Electronic Data Processing of the Population and Housing Census in Romania

(A paper indited by a group of research workers from the Central Statistical Board—Romania)

The population and housing census of March 15, 1966 was the first census in Romania processed by electronic means. This paper presents the main aspects and the experience gained in using electronic equipment for that purpose.

1. PREPARATORY WORK FOR ELECTRONIC PROCESSING OF CENSUS RETURNS

Preparation of the program for census returns processing

Much more extensive and comprehensive than in the previous censuses conducted in our country, the program for processing the population and housing census returns of March 15, 1966 includes a series of 123 basic tables, of which 58 for population characteristics and 65 for the characteristics of buildings, dwellings and households.

If we also add the variants of tables generated either by the structure of tables or the programming solutions adopted, the total number of processed tables amounts to 278.

In preparing the processing program materialized in tables, account was taken of the data requirements for the population, buildings, dwellings and households number and structure needed for the substantiation of the socio-economic development plans. The model tables presenting census returns were finalized after taking the advice of principal users: the state central bodies, the specialized research institutes and other authorities interested in the population and housing census returns.

The tables included in the processing program were classified according to certain characteristics.

Thus, the tables on population may be classified as follows:

— Tables showing total resident population according to geographical characteristics (usually included in the stub of tables), some demographic characteristics (sex, age, marital status, nationality, etc.), mother tongue, citizenship and place of birth;

— Tables showing the fertility of the female population aged 15 and over;

— Tables showing the educational level of the population aged 12 and over;

— Tables showing the economically active population by demographic characteristics (age, educational level), the branch of economic activity, the occupational category and group;

— Tables showing the economically inactive population according to source of livelihood, sex, age, social category, educational level, etc.

The tables presenting the buidilngs, dwellings and households were classified into:

Tables relating to residential buildings, non-residential buildings which contain dwellings and collective buildings classified according to various characteristics;

— Tables relating to all occupied and unoccupied dwellings;

— Tables relating to occupied dwellings;

— Tables relating to housing units where people are living by dint of necessity;

— Tables relating to the structure of households and the way they occupy the dwellings they live in, etc.

Depending on the classification adopted, the tables were prepared according to two criteria:

a) Tables having a variable structure of the lines;

b) Tables having a fixed structure of the lines.

The tables with variable structure of the lines showed the classification of characteristics according to their territorial (geographic) structure. For such tables, this classification provided nine classes according to their structure, the geographical level, the number of indicators included in the stub for each geographical level and the arrangement within the page (number of spaces between lines).

In order to obtain the nine classes it sufficed to have only six primary classes from which the remaining ones would be obtained.

The primary classes were the following:

Class 1 : region, district, commune, village;

Class 3 : region, district, town, including sex items;

Class 5 : region, district, town, including building items;

Class 6 : region, district, commune;

Class 7 : region, town;

Class 9 : region, district, commune, including dwelling items.

The way of obtaining the derived classes is:

Class 2 is obtained from class 1;

Classes 4 and 8 ore obtainend from class 3.

In preparing the tables the following was taken into account: the class size, the need for a detailed presentation of characteristics, the number of pages of the printed table, etc.

It should be mentioned that some tables included both absolute and relative (e.g. percentages) figures. There also have been prepared some special tables having a heading with a fixed structure while the body contained an enumeration of the towns of our country.

Elaboration of census registration forms and of the coding system

The quality of census results is determined to a high extent by the quality of the registration forms. While the errors that have slipped through in the other stages of census operations can, in the majority of cases, be corrected, those resulting from an inadequate design of the registration forms and from the questions included in them are not generally liable to corrections.

In constructing the forms, account was taken of the need to make them adaptable to the registration method which is to be used, on the one hand, and to the processing techniques and type of equipment with the help of which the processing is to be carried out, on the other. They should also secure the collection of all the information required in preparing the tables provided in the processing program.

As to the types of forms which may be used for registration, such as: individual forms, forms for a single household or dwelling and forms for several households or dwellings, it is difficult to say which of them is best suited to electronic processing, for the practice of countries having used electronic facilities for processing census data shows countries having used individual forms, as well as countries having used forms for a single household or dwelling or for several households and dwellings.

The same applies to the way of writing the questions on forms — on lines, in columns or in boxes. Each way has its own advantages and disadvantages which should be taken into account in constructing the forms. Two ways of entering the data were simultaneously used in the census conducted in Romania on March 15, 1966: in columns (in the form for recording population characteristics) and in boxes (in forms for recording buildings and dwellings).

Taking into account the material resources available as well as the specific conditions of our country, the forms used for recording the answers were constructed in such a way that the following requirements should be met:

a) to be filled in by the census personnel specially trained for this purpose (census takers) from the statements made by individuals and not by self-enumeration;

b) to enable the coding of answers to be made on forms with a view to transfer the results on punched cards;

c) to be constructed so that the transfer on punched cards by the system of codes adopted could be used in the electronic processing of results.

The three basic forms were the following:

Form P_1 — "Lists of household members", where all persons belonging to a household were recorded as well as all persons temporarily present in that household at the time of the census;

Form L_1 — "The building", filled in for all residential buildings or non-residential buildings which contain dwellings (including buildings which contain collective dwellings);

Form L_2 — "The dwelling", filled in for all dwellings regardless of whether they were occupied at the time of the census or not.

Unlike form P_1, where the majority of charactristics were coded in a centralized way, the recording of the great majority of characteristics on forms L_1 and L_2 was made through direct coding by the census taker during the registration.

Thus, a number of codes and lists of the geographical units, the classification of branches of activity, the occupational classification, the socio-professional group, the type of ownership of buildings and dwellings, the number of storeys, etc., were worked out by taking into account the specific features of the processing work.

It should be recalled, for instance, that the coding of branches and sub-branches of national economy (which included 5 digits) enabled by its very structure a general classification by branches of national economy (making use of the first two digits) as well as a more detailed classification by branches and sub-branches of economic activity (making use of the 5 digits of the code), each classification being used in the various tables of the processing system according to the level of detail of the respective classification.

For example: Industry .....Code 01
Fuel industry .....Code 01.02
Coal industry .....Code 01.02.1

In order to enable a rational processing of data to be made, the following types of cards were prepared by taking into account the se tof tables presenting census results:

K_1 — the individual card	(19,103 thousand)
K_2 — the building card	(4,518 thousand)
K_3 — the dwelling card	(5,670 thousand)
K_4 — the household card	(8,069 thousand)

In order to obtain combined results (dwellings—buildings, dwellings — households) a system was developed, which enabled the inclusion into the population of a unit of observation-information on units of other populations as well.

