

SOCIO-DEMOGRAPHIC RESEARCH UNIT  
CENTRAL BUREAU OF STATISTICS OF NORWAY

January, 1972

Erling Berge:

PROGRAMS AND SYSTEMS IN THE CENTRAL BUREAU OF STATISTICS OF NORWAY

Introduction.

Some of the programs and program systems developed to meet the special needs of the Central Bureau of Statistics of Norway may have a wider interest. An institution like the Bureau usually handles very large data bases, and we have utilized automation for the production of statistics in a high degree.<sup>x)</sup> Our computer is a fairly common one, namely the IBM 360/40. The most interesting item reported here may be the DATSY system. DATSY was developed to facilitate work with models of economic analysis and planning, but its general characteristics make it easily adaptable to both similar and dissimilar purposes. It is now implemented on a UNIVAC 1108.

Description of the computers.

a. IBM 360/40.

The machine has 128 KB of main storage, 2 disk-drives (Memorex 660) with 140 mill. bytes of direct access storage, 6 tapedrives of which 1 are 7 track and 5 are 9 track together with one card-reader and one line-printer. The operating system is System 360/Disk Operating System release 26-1 with the spoolsystem GRASP II.

b. UNIVAC 1108.

The machine used by the Bureau is a commercial installation used both for batch processing and for time-sharing applications. It has 192 KW (36 bit word) of main storage. The direct access devices contain a total of about 400 mill. bytes of storage (the devices are four FH-432, two FH-880 and some Fastrand drums). A total of 10 tapedrive are

---

x) P.J. Bjarve and S. Nordbotten: Automation of the Production of Statistics. Artikler nr. 28, Central Bureau of Statistics of Norway, Oslo 1969.

available (of which 2 are 9-track), each reel of tape capable of storing 20 mill. bytes. Beside this there are papertape-equipment, a calcomp plotter, and of course card-reader and line-printers. About 15 remote job entry terminals and 40 terminals of the TTY type are attached at present. The operating system is EXEC-8 level 26.

### Description of the systems and programs.

#### a. DATSY.

Data Treatment System is a user-oriented data-language developed at the Norwegian Computing Centre in collaboration with the Central Bureau of Statistics under contract with the Planning Division of the Ministry of Finance.

Most of it is written in FORTRAN (with some extensions), and it is now implemented on UNIVAC 1108 under the EXEC-8 operating system.

In a sense, DATSY combines properties both from a programming language, a program package and an archive system. The least developed part of these is the program package.

One class of statements is called directives. The function of a directive may be compared with the use of a subroutine within FORTRAN (e.g. from IBM's Scientific Subroutine Package). It is fairly easy for an experienced programmer to implement new directives either through a FORTRAN program or through a combination of existing directives. DATSY might therefore be developed to contain features from existing packages. The archive system is closely integrated with the computational part. It contains all information relevant for processing without explicit reference, and the introduction of new data is straightforward whether it comes as results from a DATSY-program or from outside the system. Output is not standardized, but easily managed through the print directives. For example:

PRINTMAT matrix list list record

will print a table with heading from record, names of rows and columns from list and list respectively, and figures from matrix. All this data may come from the archive. From the inexperienced user's point of view it seems easy to learn. The system was developed to meet needs of research teams where everyone need not know what the others have done. Therefore, the need for both program and data as self-defining documents was stressed.

Today the system seems to offer the best opportunities for those working with development of new methods and models requiring large data bases, but, as we mentioned, implementation of new directives is fairly easy and the tailor-made package may be just around the corner. The minimum configuration for implementation on UNIVAC 1108 is at least 100 KW of main storage (allowing 65 KW for EXEC-8), 150 000 bytes of direct-access-storage and 3 tape-drives. The program itself requires approximately 35 KW of core to operate with the present overlay (i.e. segmentation). At present, it utilizes 1 mill. bytes of direct-access storage and up to 16 mill. bytes of sequential-access storage.

b. NATBLES<sup>x)</sup>

NATBLES has been developed in the Central Bureau of Statistics to produce edited tables of time series ready for publishing. The programs are written in COBOL and System 360 Assembler. It was first implemented in 1969 on our IBM 360/40, and it has been subject to improvements and extensions since then.

