

Data bases—Uncontrollable or uncontrolled?

by CHERYL M. TRAVER

Stanford University
Stanford, California

In the last two years the "data base" concept has received as much management publicity as did "Management by Objectives" in the years just prior. The popularity of and recognition given to the power of computing in recent years, coupled with increased awareness of the role the enlightened manager should assume, have encouraged one installation after another to acquire a data base system in the pursuit of modern techniques for information retrieval. As both hardware and software have become cheaper and more widely available, the decision has become easier to make; thus the apparent need for foresight and planning has declined. Unfortunately, all too many people are now seeing the entire data base technology as an impossible dream and discrediting the concept as an expensive toy devised by the ever-hungry computer vendors. This trend is summarized in a brochure for a current seminar on data bases: "No subject in computing history has had the import of data base, nor disaster stories of equal magnitude."* And yet the "disasters" are not necessary. With proper foresight and planning and continued supervision, a data base system can produce all the results it promises.

As long as the concept was new, the software expensive, and the expertise nonexistent, much planning was done before anyone ventured into the new territory of the data base systems. But as the successes were publicized and the software became more available, many installations began to acquire such systems with little more thought than new compilers and card readers. Is it any wonder then that fiascos occurred and disillusionment set in? It is unfair, however, to attribute these disasters to the data base technology. One must look instead to the people behind those systems, and in particular to management. For, in the new era of data base technology, management cannot blame the system, or the computer, or the people pushing the computer buttons or coding the computer programs. Data base systems die (or become diseased) through lack of management interest, lack of controls and standards, lack of committed resources, and the absence of a cohesive bond which encourages the users of data to define their needs and communicate their desires.

As stated in the abstract, data bases are a lot like chil-

dren. Both carry a great deal of appeal to the casual observer, probably due to the fact that they are usually presented at their best by a devoted set of admirers who extol their virtues and minimize their shortcomings. Vendors out for a sale will, of course, never stress (perhaps not even mention) the limitations of a system. The owners of such systems, on the other hand, are similar to the man who, having nurtured back to health a scraggly mutt that was left on his doorstep, becomes convinced the bastard is at least equal in worth to the pedigreed schnauzer next door. Disaster stories and problem children are presented in third-party terms and are interpreted as happening only to the "other guy".

Furthermore, the owners of many data base systems are a lot like many parents. They are unhappy, disillusioned, and cannot understand how they acquired an uncontrollable and unmanageable offspring. They complain of the exorbitant investments of resources, time, and money, and wonder why they ended up with a rotten apple. They blame the environment, God, and country—everything but themselves.

Fortunately, though, the concept of parenthood has been around long enough for us to realize that model children are not accidents, but rather the product of hard work and concern. In recent years many books, seminars, and classes have appeared which aim to help parents and potential parents acquire the necessary knowledge. On the other hand, typical of a field that is growing very rapidly, the technology for supporting the sophisticated data base was available long before the guidelines on how to use it. Indeed, the early users were often not even aware of the potential problems. Many of such systems were dismal failures or at least disappointments. There were some, of course, that survived and even prospered, usually as the result of a gifted and farsighted individual. The ratio of successes to failures, however, strongly indicates that data bases are a risky business. The publicity of the failures has frightened many and caused others to blame the computer and its software. For the farsighted, however, the successes prove the benefits can be had—if one can learn how to obtain them.

THE DECISION

First of all, it should be recognized that data bases are not for everyone. Compared with the traditional fixed

* *Data Base 1974 Seminars: Fact vs. Fiction*, Performance Development Corporation, 1974 brochure.

input/output tape systems, data base systems are expensive. The very sophistication and complexity which permit the flexibility and versatility of the new type of system also make it difficult to work with. When there are many options, mistakes are more frequent, since there are also more wrong solutions. Mistakes mean money since they require rework—manpower and machine time.

Data base systems usually include online terminal facilities for retrieval and/or maintenance of data. These facilities encourage increased cognizance at the user level. Since the data are directly available, the operating environment of the computer installation is forced to become more polished. Mistakes can no longer be covered by reruns without user knowledge. The support for an online system requires actual dedicated machine time and thereby substantially reduces the wall-clock hours available for batch work. The batch capacity of the CPU is therefore greatly reduced.