Thus, form L_2 — "The dwelling form" included (in the form of codes) some characteristics of the building which contained the dwelling: type of building, type of residential building, building material of walls and floors, etc. Also in form L_2 some information about the characteristics of the household were taken over, in the form of codes, from form P_1: the number of economically active members, the type of household, the social category of the head of the household, the number of wage-earners in the household.

The duration of a census processing is also influenced by the manner of data collection; it is worth mentioning the technique of total observation for a minimum number of main characteristics, the remaining ones being covered by a sampling program.

Thus, a rational combination of thorough processing and sampling processing would be effective in reducing time, labour and material expenditure arising in electronic data processing operations. At all events, it is obvious that an intensive use of the sampling method is liable to provide important advantages both in securing a rapid processing and in the quality of data subject to processing. In our case a sampling census was made in order to study the living conditions and the equipment of households with durable goods. In this way some details of the census could be studied more thouroughly.

2. ANALYSIS AND PROGRAMMING OF THE CENSUS DATA PROCESSING

General description of the system of programs

The processing (Appendix) includes the following stages:

Conversion of cards on magnetic tape

The first stage of the setting up of the exact copies is necessary in order to avoid the frequent handling of a very big quantity of punched cards, this being most inconvenient and bringing about serious difficulties in the sorting of the records.

The logical control

Errors in the data in the punched cards are unavoidable, they are due to wrong recording, coding or punching. Such errors will lead to incompatibilities in the tables. It is therefore necessary to establish a system of data control and data correction. The basic idea of logical control programs consists in the rejection of the records which do not contain correct information for several principal indicators which afterwards lead to the fixing of values for auxiliary indicators. Furthermore, records which have undergone more than four corrections are not being accepted because it is supposed that in such cases the information contains too many inaccuracies to be accepted. Eventually with a view of carrying out a more serious control of the records, certain indicators which have a very low frequency (e.g. records of individuals aged over 100 years or having more than ten children) have been rejected. Rejected records were corrected manually and, if erroneous, re-punched, passed through the programs of logical control and merged with the file of correct records.

There are two types of automatic corrections which are being carried out with the not accepted records:

— *systematic corrections*, which correct an indicator corresponding to the value of another indicator on the basis of an established algorithm;

— *random corrections*, which either fix the value of one indicator in accordance with the value of a corresponding indicator of a previous record having certain characteristics similar to the record which is being corrected (hotdeck method), or completely random corrections. An example of of hotdeck correction is the correction of the birth date (year, month and day) which gets the constant value of the preceding record. An example of random correction is the correction of the sex of persons aged under 12 years which have been alternately classified male or female.

It is clear that a difficult question in the logical control is the flow of controls depending on their priorities.

Another important problem is the establishment of a correct sequence of the records in files. In all cases the cards have been sequenced according to the identification characteristics. However, generally this flow was not always ensured owing to the existing errors in handling and punching. The logical control programs had to provide signalization of cards which were out of sequence. These records were sent onto a separate file, afterwards they were sorted, passed to the logical control program and merged with the main file.

The logical control carried out for houses, dwellings and households consisted of two stages:

In the first stage, every single file was checked, which means that the compatibilities of the information within each record were checked. In the second stage, the information contained in diferent files was checked: e.g. the household file was correlated with the dwellings, the dwelling file was correlated with the building file. This was due to the fact that certain data appeared in several archives (e.g. the number of dwellings can result both from the building file and the household file). This intercorrelation presented rather difficult problems. However, this difficulty could have been avoided if the printed forms had been laid out otherwise. But this has not been possible, because at the time the printed forms were laid out we did not know the type of equipment which would process the census.

The logical control programs made only a check of the compatibility at the level of each record. A global check of the validity of the data was also necessary and in our case it was done working out control totals for a limited number of indicators (number of persons, nationality , etc.). Only after these totals had been checked, were the files considered correct. The checking of these totals was carried out by comparing them with statistical data known beforehand or with preliminary data. A further control has been possible by comparing these data with figures resulting from the cards which had been processed mechanographically.

Elaboration of raw tables

The working out of the tables was principally made by means of two different types of programs in accordance with the variable or fixed structure of the tables. As most of the required tables could be stored within the case store, the programs worked out many tables simultaneously. Each table was sent on a different file on magnetic tape. An average of four tables was worked out by a tabulation program. On the table file each row of the table represents a single record.

As the time for the programming was very short, it was not possible to write a generating program which would generate the tabulation program starting with a description which should be accepted by the generator in a parametrical form. Therefore, different programs were written for several groups of tables. However, in order to facilitate the programming, all the programs were written using the same general scheme.

The tables which had been worked out for several geographical levels were separated on the table tape by special small-sized blocks which contained data for the identification of the table.

The working out of the final tables

In this phase the final tables were worked out on the basis of the tapes containing the gross tables which had resulted in the previous phase. The processing in these phase consists of the following principal operations:

— the concentration of several tables on a single file according to the type of processing which was to be carried out;

— the totalization among the different geographical levels of the same table;

— the totalization within the geographical level of a certain table;

— the computation of auxiliary rows and columns;

— the computation of percentages;

— the rearrangement of tables within a given file.

At the end of this phase files containing many tables were obtained, exactly in the form required by the printing programs.

The programs for this phase are general and they use parameters according to the tables which are being worked out. For instance, there was only one program for the computation of percentages.

The printing of the tables

The printing of the tables is a very delicate problem because the print-out was required in a rather sophisticated form in order to allow the publishing by direct multiplication of the print-outs.

The printing program used two files on magnetic tape: the control or text file and the table file. The text file contains data about the printing of each page heading and the printing of every line and general information about the way of printing. The format of this file is standardized and thus by means of two printing programs it was possible to print all the tables.

For the creation of the control file a package of programs has been used, which generates a tape using as parameters coded descriptions of the way of printing. These parameters are written in standardized forms quite simply, so that they may be written by a personnel having no special qualification. The same package of programs contains programs for the correction of the control file for its alteration and for its listing. With this system it was possible to carry out the printing operation with a high degree of flexibility.

Management of the analysis and programming tasks

The elaboration of the programs comprised the following stages:

Analysis of tables and cards

In the first stage, the analysis of the tables consisted in the preparation of a number of symbolical descriptions of the tables in order to avoid ambiguities in their lay-out. For this purpose a symbolic system corresponding to these definitions was worked out. Furthermore, it was necessary to define a number of notions in order to avoid ambiguities. Also in this phase a list of indicators and of their values was established in order to verify whether all the data in the tables were to be found in the cards.

