Dr. eng. V. PESCARU
Center of Informatics
and Cybernetics for Planning.

Information Structuring for the Automatic Processing Systems

The article presents author's opinion regarding the main requirements of the information structuring with the view to processing the information by means of electronic computers. At the same time, the author gives a short account of the factors that hindered the evolution of the information structuring methods and techniques.

In the second part, it is pointed out the qualitative leap scored by the electronic processing systems as an outcome of changing the traditional files for the integrated files and data banks.

1. INFORMATION — THE OBJECT OF PROCESSING

Before dealing with some aspects concerning the methods and techniques meant to stucture the information which is to be automatically processed, we are of the opinion it should be timely to point out some principles that have to underlie the structuring procedures so that they get accomplished in keeping with the goals as to the improvement of the information systems.

The general goal aimed at by the information systems, at the same time taking in the goal of those systems based on the electronic techniques of data processing and teleprocessing, is that of promoting the increase of efficiency in the basic activities (production, trade, transport, etc), while the particular goal of the system is aimed at providing the necessary, sufficient

and qualitatively adequate amount of information for the decision and/or informing levels both inside the given system and the hierarchically superior information systems to which it belongs.

Hence, any information structuring has to adequately meet the demand of accomplishing those goals. The techniques and methods of information structuring when using electronic processing systems are not to be treated as an aim in itself.

One of the principles one has to take into consideration in the information structuring process is that the specific structuring methods and techniques are meant for the information, for its preparing and processing with a view to turning it to the utmost account.

The quality of these methods and techniques becomes acknowledged only by analysing the extent to which they have met and are still meeting the requirements of efficient processing and revaluation, after the direct contact with the information. Information stands for the object of processing, and, at the same time, it represents the main cathegory among the resources employed in the information and informatic systems, along with other cathegories of resources: specialized staff, processing equipments and tools for operating on these equipments (analysis and design methods and techniques, programming and inquiry languages, application standard software packages, operating systems, methods and techniques meant to organize and process information, etc.)

The other three cathegories of resources, except information, are exclusively used to design and/or elaborate the information generating, recording, transmission and processing systems with a view to turning information to good account.

At the same time, one may draw the conclusion that information structuring has to cover the entire information flows, starting with its design and generating and ending with its turning and re-turning to good account.

Tackling the information structuring so that it should meet only the technical requirements determined by electronic data processing and teleprocessing is but an incomplete solution and, to a great extent, it is severed from the practical aspects of revaluation, generation and preparation of the information with a view to its being processed. This type of approach moving away from the concept of systemic structuring of information and trying to organize information only by one of the flow sections by means of which it is conveyed, processed and turned to good account, inevitably causes much troubles in implementing and operating the specific processing systems.

Files and data banks as structuring types of electronic processing, in most cases, are neither built nor updated by starting with the information flow at the entrance of the electronic computing centre. They are built and updated beginning with the moment information is generated; in its turn information generation is directly conditioned by the informing (revaluation) requirements.

Another principle that has to be taken into consideration during the process of information structuring is that by which information put in, generated and processed by means of the given information system and revaluated by it or by other external systems, is characterized by a significant frequency of adjustment (it refers to a good part of the whole system). It

is the volume and structure of the information put in and/or generated within the given system and the frequency of these information inputs and/or generating that determines, particularly, the main technical and operating characteristics of the data processing systems and, generally, the characteristics of the specific information systems.

Information is the most dynamic element of those elements belonging to the information systems. It is able to render life to those systems. Hence, the need to select and use these methods and techniques of information structuring in full concordance with the dynamic character of the information.

So, several fundamental requirements are brought forth; they have to be taken into consideration when one is structuring information within the contemporary information systems and those of electronic data processing, implicitely. Among these requirements we mention the following ones:

— when structuring the information one has to start from the concrete present and prospective needs as to informing the decision and / or executive levels (one has to start by more accurately defining the outputs);

— the information input of the specific information system and/ or the information which is being generated (primary information) in order that it may be turned to account has to be determined from structure and volume point of view in keeping with the information input requirements and the possibilities the electronic processing systems can offer in this respect;

— providing information with a good quality during its conveyance flow (starting with its generating point and ending with its turning to good account) as well as a minimum cost of the information output along with *the minimum response time of the information system, represents a priority* task of the process of information structuring;

— fact being given that the whole amount of information generated and / or put in, processed and/ or turned to account within the information system is not entirely covered by the system of electronic processing and taking into consideration the demands as to continuously extend the latter systems, it is necessary that information structuring should ensure a compatible and functional interface among manual processing, mecanographic and electronic systems as well as an easier transition to the electronic processing.