Ongoing support

The new sophisticated systems also require ongoing support. They are not like the "canned" payroll packages or budget accounting systems which, although usually requiring an initial conversion of input/output documents, may be executed week after week with minimal intervention. The good data base system is alive. It needs supervision and care in order to be molded and tuned to react better to the needs and desires of its user population.

One could compare the data base system to the rose. Although universally recognized as one of the most beautiful flowers, the rose does not appear in every garden, even though its cost is not prohibitive. The fact that, unlike the tulip or chrysanthemum, the rose bush requires constant care and attention makes many people give up that beauty and settle for other pretty flowers which require less effort.

Data base systems are open-ended—if handled properly they can be expanded to include an ever-widening spectrum of information needs. However, this expandability can be expensive. The tendency is strong to permit such systems to grow too rapidly. User desire and user need may become confused and often the installation which can barely afford to support the user needs finds itself in a costly situation where the desires of some users are being catered to before the needs of others have even been addressed.

These new systems can even be expensive in terms of computer hardware in that the search for more rapid means of data access for ever-growing online demands encourages a proliferation of different hardware peripherals.

Management commitment

However, the greatest expense of a successful data base system is management time. To be successful, a data base system requires standardization and compromises at the user level. In many environments, especially university administrative areas, time has permitted various departments to become semiautonomous. This characteristic is

usually very evident in the automated systems of the installation where each application is an independent system with its own editing rules, formats, and procedures. In a data base system data must conform across areas in order to be able to form a base for information retrieval.

This need for conformity or standardization is similar to the process of American urbanization.* In colonial days the pioneers were self-sufficient and independent; they had no need for standardization. The shift toward urbanization and community living forced a certain amount of standardization: people had to learn to cooperate in order to live together in harmony. Thus was born a government of checks and balances. Likewise, the desire to make common use of data necessitates an acceptance of a form of "government" and this government needs definition of, commitment to, and enforcement of "laws". Although the rules can be (and all too often are) formulated and enforced strictly by the computing organization, our society seems adapted to a more democratic form of rule. Hence, the management of all areas contemplating use of data base information must become involved.

Since the needs and desires emanate from the users, those are the people who should be determining the nature of their data and of the system to transform those data into information. This commitment of management time is heaviest in the early stages and will be difficult because no results will be immediately apparent. Many of the discussions may even seem absurd. For example, agreement on a standard format for a computer-stored name may seem trivial. However, without that standardization the resultant computer systems may not be able to search and match satisfactorily. Worse yet, the computer system may be burdened with the weight of trying to resolve through software those problems which should have disappeared through user discussion.

Unfortunately, most installations are not first entering computerization when they contemplate a data base environment. They already have precedents, conflicting standards, and users who have become accustomed to their own way. It is the responsibility of management to see that these users come to agreement. Otherwise, they should not show surprise if the computer administration delays delivery of the data base system or if that system seems superficial or incomplete.

PREPARATION

Even when the commitment of time, money, and manpower has been made, not all data base systems are successful. The biggest reason is lack of planning. Having made a large commitment, the user communities understandably want results. The computer personnel, in an attempt to satisfy their users, too often rush headlong into the acquisition of a data base vehicle and the implementation of the

* John K. Lyon, "The Role of the Data Base Administrator," *DATA BASE*, Winter 1971, pp. 11-12.

first base systems. Somehow the estimates are always low and corners are cut in the attempt to placate the expectant communities. The first systems may even seem fairly satisfactory. This approach, although reasonably successful with the independent systems of the past, portends disaster for a data base system. The data base system is a self-propagating animal which feeds on the insatiable desires of its users. As more is learned about the capabilities of the system and the characteristics of the data contained therein, more is demanded. Widespread and diverse areas in the organization will become involved; new sources of information will be discovered. These conditions are healthy. However, if there were not strong foundations upon which the original data base was built, the additional demands may overtax the system. To avoid a short life for the data base, the long range goals, needs, and desires must be carefully analyzed even before the selection of the data base vehicle—and obviously before the design of the first application systems.