Analysis of indicators and logical control specifications

In this second stage the defining of the specification for the logical control was established. Therefore, a standard form has been worked out wherein all these conditions were put very clearly. The conditions were fixed by the Analysis Department and checked by statisticians.

Definition of the general processing flow-chart

In this phase it was possible to determine the general concept of the computer processing and its framing into all the operations which had to be carried out. Also at this stage the programs were established and the tasks of each program established. A first provisional planning and also the responsible of each program was fixed.

Definition of the program tasks

A detailed specification for each program was worked out. To this end, the exact structure of all the files on magnetic tapes was determined in order to avoid incompatibilities among the various programs.

Programming

In the programming stage the subroutines were defined, written and tested at the beginning, and only afterwards the programs were written using these routines.

The chief objective in the conception of the programs was to carry them out in the shortest possible time with maximum efficiency from the standpoint of the utilization of the computer and the duration of the actual processing.

With this in mind, standard subroutines were written for the programs of logical control and table elaboration, which facilitated very much the writing of the final programs. This shortened the writing time of the programs and led to a standardization of the writing which made the debugging easier.

For the tabulations, the first idea was to write a general program which would generate the tabulation program for each table using an adequate description of the table. Such a program was written in FORTRAN and in PLAN for the input and output sections.

Unfortunately, this program was very slow, especially due to the inefficiency of the FORTRAN compilation. Therefore this idea was temporally abandoned and separate tabulation programs were written instead for each group of tables. Also in this case the method of standard subprograms was used with very good results.

Finally, the preparation of the texts, the printing, the totalizing and the additional processings were carried out using not too many programs. This was carried out only by observing a very strict discipline in working out the documentation of the programs and in co-ordinating the whole work.

The stages described above were not carried out strictly in this linear order. Of course there were changes of specifications. It must be mentioned that when establishing a system of such big dimensions it is necessary to avoid numerous changes in the final stage of the lay-out of the work, because such changes are rather costly and sometimes they lead to unforeseen and disagreeable consequences.

It has been found that when utilizing an electronic computer the specification of the task must be much more accurate and rigorous. However, such specifications could not be obtained in the initial stage from the statisticians.

Program testing

The testing was done mainly in multiprogramming; this shortened the testing time 2 or 3 times:

The performances of the written programs are the following:

	Internal storage (K words)	Processing duration (Hours)
Logical control	$M = 13.5 \ldots 17$ (average value for approximately 35 indicators/record)	$T = (1.6 \ldots 2) \times 10^{-3} . I . t_a$ (average value for approximately 35 indicators/record)
Working-out of tables with constant structure	$M = 12.5 + 0.4C$ $^{-30\%}_{-25\%}$	$T = (4.9 + 0.22C)10^{-5} I_{ta} \pm 25\%$
Working-out of tables with variable structure	$M = 3.5 + 2.5 N_e$ $^{-8\%}_{-10\%}$	$T = (2 + 4.5 N_e) \times 10^{-5} \pm t_a \pm 50\%$

NOTATIONS:

M = internal storage
T = processing time
C = number of cells of the table (K words)
I = number of records
t_a = average time/operation (μs)
(12 μs for ICT-1905)
N_e = number of tables with variable structure simultaneously tabulated.

3. THE RUN OF THE CENSUS PROGRAMS*

General processing scheme

The processing of the population census data comprised the following main operations:

Punching and punching control

The cards were punched by monoperiodic card punching machines ICT.

The punching scheme for the population comprised 67 columns of the 80 columns, out of which 13 were punched automatically.

About 20 million cards were punched with an average of 140 cards per hour on a punching machine.

The errors amounted to about 0.5 to 2.5% of the total number of cards.

In the last stages this percentage was smaller owing to the fact that at the beginning untrained personnel for the punching and the punching control was used.

The translation of the data onto cards was checked by control machines with an average rate of 210 cards per hour. In order to check that the control instructions had been exactly carried out about 5% of the total number of the cards were checked once more.

Obtaining the total control figures

In order to make certain that all the printed forms were punched onto cards, control figures were obtained using the cards for the various geographical levels.

The control figures were obtained by handling the punched cards mecanographically. These figures were compared with the figures worked-out in the coding stage, with the figures obtained in the conversion stage and with those of the preliminary results of the census.

Logical control and automatic correction

In the beginning, the conversion of the cards onto the magnetic tape on the basis of a single program had to be carried out at the same time with the operations of logical control and automatic correction. Later on, it was decided, in view of the results of the experimental processing which had been carried out, that it would be more efficient, in the case of a big quantity of cards having been already punched, to separately carry out the conversion operation. With this in view a special conversion program was written in the following way:

---

* This section refers only to the actual processing of the population census which was completed before writing this text.

— The conversion of the cards and obtaining a multireel archive (with two copies) which are an exact copy of the cards. In order to substitute, when necessary, several reels of the first copy by reels of the second copy, the program recorded a fixed number of records, and thus a strict correspondence between the two copies of the archives was established;

— The counting of the cards and the printing of the total number of persons for the smallest geographical level in order to compare the results with the control figures.

In order to avoid the restarting of the whole conversion operation in cases of technical break-downs or other types of errors every half hour a dump was done on magnetic tapes.

The conversion operation was always processed by a multiprogram system and, in most cases, at the same time with the logical control at a speed of 550 cards per minute.

The logical control and the whole processing was carried out separately for each region in about 2 minutes for 1,000 inhabitants. The flow-chart of the logical control and of the automatic correction is shown in the Appendix.

For carrying out the logical control, the following parts of the electronic data equipment were used:

Internal storage	23,344 words
Magnetic tape decks	3 units
Line-printers	2 units
Card readers	2 units

At the end of the processing of each region a centralizing table containing general features of the process as number of cards processed, number of cards rejected, number of corrected cards, number of out of sequence cards, etc., was printed. The average figures are as follows:

Total number of cards processed	100%
Cards	0.1%
Cards out of sequence	2.8%
Automatically corrected cards	1.5%

In certain cases at this stage the possibilities of the logical control program were used in order to list the cards which had been out of sequence.

As the run of the logical control program for the population of a region lasted tens of hours, a core store dump was taken periodically.

In this way it was possible to restart the program whenever it was necessary, after a technical breakdown of the equipment or the magnetic tape or after some operator's or programming errors.

It is necessary to point out that the carrying out by multiprograms, the conversion and logical control programs was one of the most efficient cases of utilizing the computer in multiprogramming.