2. THE DEVELOPMENT OF METHODS AND TECHNIQUES AS TO THE INFORMATION STRUCTURING

Information structuring methods and techniques have developed within the process of improving the information systems and the process of improving the structure of these systems.

The improvement of the information systems as tools necessary to manage and implement the economic and social activity is a continous process, objectively determined, since the development of the society is an objective process, too.

The evolution of the information structuring methods and techniques was directly influenced by a series of factors, among which we mention the following ones;

— the ceaseless increase in the complexity of the social and economic activity which determined an increase in the volume of information that was generated, conveyed, processed and revaluated;

— the growth rate of the social development which determined the way time resource is used would become an ever more conspicuous factor for this development; the size of the response time of the information systems has also become one of the main parameters for estimating the degree of structuring as well as the degree of efficiency these systems acquire;

— the influence scientific and technological development exercises over the evolution of the ways and methods to find, record, convey, process, store and retrieve the information.

Against a background covering a great variety of information structuring methods and techniques that are to be traced in the history of social development, the main structuring forms are: files and data banks.

The datafile is a cluster of information, homogeneous from the point of view of their meaning (determined mainly by revaluation requirements) and processing requirements.

Information structuring by files could be traced in the historical past, far beyond the limits of our century; by then, the physical recorders were stone, metal and wooden plates, walls and plates made up of various binding material, papyrus, etc.

Among the first stages regarding the development of the information structuring methods and techniques under the conditions of employing the electronic computers, one is mainly characterized by adapting the types of the information structuring that were already existing within the systems of manual and mechanographical processing to the technological requirements set by the electronic computers. The main type of information structuring used during this stage as to the electronic processing system is the file.

People who identify, conceptually, the origine of the file with the beginning of the electronic data processing are making an error. Generally, a file as well as the main structuring types known so far, were built and used within the systems of manual data processing. The mechanographic and electronic data processing have taken over this structuring manner from the manual system without bringing about fundamental improvements. The files that are sequencially, indexed-sequencially and randomly *) organized are to be found in the manual processing and/ or searching systems structuring, too (see the files that are recorded on paper cards characteristics for administrating the materials/and raw materials, for libraries, for polyclinics, etc.) The selection of one of the types as to files structuring is mainly performed according to the amount of the data recorded, the requirements as to search and response time, the complexity of processing and / or searching, the number of selection and / or processing features.

There are frequent cases when, in the systems of electronic processing, the files recorded on the paper cards of the type that has been already men-

*) The author considers the notion of "file by direct structuring as being improper since a direct access to the information belonging to a file of this type is not possible to be attained; this is valid for the electronic processing systems, too. As a matter of fact, the "direct" access is performed by an index (key) which, logically speaking, is not a direct one. In this paper, the author uses instead of the term direct access, that of random access.

tioned ("traditional files") are re-written on punched cards (other paper cards) and then they are recorded on magnetic media (tapes and disks).

These two "re-writting" procedures are performed in order to reduce the processing and response time of the information system in question, by the high speed ensured by employing an electronic computer. In these cases it is quite natural that one should estimate the time and cost required by the "re-writtings" (the time for finding out and correcting the errors generated by the manual re-writting of the data from "traditional" to punched cards), as well as the time and cost required by information search and / or processing under the different alternatives of the processing system. Only after the comparative analysis of the estimations regarding the different alternatives of the foreseen processing has been performed, the most efficient solution can be selected.

In this stage, significant evolution was accomplished in the field of the electronic processing systems by the relevant decrease in the volume of physical devices used to record the information that existed on files as well as a diminishing in their processing time, but the field of file structuring methods.

During this stage sequencial files are mainly used; usually, the physical structure overlaps the logical one; batch processing is exclusively used; application programmers design the physical structure of the recorded data; the changes occuring in the data structure or storing techniques brought about some alterations inside the programme which finally led to their recompiling. During this stage the data structure was usually designed in order to serve only one application; it can be noticed that there existed a great deal of redundance in the data of the files that had been designed to serve an application cathegory.

For a long period of time the philosophy and practice of the information structuring process within the systems of electronic data processing has not managed to keep pace with the possibilities offered by the electronic computers; this is one of the main factors which affected to some extent the efficiency of applying these means to the economic and social information systems which are characterized by a large volume of input and output information within the processing systems.