The data

The first element to be analyzed is the data themselves. The various types of data, both those which currently exist and those which would be desirable to obtain, should be analyzed. In-depth descriptions need not be detailed in the early stages of planning. However, those data which span traditional file boundaries must be isolated. (An example of this might be university instructor course load, which is used for personnel reports and also for student study lists.) These interrelated data may not seem critical in the early phases of a data base system. However, due to the tendency toward the development of more complex analyses and correlations which were unavailable in the past, these interrelated data must be structured within the data base in such a manner that future demands do not overburden the retrieval structures and/or security schemes of the new system.

All data should be classified according to their "owners" and "users". The owners are those people who can modify the contents or nature of the data—the traditional controllers of the file. The users are those people who are permitted to peruse the data. The owner/user relationships must be carefully studied to ensure that an appropriate security scheme can be evolved.

Usage patterns

Once the data have been analyzed and isolated, the retrieval and maintenance patterns of the data should be identified. These patterns will be used to determine the spectrum of data base features that are necessary and desirable and will help to determine the need for online retrieval. The 'currency' requirement of the data is also important: how current the data in the base must be to a large extent indicates the real need for an online maintenance environment. If the data can be a day or a week old, a pressing requirement does not exist. (Of course, online

maintenance, given that online retrieval has already been justified by need, may be more cost effective and is usually more desirable than overnight batch maintenance.) Since terminal equipment is still rather expensive, it must usually be carefully justified by need (or by enough available funds to justify desire). In addition, the establishment of an online environment is usually a one-way street: very few organizations revert to batch-only systems.

The data base vehicle

The characteristics of the retrieval and modification patterns should be summarized and condensed into a spectrum of required features (those functions/characteristics which must be facilitated) and a spectrum of desired features (those functions/characteristics which would improve usage and expansion of the system). These spectra may then be compared with those features provided by the available data base vehicles.

A thorough analysis of the available budget and manpower for programming and support must be done, both for the current time period and for several years hence. By comparing available resources and desired characteristics with costs and offered features, the organization should be able to find a suitable data base system (or to decide that any such system is beyond the current scope of feasibility).

DATA ADMINISTRATION

After the organizational needs and desires have been defined and a data base vehicle acquired, a live data base should not be established until proper controls over that system have been established. It should be clear from the records of those data base vehicles currently available that each of them has the potential for a successful data base system—and the possibility of failure. The difference is not inherent in the data base vehicle itself but rather in the manner in which it is used—or misused. The data base system must be controlled, or it becomes uncontrollable.

The specific method employed to produce the controlled environment is not important: the presence of such a method is. However, the trend in recent years has been to focus the responsibility for this supervision in one individual within the computing organization. He may be called a data base administrator (DBA), data manager, information specialist, systems controller, information scientist, communications administrator, or any of a number of other titles. His functions may vary from the purely technical to the strictly political. His responsibilities are often almost infinite, his authority all too often nonexistent or only token.

If the responsibilities and functions attributed to the DBA are even cursorily examined, it should be obvious that no single individual, not even a "superman", could adequately handle all of the work. It should also be apparent that the specialties required vary from those associated with computer technicians to those needed to converse with and instruct clerical personnel to yet others

directed at the involvement with middle and upper management. This diversity and breadth of functional responsibility is finally being recognized. A few installations have already begun forming a Department of Data Administration so that the diverse needs may be satisfied via more than one individual: perhaps a teacher, writer, technician, analyst, and user representative. A strong manager is mandatory to ensure consistency and the strong commitment that the user is all-important.

However, regardless of what the organizational structure is, a consistent attitude is shared by these installations: the recognition of the need for control over and monitoring of the data base. A summary of the responsibilities and functions follows.

Maintenance of the data base and its directories

One of the most recognized characteristics of the data base environment is the fact that control must be exercised over the creation, deletion, and reorganization of files. Most data base systems have sophisticated file structures which combine traditional files into a single data set, often on the same physical device. Many support chaining between logical files and even permit references to multiple records in a manner that is transparent to the requester. Caution must be employed when deleting or reorganizing one logical file since other systems may depend on its contents. The allocation of space for new files must be coordinated to ensure that overlapping does not occur and that usage on the various devices in the data bases can be balanced across all user areas.