Obtaining the control totals

The control totals, which had been obtained from the data archives before the processing of the tables, were submitted to a control group. This group either accepted the figures or pointed out the discrepancies which appeared. In such cases certain programs which had been specially prepared in order to find out and correct errors were used. After this operation the program for obtaining the control totals was continued.

Examples of using special utility programs for the correction of the final archive are the IZOL and COAR programs shown in the Appendix. The carrying out of the logical control programs and the obtaining of the control totals including the special utility programs amounted to about 30% of the processing time of the population census data, i.e. about 25% for the logical control and 5% for the control totals.

The cases of restarting the control totals (approximately 15%) were due to:

— double entries of individuals on the tape;

— wrong coding of certain codes indicators;

— alteration of certain indicators in the store.

Obtaining and printing the tables

The tabulation of the tables was carried out simultaneously for a group of tables. This required a scrupulously careful organization of the handling of the very big number of magnetic tapes.

For example, for the tabulation of the simple linear tables by the VENL program (as shown in the Appendix), a number of 11 reels of magnetic tapes were obtained, corresponding to the 11 linear tables *.

The tabulation of 1,000 persons in a table required about 2 seconds.

The time consumed with re-runs of the tabulation programs due to system faults was of about 15% of the whole processing time.

The magnetic tapes at the output and of the centralization program are a very important archive because they contain the final data of a large number of tables, obtained after spending very much computer time. Therefore they had to be handled and kept with utmost attention.

Among the security and utilization measures of these archives of tables, the following could be mentioned: copying a file immediately after its creation and the creation of two copies of these archives, the wiping off of input tapes only after a careful examination of the output ones, etc.

---

* In order to simplify the processing scheme as shown in the Appendix, the names of the programs are conventional and in a few cases they belong to a group of programs whose object is similar: VENL in the Appendix, corresponds to the NATA, NATB and NATI programs, each of them tabulating a group of tables.

The carrying out of the totalizing and centralizing amounted to about 10% of the time needed for the execution of tabulation programs.

The printing of the two pages of the tables which are published as an "open book" was carried out at the beginning, using the two printers of the system, but later the possibility of first printing the first part and later the second part of the table was used.

The tables were multiplied by the "xerographic method" and this was done at low speed by the printer, in order to obtain a good quality of the printing (680 lines per minute).

Some tables had to be reprinted owing to poor printing due to the deterioration of the paper by the printer. This amounted to 10% of the printing time.

In most cases the printing of the tables was carried out by the respective program in multiprogramming.

4. CONCLUSIONS AND FINAL OBSERVATIONS

The use of electronic equipment requires a change in methodological orientation of the census conceptions and their statistical processing.

The 1966 population and housing census has been processed in good conditions by electronic equipment for first time in our country. The efficiency has been superior to the classical processing systems.

Although most of the problems regarding the lay-out of the printed forms for electronic data programming and processing were entirely new for our statisticians, this did not influence quality of the processing nor the time schedule for the publication of the final results of the census.

The volume of programming and processing was as follows:

— Total number of written instructions	150,000
— Number of programs	100
— Number of compilations/program	10, ..., 15
— Average number of errors corrected/compilation	2, ..., 3
— Average number of errors/program	30

Computer time used for:

— testing (first 1.5 months)	14 hours
— testing (last month)	7 hours

Several conclusions obtained from the experience earned by the electronic processing of the population census can be applied to other large statistical works.

In the case of electronic processing the preparatory operations (coding, code verification, punching, punch checking) amount to more than 99% of the time spent for the whole processing.

Thus for one hour's work of the computer 100, hours-man were necessary for punching and checking.

This means that the possibilities of reducing the amount of work must be looked for mainly in these preparatory operations. It is necessary to replace the classical methods of preparing the data for automatic processing by new methods based on special printed forms.

These printed forms, as a technical information carrier, should enable the direct introduction of the data into the computer and should ensure the automation of the operation.

It is necessary that programs with long execution term and especially those which operate with files formed by several reels should use the DUMP and RESTART procedures in order to avoid possible big restarts due to certain system or operator's errors.

In order to ensure a continuous flow of processing it is advisable to lay out specific utility programs. Such programs are, for instance, those for searching certain records on the magnetic tapes, isolation and record correction programs, etc.

One of the characteristics of the present census is its statistic conception. However, the electronic data processing systems allow another conception which may be characterized as "dynamic". In the electronic processing flow the initial information is filed in archives printed on magnetic tapes. If a corresponding system of programs has been laid out, any table can be obtained in a few hours. Under such circumstances it would be necessary for the initial stage to lay out only a minimum number of tables containing general features, while highly specialized aspects would be presented only on request.

Thus, the potential capability of the computer to produce tables very rapidly allows for the conception of small tables liable to be easily consulted.

It is necessary to also mention the capability of the computer to work out regressions, correlations, etc., starting from the tables or the data in the archives, which will simplify considerably the interpretation of statistical results.

This dynamic conception evidently goes hand in hand with the idea of a permanent population file being continuously updated.

In order to carry out these objectives it is necessary to prepare systems of programs which allow the working out of tables at short notice. In view of the fact that for the working out of a table by the computer very much time is spent on programming, it is necessary to automatize programming.

It has already been mentioned that such an unsuccesful attempt was made by using a program written in PLAN and FORTRAN.

The system of programs necessary for this purpose is a compilative system which uses a correspondingly specialized language.

Furthermore, it is necessary to create both a language for the specification of the conditions of logical control in an archive and a general system

of programs for the carrying out of the logical control. With this purpose in view, our analysis and programming department has started preparatory work for the construction of a system of programs for processing statistical tables.

Owing to the changes which have taken place in the administrative structure of our country, it has become necessary to do over the whole of the processing again. This has been carried out by using the data existing in the basic archive. Owing to the fact that the tables requested in the new variant were different from the initial ones, other more efficient tabulation programs have been written. The speed of the new programs has been increased by 100% as compared to that of the initial ones.

TRAITEMENT ELECTRONIQUE DES DONNEES DU RECENSEMENT DE LA POPULATION ET DES LOGEMENTS EN ROUMANIE

Le recensement de la population et des logements du 15 mars 1966 est le premier de cette sorte en Roumanie, traité par des moyens électroniques.

Dans cet ouvrage on présente les principaux aspects du travail de préparation et d'exécution du traitement électronique des résultats de ce recensement, l'expérience acquise en ce qui concerne l'utilisation de l'équipement électronique.