Among the factors which have to some extent obstructed the faster development of the information structuring methods that could take place in keeping with the characteristics offered by electronic computers, there are to be mentioned the following ones:

— the electronic processing systems have been and are still used within some information systems conceived and structured mainly in order to meet the requirements of the manual and mechanographical processing. It is to be noted that for many years a more stronger phenomenon has become relevant as to adapting the electronic computer utilization to the information systems conditions, while the process of re-designing and re-structuring of these systems in keeping with the possibilities offered by the up-to-date technique of electronic data processing and tele-processing was less conspicuous.

— during the first stage of electronic computer utilization, the overwhelming preponderence was represented by performing some technical and scientific calculations which did not raise special problems as to information structuring.

— the output of computers has grown at a very rapid rate and exceeded to a great extent the rate of re-designing, restructuring and adapting of the information systems to the new conditions. It can be stated that the information boom that has occurred during the last decades determined that the electronic computers be used, while the philosophy and practice as to improving the information systems were unsufficiently prepared to cope with the requirements of the new stage.

Nor the activity of elaboration and usage of the application standard software packages, standard software systems meant to organize and process the information has developed in accordance with the demands and the rate of improving the electronic computers, from construction point of view.

— The training system of the personnel so that they should be able to work with an electronic system have proved to be unsatisfactory from the point of view of acquiring the knowledge of the concrete demands and methods as to the economic and social information systems, the continuous raising of their efficiency as well as the ways meant to effectively integrate the electronic computers in the information-decision activity.

A significant qualitative leap was scored in developing the methods and techniques of information structuring within the systems of electronic data processing by elaborating and commissioning the algorithms and standard programmes systems concerning file linking, both as their structuring and using by means of the electronic processing systems.

The demand for file integration was an outcome of the continuous, accelerated increase in the complexity of the information systems which is mainly illustrated by the sensible growth in the number of relationship, the correlation ones included, among the information (information group) belonging to various activities or applications involved in an information system or more than one information system.

At the same time, the urgent need to diminish the influence exercised on the application programmes by the various changes performed in the equipment, led to the further development of the information structuring methods and techniques; thus, it was possible to foreshadow the characteristics of the next stage, its beginning being marked by the period that comes immediately after 1960.

Some advanced elements acquired by the information structuring methods and techniques with in the electronic processing systems were implemented by the end of the first stage; they made it possible that the qualitative leap occur more easily during the second stage. Thus, during the transition years between the two development stages, the interest in severing the logical level from the physical, data structuring one became intensified, while the tendency was maintained as to design the data so that they should mainly cope with only one type of application.

During this stage the files indexed-sequencially and randomly structured have been ever more frequently used; in addition, the conversational mode and real time operation system began to be used.

The complex requirements assigned to the up-to-date information systems has given rise to the necessity of considering the desideratum that more users have access to the same data sets, since each user requests that the data would be organized so that they would meet the specific requirements of processing.

The new situation being created, the employment of the traditional methods and techniques of information structuring would have led to the continuous growth of data redundances, since the electronic processing system requires the representation of the same data by two or more files in order to optimally assist various application programmes. This solution would have implied a decrease in the efficiency of the informatic systems illustrated by an increase of the period of time necessary to design and program the applications as well as by an increase of the time and technical storage capacity that are necessary to structure and record the data.

Under these conditions, the need to control and minimize the redundances occurring when data are presented became apparent and this is called the need of integrating the data file.

So, there emerge the information structuring techniques which allow data representation as well as the representation of the inner relationship among the data; their final aim is the utilization of the integrated files.

The systems BOMP (Bill of Material Processor) and SCF (Systeme de Chainage de Fichiers) may be looked upon as belonging to the techniques type mentioned above; SCF is characteristic for the Romanian computer FELIX C-256.

The emergence and development of the data files integration techniques marked a significant qualitative leap towards implementing some efficient informatic systems that yield immediate results, especially in the field of output management and planning.

Using the integrated files did not manage to solve a very important problem, namely that of making application programmes free of data structuring mode. This requirement becomes more and more stringent with the increase of the complexity acquired by the informatic systems and, consequently, with the need to organize new types of files, to alter the structure of the existing file recording without changing the application programmes that are being used.

The solution for accomplishing this desideratum was to include the characteristics of the files as well as the relationship among them in the application programmes.

During the period immediately after 1965 it is to be noticed that the producers of application software made their utmost effort to accomplish application programmes free of data structure.

One of the first solutions that was approached and implemented was that of accomplishing an interface between the files structured by BOMP system and the application programmes; the primary task of this interface was to make the application programmes free of data structuring mode. By this way, the alterations performed in the data structuring led to changes in the interface, without affecting the application programmes that were being used.

During the next stage, the systems of specialized programmes is developing; they are meant to ensure the management of the data required by several users, so that application programmes will not depend on the way data are structured and the way their management is performed.