The system files used by the data system must also be controlled. This usually involves the maintenance of a system dictionary or directory file and a file which contains data on the space utilization and boundaries of user data files in the data base.

As will become increasingly more obvious, the very sophistication which makes a data base system flexible and powerful also permits accidental catastrophes and seemingly self-destructive files. Many an installation has been forced to recognize the need for centralized control on major modifications to the data base when two programmers build a file in the same allocated space or when the system directory is destroyed by an inexperienced operator.

Control over the contents of the data base directory is also important, since that is usually the user means of access to his data. Because the definitions of all data items within the entire data base are stored within one directory and users may often gain access to more than one file, the names and structures within the dictionary must be monitored to ensure compatibility, uniqueness, and consistency with the names in the overall system. Revisions to the directory must be controlled for the same reasons demanded by the physical data themselves.

It is also necessary to organize the printing and distribution of the contents of the dictionary. Frequently, Data

Element Dictionary manuals, based on the stored contents of the physical dictionary, are prepared and distributed to the authorized users; it is, of course, very important to ensure that each user gain access only to those Dictionaries which contain data he may retrieve.

Security

Probably the duty most widely associated with the data base administrator is that of controlling the security aspects of the data system. Even in those environments where all the information in the data base is published in public documents the security requirements are stringent. This characteristic is no doubt precipitated by the new retrieval capabilities commonly offered the user: the more ease he enjoys in being able to analyze his data, the more apprehension he has about the ease others might share in accessing *his* data. As will be stated again, those requirements placed on a data base system because of user apprehension should receive the utmost consideration in that the primary purpose of most such systems is to provide users with more and better information.

Included under the enforcement of security are the assignment, reassignment, and modification of passwords and/or passkeys. Although the DBA normally supervises the password procedures, the policies and procedures governing the issuance of these passwords should be the responsibility of higher management or a consortium of users. Procedures for detecting the compromise of security keys and control procedures within the actual machine room may also be assigned to the DBA.

Integrity

The integrity of his data is very important to the owner of a file. The stability of the system with which he interacts is critical to the user of a data base. These needs are fairly easily met in the traditional computerized systems—and, even if not quite met, infringements can usually be minimized before reaching the user level.

In an online system, especially one which permits online updating, stability is a very elusive quality. Procedures on backup and recovery, logging systems and audit trails must be carefully established and even more carefully controlled. Backup hardware may even be needed, since the user who trusted his data processing department enough to discard his manual files will want to revert to them when he cannot get at his data because of a malfunction. Because of the integrated nature of the new files, testout procedures must be established and monitored for both software system modifications and new application systems.

The standardization of the codes among files is also related to data integrity. Inconsistencies between traditional files were easily tolerated in that only the computer program need know. But when the user begins to input his own retrieval requests he becomes easily discouraged if con-

fronted by such situations as name being stored surname first in an admissions file yet surname last in the student master file. Unfortunately, although it is easy to convince users that consistency in codes and data content is critical, it is almost impossible to get them to agree on a change to their standard.

Quality control and performance measurement

In addition to the need for standardization of data content, the data base systems require a certain amount of standardization of application design, especially where terminal interaction with a user is involved. To the user all communications derive from a common source—the computer—and he will not respond happily to major differences in the responses requested of him.

In addition, since data base systems frequently support multiprogramming and concurrent accessing of data, lack of efficiency in a given program may affect the performance of another application or even of the entire system. Thus, standards must be established for application systems and these standards must be enforced by someone outside of the environment which always demands too much in too short a time.

Because of the user cognizance and the tendency toward sophisticated system activity, performance measurements become an important responsibility of the data base administrator. Hardware and software monitors may be used to measure CPU and hardware utilization and the distribution of access. For those systems supporting interactive devices, system and user response times should also be measured. The activity by terminal user is informational and often required for accounting interfaces. Such measurements may be used to assist terminal users who show abnormal usage. The performance of the system expressed in quantitative form often dispels user convictions that the system is slow or that “everybody else” is benefiting more.