C'est de ce point de vue qu'on étudie les travaux destinés au traitement électronique des données: l'établissement du programme de traitement, l'élaboration des formulaires d'enregistrement, la transposition de l'information sur des porteurs techniques d'information; on analyse ensuite les principaux aspects du traitement, à savoir: les caractéristiques d'exploitation des programmes, la manière dont on a planifié les travaux, le temps affecté à chaque opération, etc.

ЭЛЕКТРОННАЯ ОБРАБОТКА ДАННЫХ ПЕРЕПИСИ НАСЕЛЕНИЯ И ЖИЛИЩ В РУМЫНИИ

Перепись населения и жилищ, проведенная 15 марта 1966 г. в Румынии, является первой переписью такого рода, обработанной при помощи ЭВМ.

В данной работе представлены основные вопросы подготовки и проведения электронной обработки итогов данной переписи а также накопленный при этом опыт использования ЭВМ.

В этом отношении, рассматриваются подготовительные работы для обработки на ЭВМ — составление разработанной программы, разработка формуляров регистрации, запись данных на технические носители —, после чего проанализируются основные вопросы самой обработки, а именно: эксплуатационные характеристики программ, порядок планирования работ, использованное время на отдельные операции и т.д.

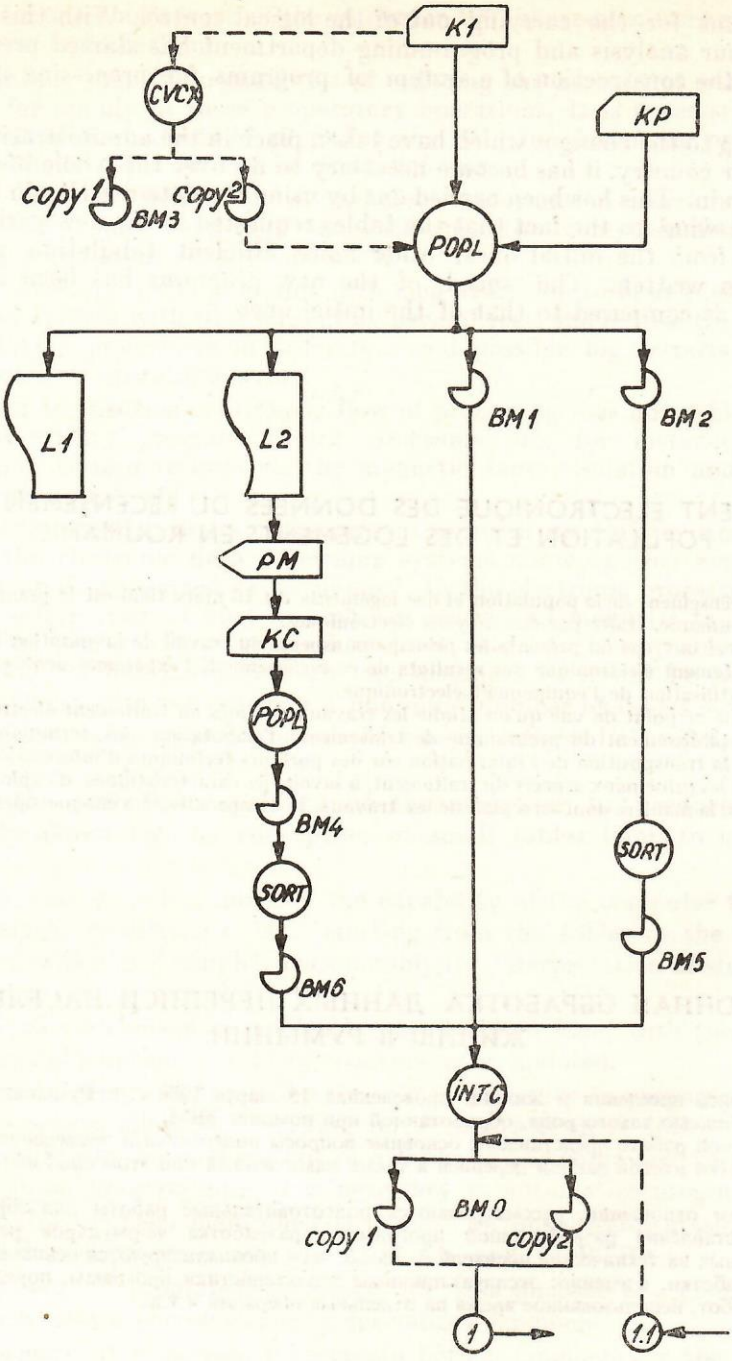

K1
CVCX
KP
copy 1
BM3
copy2
POPL
L1
L2
BM1
BM2
PM
KC
POPL
BM4
SORT
BM6
SORT
BM5
INTC
BM0
copy 1
copy2
1
1.1

Fig. 1

Appendix

DIAGRAM OF THE POPULATION CENSUS DATA PROCESSING

At regional level

Logical checking

K1 — Data file of population census on punched cards
KP — Parameter card comprising the list of localities in a region
POPL — Program for logical checking of the population census data
CVCX — Program for recording the K1 file on magnetic tapes (Example 1 and example 2 — two identical files)
BM3 — Population census data file on magnetic tapes
L1 — The list of records outside the sequence
L2 — The list of the records rejected at the logical checking, the list of records with more than 4 corrections and the centralizing list of the records distributed by sorts of corrections
BM1 File of correct data existing in the sequence
BM2 — File of correct data, outside the sequence
PM — Correction of the rejected records and their recording on punched cards
KC — File on punched cards, with corrected records
BM4 — File KC recorded on magnetic tapes
SORT — Sorting program
BM5 — File BM2 sorted
BM6 — File BM4 sorted
INTC — Interclasifying program
BMO — The basic file in the sequence, corrected after the logical checking (Example 1 and example 2 — two identical files)

NOTE. The dotted lines illustrate the possible variants of the processing flow.

