EDP ANALYZER

MAY, 1978
VOL. 16, NO. 5

PLANNING FOR DBMS CONVERSIONS

It is estimated that less than twenty percent of today's mechanized data is stored in data bases, under the control of data base management systems (DBMS). But the rapid growth in sales of DBMS and related software (such as data dictionaries) leads one to believe that by 1985, this percentage will be much higher. Further, the organizations that *are* using DBMS today are finding that they want to (or have to) upgrade to a new DBMS every few years. So conversions from non-DBMS to DBMS, or from one DBMS to another, will be very much a part of the near future workload for many computer users. Here is a summary of two working conferences that have aimed at developing management guides for these conversions.

Mayford Roark, Executive Director of Systems for the Ford Motor Company, Dearborn, Michigan, was the keynote speaker at a working conference held last November and sponsored by the U. S. National Bureau of Standards and the Association for Computing Machinery. The title of the working conference was *Data Base Directions, The Conversion Problem*.

This was the second such working conference on data base directions sponsored by NBS and ACM. The reports of the two conferences will be found in References 1a and 1b. For ease of reference, we will call the October 1975 conference 'DBD I' and the November 1977 conference 'DBD II'.

As the keynoter at DBD II, Roark addressed the problems of converting to new hardware, operating systems, and DBM systems, based on the experiences at Ford. His company had recently made a catalog of its computer applications throughout the North American divisions. In total, the survey catalogued some 50,000 computer programs, most of which used their own data files. Some of these data files were huge, measured in billions of characters. A typical division of Ford, however, is very much like a medium size business, said Roark; it might have between 1500 and 3000 computer programs.

While the survey did not explicitly study the amount of the data processing that is data base oriented, some clues were to be found in the catalog, Roark said. From these indications and from personal contacts elsewhere, he estimated this amount to be only 10% to 20% of the total data processing for U.S. industry as a whole. This indicates a much slower acceptance of data base technology than was expected five years ago. Roark saw several possible reasons why this was the case at Ford.

Some Ford divisions were saddled with a heavy system maintenance workload during the past five years. These divisions felt that they could not spare the resources that would be needed for converting to a data base.

A few divisions had tried to convert to a DBMS and had run into conversions traumas. These divisions had recovered from the difficulties but then decided that 'once was enough.' They did not undertake any more conversions to a data base during the five year period.

Some other divisions had installed DBMS successfully and have continued to extend their use of data base technology. One Ford manager, perhaps the most enthusiastic advocate of data bases, told Roark that he thought about 40% of his division's data was under DBMS control, after five years of development, and that by another five years, the amount might reach 70% to 80%.

In general, Roark observed, the biggest obstacle to the adoption of data base technology is the historical accumulation of systems. If they could, most companies would like to get rid of their 'old mess' and start over. But the existing systems represent an investment of hundreds of millions of dollars at Ford, and many tens of billions of dollars for computer-using organizations collectively. "Where does even a moderate size company with, say, 1500 to 3000 computer programs, begin when converting to data base technology, without having to scrap its investment in programs and data files?" he asked.

But even if an organization does successfully install a DBMS and gradually extend its use, the conversion problem does not end there. Roark saw three continuing problem areas that are affected by the existence of the DBMS.

Hardware conversion. Historically, the capacity of most data processing computers is saturated every three to five years and the computers are replaced. Roark cited three situations. In the first case, if the replacement involves a change to a *new supplier*, and if a DBMS has been used, then a shift to a new DBMS probably will be required. Such a DBMS conversion may or may not be a problem, he said. One Ford division converted from a Honeywell computer, using IDS, to a Burroughs computer that used DMS II, two quite different systems. But these differences in the DBMSs did not materially affect the conversion effort. However, such smoothness cannot always be assumed, he added.

In the second case, if the hardware change involves a switch to a *new* family of computers of an *incumbent supplier*, which seems to occur every ten years or so, it is hard to say what trauma will be caused by the DBMS. This kind of change, involving a DBMS, has not yet happened at Ford.

Finally, if the hardware change involves an upgrade within an existing family of computers, the existence of a DBMS has not been much of a factor in the conversion, Roark reported.

Operating system conversion. Roark said that he expects major operating system upgrades or conversions from any one supplier every five to ten years, regardless of how 'permanent' the supplier claims the current operating system to be.