So, the first data banks are elaborated; they marked a qualitatively new stage as to data structuring within the electronic processing systems — the third stage. By sinthesizing several definitions that have been given to the data banks in the specialized literature, we think that the following definition is worth mentionning:

The data bank represents a system meant to structure, process and teleprocess the information; it is made up of the following three elements:

1) a set of interrelated data;

2) the description of the data and the relationship among them;

3) a system of programmes for data management which can ensure that the programmes be free of the data structuring mode; also it ensures a minimum redundance which is controlled during data storing as well as a minimum response time as to user's requirements.

The concept of *data base* is used to designate, by group, the first two elements of the data bank. Hence, the data base is used but as a component element of a data bank, since it does not stand for a distinctive stage in the development of the data structuring techniques. Otherwise, trying to consider the data base as a stand alone element, out of the need to use the data existing within the data base, one can obtain the specific use of the integrated files (without making a distinction when describing the data, etc) or the construction of a data management system having the characteristics given above; then, the data bank is accomplished.

In case there are several data banks, the data set is devided into distinctive data subsets by their information content (many times out of the need to accomplish the data bank by stages) to which the corresponding descriptions are included. Hence in these cases we get a set of bases we call *the system of data bases*; the management system of these data bases acquires an additional valence, namely that of managing the relationship existing among the information belonging to different data bases.

The third component part of a data bank, i.e. the system of programmes with specific tasks, is called the data bank management system (DBMS). Hence, the DBMS ensures the management of the data base or the system of data bases covered by the data bank.

So, it could be also stated that *the data bank is made up of a data base or a system of data bases with a DBMS included.*

When considering the DBMS, there are to be noticed two different stages from the point of view of their quality. In the first stage the user has to know the description of the entire set of data which are belonging to the base and sometimes (the case of TOTAL case) he has to use the name of the description fields in order to perform the logical recordings specific to the applications.

In the second stage, the main preoccupation was how to dispense the user with the task of acquiring as much as possible a knowledge about the entire structure of the data base, having in view the fact that these structures

have recently become extremely complex. So, an independent management as to the general structure of the data base is accomplished. This structure is called virtual level of the data structuring.

Hence, there appear two levels of independence, namely the logical and the physical.

By logical independence it is meant the possibility of altering the virtual structure without performing changes in the application programmes, and by physical independence it is meant the possibility to alter the physical structure of the data, without performing changes in the virtual structure and, thus, without changing the application programmes.

Typical for this stage is also the emergence of that function as to the data bank management, the emergence of some improved procedures to ensure protection, differentiated access to data and getting responses to unanticipated questions.

The DBMS that were accomplished during this stage usually have a language used to describe the data and the relationship existing among them, a control language which is employed by the application programmes and a language meant to directly inquire the data bank. One of the present tendencies that has become manifest in the evolution of the inquiry languages as to the data banks is the strengthening of the non-procedural character which determines that these languages get nearer the naturel languages, so becoming much more accessible to the user.

Also, the need to implement the informatic systems by regional distribution determines now the emergence of the distributed data banks based on the computers networks.

In conclusion it could be underlined the aspect that, if during the first aproximately twenty years of using electronic computers, the methods and techniques of information structuring did not raise at the level of the technical operation possibilities offerred by these up-to-date processing means, the last period (especially after 1965) was characterized by the continuous and rapid development of the methods and techniques as to the structuring of information with a view to process it. By this way the social and economic information systems and the informatic contemporary informatic ones have become highly efficient.

L'ORGANISATION DE L'INFORMATION POUR LES SYSTEMES DE TRAITEMENT AUTOMATIQUE

Dans l'article on présente les opinions de l'auteur concernant les principales exigences de l'organisation de l'information en vue de son traitement à l'aide des ordinateurs électroniques. On énumère aussi quelques uns des facteurs qui ont constitué pendant une certaine période des freins dans l'évolution des méthodes et des techniques d'organisation de l'information.

Dans la seconde partie de l'article on met en évidence le saut qualitatif enregistré dans les systèmes de traitement électronique par le passage des fichiers classiques aux fichiers intégrés et aux banques de données.

ОРГАНИЗАЦИЯ ИНФОРМАЦИИ ДЛЯ СИСТЕМ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ

В статье освещены мнения автора относительно основных требований организации информации в целях её обработки с помощью электронных вычислителей. Перечислены, в то же время, и несколько факторов, которые тормозили определённое время эволюцию методов и техники организации информации.

Во второй части отмечается качественный скачок, отмеченный в системах электронной обработки путём перехода от классических карточек к интегрированным карточкам и банка данных.