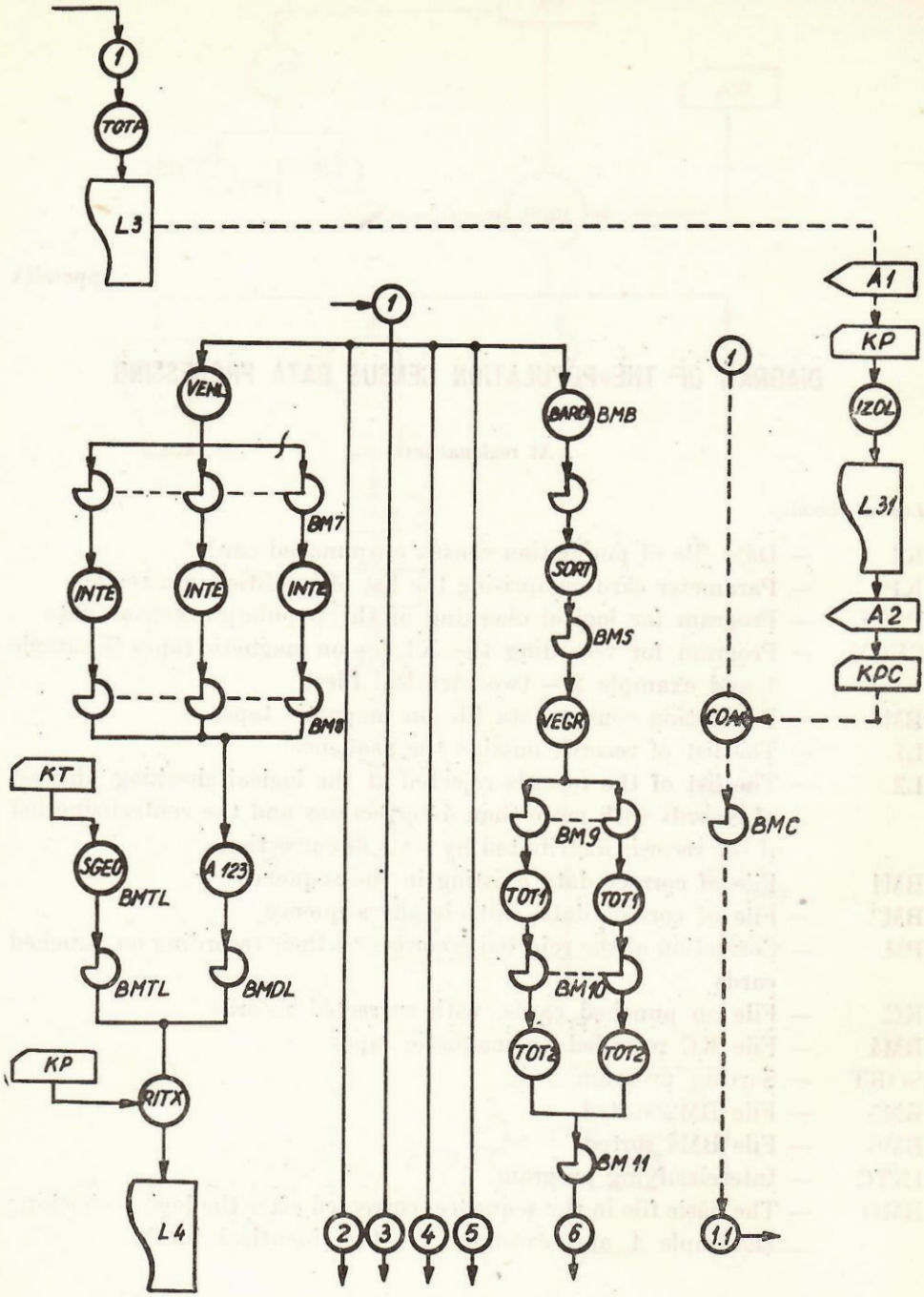

1
TOTA
L3
A1
KP
IZOL
L31
A2
KPC
1
VENL
BARC
BMB
BM7
INTE
INTE
INTE
SORT
BMS
BM8
VEGR
COAX
KT
BM9
BMC
SGEO
A 123
BMTL
TOT1
TOT1
BMTL
BMDL
BM10
KP
RITX
TOT2
TOT2
BM 11
L4
2
3
4
5
6
1.1

Fig. 2

Data tabulation for obtaining and printing the linear-simple tables from the reduced archives

TOTP — Program for obtaining the checking totals tables
L3 — Table of checking totals at regional level
VENL — Tabulation program for the obtaining of linear table data, corresponding to the territorial units of a region
BM7 — Magnetic tapes containing a linear table
INTE — Program for arranging the table data in the requested order
BM8 — Magnetic tapes containing a linear table
A123 — Program for centralizing the linear tables on a single magnetic tape
BMDL — The output file of the program A123
KT — Punched cards with the table text
SGEO — Text recording program from the linear tables on magnetic tape
BMTL — Output text file of the SGEO program
RITX — Linear tables printing program
L4 — Linear tables at regional level
BARD — Program for choosing the active persons from the basic file
BMB — Data file obtained at the BARD output
BMS — BMB file — sorted
VEGR — Data tabulation programs for obtaining big-sized tables, resulted from the BMS reduced file
BM3 — Magnetic tapes containing data that correspond to the indicators values, without TOTAL
TOT1 — Totalizing program according to the values of an indicator (tables that surpass the memory capacity)
BM10 — Data files obtained at the output from TOT1 program
TOT2 — Final totalizing program and arranging of the table sections in the recommended order, recorded on the BM11 magnetic tape
A1 — Analysis of the checking totals and establishing of the records that are to be extracted when acceptable
IZOL — Program for extracting the records established by the analysis of the checking totals
L31 — List of records extracted by IZOL program
A2 — Analysis of the extracted records and establishing of the cases when the basic file must be corrected
KPC — Parameter cards for correcting the basic file
COAR — Program for correcting the basic file
BMC — Basic corrected file

