary languages presently available are FORTRAN and COBOL,

62-04504

the former designed primarily for scientific use, the latter for statistical and business use. The manufacturers of most different makes of computers can supply compilers for translating one or both of these summary languages into object programmes for some or all models of computers they produce.

4. The summary languages are a mixture of symbols and a sort of highly specialized, and often abbreviated, English. They cannot be used without special training, and are no doubt very difficult to learn for people who do not speak English. Moreover, the English phrases must be written exactly as prescribed or the compiler will be unable to interpret them. For example, the instruction in the COBOL language to multiply reads "MULTIPLY a BY b GIVING c". If this were written "MULTIPLY a BY b GETTING c", the compiler would reject the instruction. Disadvantages of this sort are no doubt unavoidable in a language which can, on the one hand, express a very large variety of operations and, on the other, be converted by a computer into detailed instructions for its own use.

5. It has appeared from the studies of the Statistical Office of the United Nations that the stage has been reached when most of the normal operations of statistical offices engaged in the collection, compilation and tabulation of economic and social statistics can be described for computer use in a far simpler way requiring neither specialized language nor specialized training. This is made possible by the fact that the types of processing normally performed by a statistical office form a set far less general than that to which languages like FORTRAN and COBOL are intended to apply. The instructions for the use of such a method would be in a plain language (English, French, Chinese, etc.) and readily translatable from one to another by any really careful and conscientious translator.

6. A master programme of this type would be composed of sub-routines, each designed to perform an operation, such as those listed below, in sufficient generality to meet ordinary statistical requirements:

    (a) transforming given data into standard form,

    (b) verifying and editing data in standard form,

    (c) performing arithmetic operations on data in standard form,

    (d) transforming the codes identifying data in standard form,

    (e) merging, sorting, and selecting data in standard form,

(f) performing certain special operations, such as index number calculation, seasonal adjustment, etc.,

(g) tabulating data in standard form.

7. The application of each sub-routine to a particular case would be determined by control cards described in plain language, which could be punched in any statistical office. They would instruct the master programme how to specialize the general sub-routine to fit the particular case in question. To accomplish a sub-routine, the computer would be fed:

(a) the master programme on cards or tape for the sub-routine, prepared once and for all (perhaps through the use of FORTRAN or COBOL) by a central agency, such as the Statistical Office of the United Nations, and distributed to users,

(b) the control cards just described,

(c) the input data to be processed.

8. A statistical office wishing to use the master programme on a specific problem would first analyse the problem into a succession of sub-routines, then prepare the particular control cards required for each sub-routine, make a test on a small sample of input data to verify that all control cards are correct and, after any necessary correction, proceed with the processing. If similar data were to be processed periodically, the same control cards could, of course, be used each time with minor changes to take care of changes of date and so on.

9. The programme for the processing of data for Commodity Trade Statistics is analogous to some of the sub-routines described in paragraph 6. It accepts national trade-by-commodity-by-country cards supplied by Governments in the national format. Control cards instruct the computer as to the card columns in which the various codes and data figures are to be found, how to convert the national currency into dollars, the quantities into metric units, and the commodity and country codes into the SITC, Revised, and standard country codes. Other control cards determine what regional totals are to be made, what checks are to be applied to the data and how the final results are to be tabulated. For an input of 300,000 data cards from thirty-six countries, there are about 1,200 control cards. The latter were made, on the basis of instructions in plain English, by statistical personnel without training in the use of computers or in programming. Almost all of the control cards can be used each quarter without change. The generality already

achieved in the <u>Commodity Trade Statistics</u> programme and the success of the instructions for preparing control cards has encouraged the Statistical Office to begin trying to expand the generality so that a class of statistical operations will be covered embracing about 90 per cent of the operations normally performed by a national or international statistical service. This more general project is in the early stages, and the views of the Commission are sought on the direction it might take and the priority to be afforded it.