Consulting and training

Because of his exposure to so many aspects of data management, the DBA often serves as the consultant, instructor, and technical advisor for his installation. Programmers and analysts need assistance in design problems and may also need help in program debugging under the software system. In addition to knowing all things about the data base system in use, the DBA must keep current on the “state of the art” and knowledgeable of other software and hardware, especially the data base packages on the market.

Programmers and analysts in the installation require much training in the techniques for data definition, file design, security features, application design, coding techniques, and—of course—installation standards. All of these guidelines must be incorporated into formal documentation.

Users and management must also be educated in the use

of the data base and its features. Even using the terminal imposes a threat to many, in that they feel the pressure of the “meter running” and some even fear that “something” is looking out from inside the terminal screen.

User liaison

Communication with the user is the single most important role of the DBA, since in the long run it will usually be the reactions and attitudes of the data base users that will ‘make or break’ usage of a data base system. If the user is displeased or confused, he will not be receptive to a more sophisticated system. If he does not trust the security features or feels frustrated by his terminal, he will yearn for the old, more comfortable procedures. For these reasons the user attitudes must be considered, no matter how trivial or ridiculous. Frequent interactions and status reports, immediate assistance with terminal problems, and a log of trouble calls and user problems will help to comfort the unfamiliar user. Since interactions with the nontechnical populace are frequently difficult for computer personnel, the DBA may be the sole source of such ‘hand-holding’ talent.

IMPLEMENTATION

Once the procedures for sustaining a good data base have been established, the organization can begin to implement its own data base systems. Here, however, caution must be employed. Data base systems are not easy to design. On-line systems are a new technology and not much expertise is yet available. The new systems are used so directly by non-computer-oriented people that the primary emphasis must be placed on an audience unfamiliar to the computer professional.

The greatest problem usually becomes apparent in the use of the data base vehicles themselves. The new software is complex and capable of sophisticated applications. However, it is nothing more than a tool: it must be used with skill and expertise in order to produce satisfactory results. A data base vehicle is like a typewriter. Although it is a well-established fact that the typewriter has greatly increased the productivity of secretaries, no one would expect anything but a disaster if an untrained person were required to exchange longhand for a typewriter. Yet most installations expect far greater miracles of their computing organizations when a data base vehicle is introduced. And, unfortunately, since most computing organizations do not realize the complexity of the new technology, the users of the first system may become the victims. Furthermore, although there are many fine secretarial schools which can consistently produce speed typists in a short amount of time, there are very few ways to learn data base technology and no ready substitute for one to two years of hands-on experience in a production environment.

As a result, the disaster-prone patterns are becoming

familiar: The computer processing department plunges directly into the design and implementation of a major application—one with a high exposure profile and importance to the entire community. In an attempt to show results quickly, analysis is cut short, user interaction is curtailed, and the less visible features (such as edits and controls) are delayed “until later”. There is little time to spend in training the programmers in the use of the new data base vehicle. The resultant system is obviously weak, but the lack of efficiency and shallowness of design are not always apparent until the second or third system is added. Eventually the hastily built foundations of the data base begin to crumble and work must begin again.

Thus, it is wise to approach the implementation with caution. Work should begin with a small low-profile application and expectations should be kept low. Above all, it should be recognized that not all of the problems are due to

computer professionals. A true story illustrates the point: Having recently joined the computer professional staff, one of our analysts was especially interested in making a good impression with the user community. When one of them called to complain that she had not received her Numbers Report that week, he assured her that he would check into the situation. When none of the programmers could figure out which particular report might be the missing one, he went to the user, who promptly showed him a neat file wherein he was astounded to find page after page of *memory dumps*. It seems that one of the report programs had always encountered problems and the programmer, in order to facilitate debugging, had appended a partial dump of memory to the printouts of the reports. The corrections had finally been made and the dump option removed. The unquestioning user had simply filed the copies and worried only when the Numbers Report failed to appear.