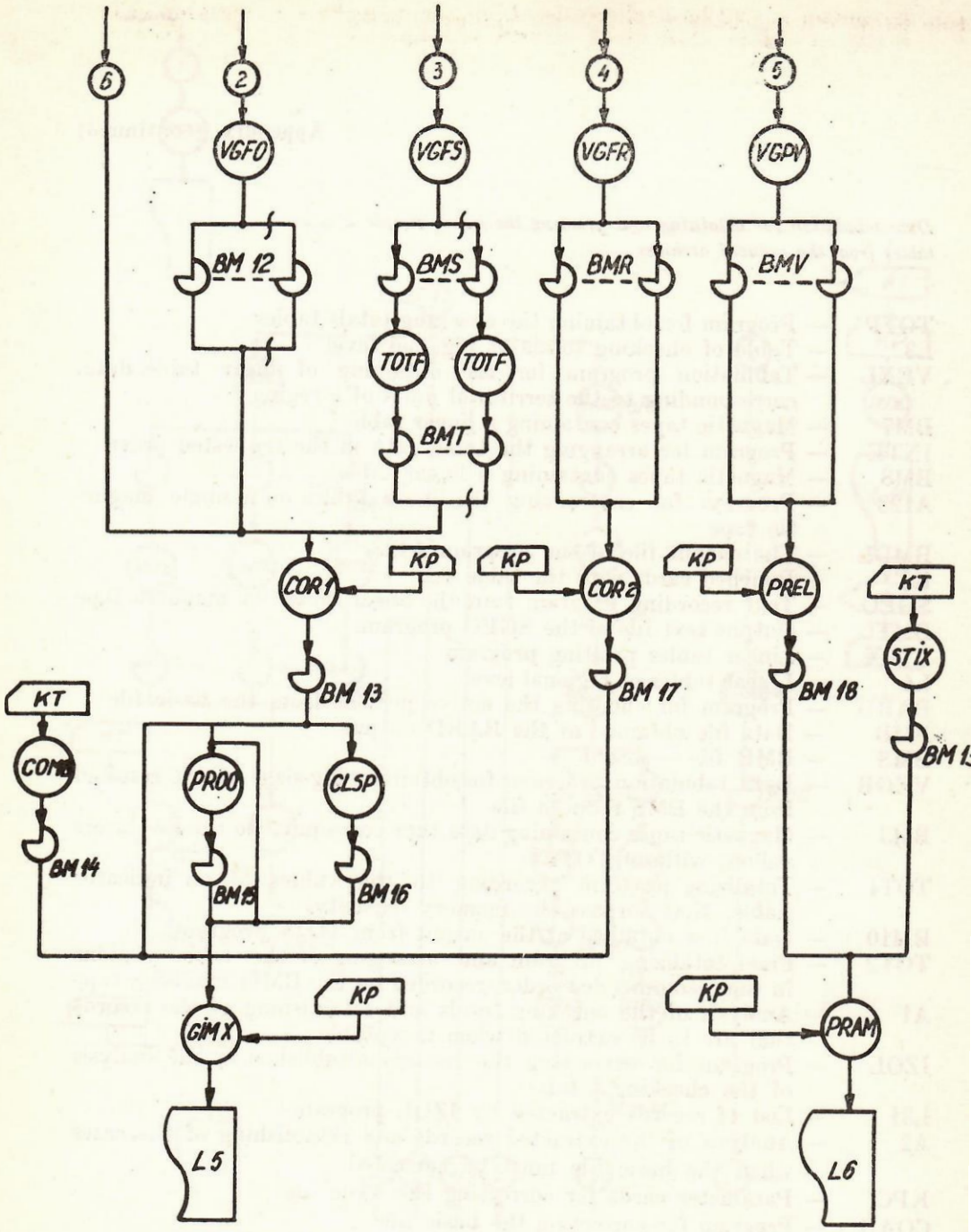

6
2
3
4
5
VGFO
VGFS
VGFR
VGPV
BM 12
BMS
BMR
BMV
TOTF
TOTF
BMT
KP
KP
KP
COR1
COR2
PREL
KT
STIX
BM 13
BM 17
BM 18
KT
COM
PROD
CLSP
BM 14
BM 15
BM 16
KP
KP
GIMX
PRAM
L5
L6

Fig. 3

Appendix (continued)

Data tabulation for obtaining and printing general correlative tables

VGFO	— Tabulation programs for tables with fixed body, containing correlations between pattern characteristics and territorial characteristics — general tables
BM12	— Magnetic tapes containing data of each general table
COR1 and COR2	— Programs for concentrating several tables on one magnetic tape — BM13 and BM17, respectively
COMB	— Programs for recording the texts for fixed body tables on the BM14 magnetic tape
PROO	— Program for percentage computation from the absolute data of a table and for recording a percentage table on the BM15 magnetic tape
CLSP	— Program for computing some additional indicators derived from the absolute data of a table and the printing of the obtained table on the BM16 magnetic tape
GIMX	Printing program for tables with fixed body
L5	— Correlative tables with fixed body, at regional level
VGFS	— Data tabulation programs for obtaining the fixed tables, containing information only on the population structure
BMS	— Magnetic tape containing tables in an intermediate stage
TOTF	— Totalizing program of a table data
BMT	— Files obtained at the output from TOTF program
VG	— Data tabulation programs for obtaining tables with fixed body, at district level
BMR	— Data files obtained at the output from VGFR programs
VGPV	— Data tabulation programs for obtaining correlative tables with variable body
BMV	— Data files obtained at the output from VGPV program
PREL	— Program for concentrating several tables on one magnetic tape
BM18	— Data file obtained at the output from PREL program
STIX	— Program for recording of the texts for the variable body tables on BM19 magnetic tape
PRAM	— Printing program for the correlative tables with variable body at regional level — L6