The information used by the system comes from its own archive and/or from cards prepared by the user. Use of the system is fairly easy if one has a detailed knowledge of the data base one wants to use.

---

x) Nasjonalregnskaps Tabeller - Brukerorientert Lagring og Ekstraherings System (National Account Tables - Useroriented Filing and Extraction System).



The tables produced by the system may be saved on tape and the numbers used in further processing. By a planned extension, both the text and the numbers may go back to the archive as new items.

Because the archive usually contains texts from earlier tables (these must now be introduced into the archive through a separate run), its main feature is the ease with which one can produce tables published at regular intervals, or many tables with fairly similar text (many variables against a common one). If one acquires data at regular intervals, it results in a time series. The NATBLES system is limited to handling this kind of data, or data which can be made to look like a time series. Generally, one may say that the more tables one wants to produce, the more obvious are the advantages of the NATBLES system.

The configuration requires for implementation on an IBM 360/40 is at least 60 KB of main storage 150 000 bytes of direct access storage and 5 tapedrives.

#### c. KOKS.

KOKS is a system developed in the Central Bureau of Statistics to detect errors in a tape file and later update the file with corrections from cards.

It is written in System 360 Assembler and is implemented on the Central Bureau of Statistics's IBM 360/40. The controls undertaken include a check of number of cards per case, a check of consistency within records (if both the sum and the items of the sum are given) and validity control (if only certain codes are permitted). The records size may not exceed 1 000 positions.

One can think of the system as a data-preprocessor, used before the data goes into existing archives.

The configuration required for implementation on an IBM 360/40 is at least 48 KB of main storage, 150 000 bytes of direct access

storage and 2 tapedrives.

d. BRAFI.

This program was developed in the Central Bureau of Statistics to meet the needs of the Interview Survey Division. It is written in the System 360 Assembler and implemented on our IBM 360/40. It gives a bivariate frequency distribution as a result.

The variable defining the columns may have 17 codes (blank and 00-15) and the variable defining the rows up to 101 (blank and 00-99). One may use up to five filter variables. The options for the output are: a) absolute numbers (from a separately defined variable), b) frequencies, c) relative frequencies (for rows, columns or total), d) means (within the cell or over the rows or columns for a separately defined variable). The data is given as a tape file with standard DOS-label and a record size less than 1 200 P. In principle, one may create an unlimited number of tables in each run (some of the parameter cards may default from table to table), but in practice one is limited by time since only one table is created for each pass of the data (even though the tape is read both forwards and backwards). The configuration required for implementation is at least 47 KB of main storage and 1 tapedrive.

e. STP1 and STP2.

These programs developed in the Central Bureau of Statistics are mostly conventional table-generating programs even though they contain features not commonly found in such programs. They are written in System 360 Assembler and implemented on our IBM 360/40. The STP2 program can take up to 9 nine-digit variables against 1 one-four digit variable (but one can have only 13 columns of output). The program allows direct recording of the column variable (bracketing if the variable is continuous), and up to five filter variables. The results from the program include any or all of the following options:

a) frequency distribution, b) relative frequency distribution (for rows only), c) the sum of absolute values of up to 10 separately defined variables, d) the mean of the 10 variables. The record size may not exceed 256 positions and the file must be sorted according to the row variables before the program starts. The difference between STP1 and STP2 is that in STP1 there is no column-variable in the usual sense. Each column must be defined by the user, and he has almost the same options as in STP2 (not relative frequencies). In addition, one may define columns as expressions between constants, variables and/or columns previously defined. The configuration required for implementation is 30 KB of main storage and 1 tapedrive.

Further references:

Inquiries about DATSY should be addressed to

Norwegian Computing Centre,  
OSLO 3, Norway

Inquiries about NATLES, KOPS, BRAFI, STP1 and STP2 should  
be addressed to

Central Bureau of Statistics,  
Oslo-Dep., OSLO 1, Norway