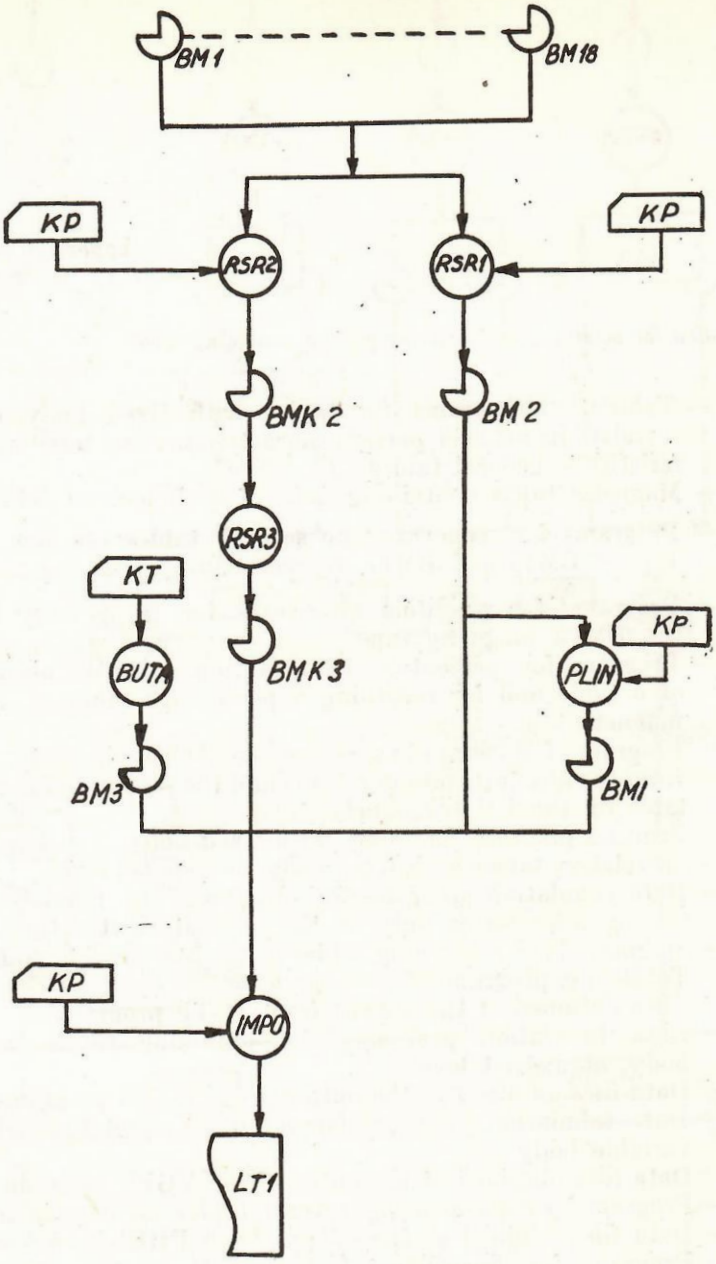

BM1
BM18
KP
KP
RSR2
RSR1
BMK 2
BM 2
RSR3
KT
KP
BUTA
BMK 3
PLIN
BM3
BMI
KP
IMPO
LT1

Fig. 4

DIAGRAM OF PROCESSING THE POPULATION CENSUS DATA AT NATIONAL LEVEL

Linear simple tables

BM1 ... BM18	— Files of linear tables data for each region
RSR1	— Program for centralizing the type I tables
RSR2	— Program for intermediary centralizing of data for type K tables
BMK2	— File of data obtained at the RSR2 program output
RSR3	— Program for final centralization of data for type F tables
BMK3	— File of data obtained at the RSR3 program output
BMI2	— Data file obtained at the RSR1 program output
PLIN	— Program for computing the percentages from the absolute data of a table and for printing the table obtained on the BMJ magnetic tape
BUTA	— Program for recording the texts of the simple linear tables on magnetic tape
BM3	— File of texts for printing the linear tables at national level
IMPO	— Printing program for the linear tables
LT.1	— Linear tables of type I, J, K

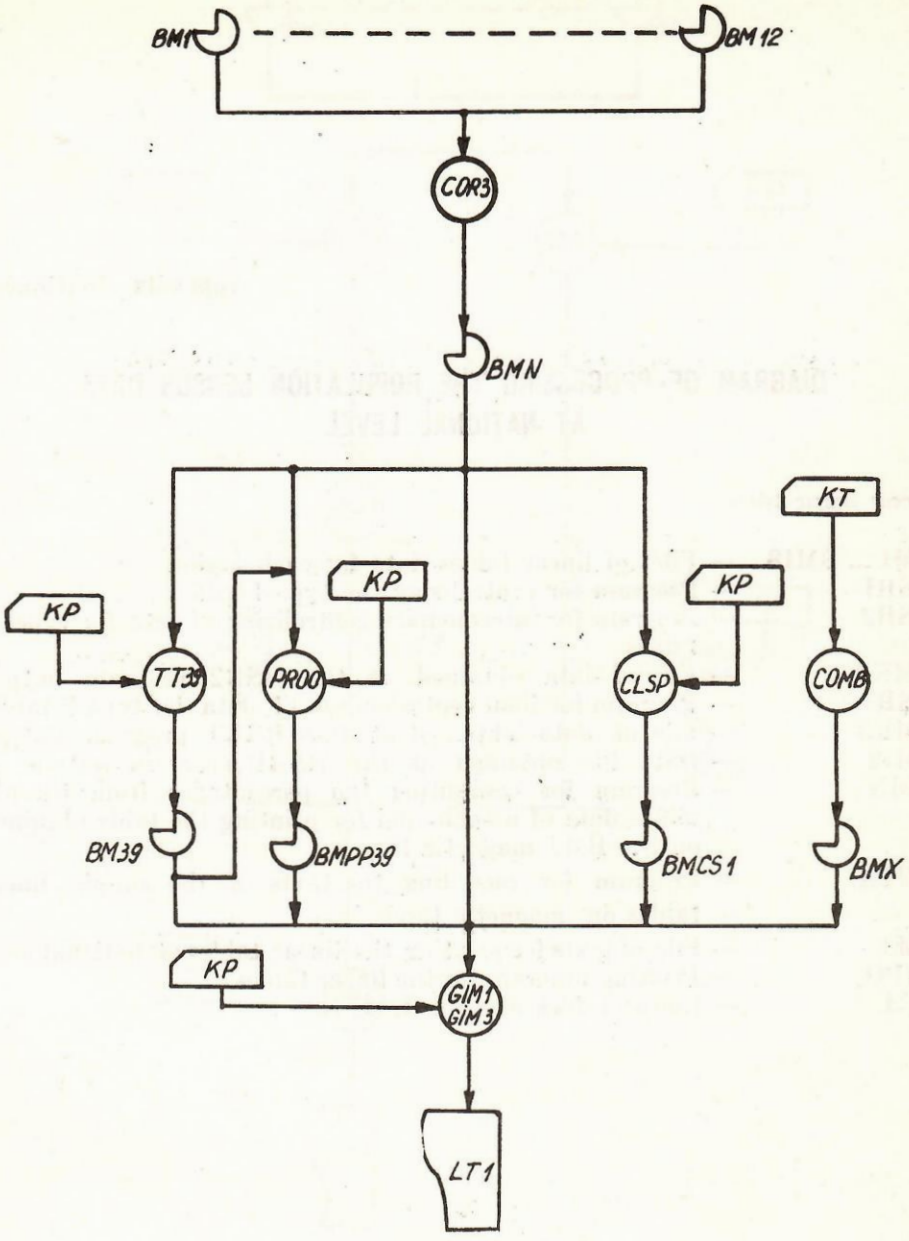

BM1
BM12
COR3
BMN
KT
KP
KP
KP
TT39
PROD
CLSP
COMB
BM39
BMPP39
BMCS1
BMX
KP
GIM1
GIM3
LT1

Fig. 5

General correlative tables

BM1 ... BM	— Data files representing the tables at regional level
COR3	— Program for obtaining the tape of fixed pattern tables at national level
BMN	— File of data obtained at the COR3 program output
TT39	— Program for centralizing the tables at national level
BM39	— Data file obtained at the TT39 program output
PROO	— Program for computing the percentages from the absolute data of a table and for printing the obtained table on the BM39 magnetic tape
CLSP	— Program for computing additional indicators, derived from the the absolute data of a table
BM CS1	— Data file obtained at the CLSP program output
COMB	— Text file of the tables at national level
BMX	— Magnetic tape of texts at national level
GIM1 and GIM3	— Programs for tables printing
LT.1	— General tables at national level