As an example of such a conversion, the large Ford data center at Dearborn is undergoing a conversion to IBM's MVS operating system. Three DBMS are in use: IMS, TOTAL, and System 2000. Data and programs used with these DBMS have to be converted to run under MVS. The conversion of the IMS applications appears to be about twice as difficult as would have been the case if no DBMS were involved, he said. Further, while one manager saw benefits from IMS/VS, another saw no advantages to his applications from it, but he had to go along with the conversion because the old IMS would not be supported under MVS.

The managers whose applications were running under TOTAL and System 2000 encountered no special conversion problems because of those DBMS, said Roark.

Evolutionary application development. Roark reported that Ford's use of computers was growing at the rate of about 20% per year, measured in terms of billions of instructions executed. He estimated that about one-half of this growth was for new applications and the other half was for changes and enhancements to existing applications. These changes and enhancements are needed because of new government regulations, changes in industrial standards, growth in product lines, corporate reorganizations, and efficiency improvements.

In theory, the existence of a DBMS should make the development of new applications and the changing of existing ones easier. In practice, the expected benefits can be hard to realize. Yes, changes can be incorporated more quickly and it is easier to develop new applications. Further, a DBMS greatly reduces the difficulty of making special analyses, since all of the needed data may be in the data base; previously, data might have to be drawn from five to ten different files with different maintenance cycles. But these benefits are obtained at a price. One study at Ford showed that some 70% of the instructions that were executed resulted from the DBMS overhead. In another instance, about half of the storage needed for a very large data base was consumed by pointers, far in excess of what was expected.

As application systems are converted to a data base, or converted from one DBMS to another, some subtle and hidden problems can come to light such as those just described. Performance may be so far below expectations that a significant redesign effort may be needed. This, in turn, may discourage the conversion of other applications to the data base.

The realization of the promise of data base technology is still a long way off, concluded Roark. The payoff will come when the decision to go data base is no longer a high risk affair that requires an all-out commitment in order to solve some of the most difficult conversion problems to be found in systems.

Some 'war stories'

Professor Edgar Sibley, of the University of Maryland, has been involved with data base technology for a number of years. Among his many computer field activities, he has been the chairman of the CODASYL Systems Committee. In a paper presented at the IFIP Congress 77 last year in Toronto, he discussed some DBMS 'war stories.'

Difficulties encountered in installing DBMS are not well documented, said Sibley. The potential authors may be intimidated by the DBMS suppliers. Also, the situations, if publicized, may be used 'politically' against data processing within the companies. Thus, the real-world experiences that Sibley discussed had to be well disguised, so the organizations could not be identified. But the cases are real ones, he said.

His contacts with many users of DBMS has led Sibley to one prime conclusion: *There are hardly any technical problems involved with installing a* DBMS, *only people problems and managerial ones.* Witness a case he described that had an abnormal complement of mistakes.

In this case, to oversee the first application of a newly acquired DBMS, the company selected a systems programmer who 'could be spared' and who was mainly interested in getting out of systems programming. Not only was this choice of the key person poor but so was the choice of the first application. The implementation of that application took much too long and it was not useful after it was running. This experience convinced user departments at this company that a DBMS was inefficient and just another programming fad. Programmers, too, looked on the DBMS as inefficient and preferred to use 'conventional access methods.'

Efforts to use the DBMS continued, and mistakes continued. Some old card oriented and tape oriented application systems were converted to the data base without redesign. Records were accessed by scanning through all records rather than by direct access. Little effort was made to develop common data definitions. Multiple, application oriented data bases were used, so that it was difficult to extract data for reports that drew from more than one data base.

As might be expected, the DBMS was not looked upon as a panacea at that organization.

At another organization, management expressed an interest in an integrated data base to serve multiple applications. But for a number of reasons—such as managerial weakness, lack of commitment, or poor communications—some significant internal problems arose that caused resistance to the use of the DBMS. For one thing, programmers resented giving up the file design function to the data administrator. Systems programmers felt that the 'glamorous' DBMS was displacing the operating system as the center of attention. User department people felt that they were losing control over 'their' data. Data processing management authorized the designers of one application system to use more system resources than the installation standards allowed; this led not only to resentment by the designers of other systems but also to a bad deterioration of on-line response time for *all* data base applications. Finally, the top DP managers recognized that *they* were a part of the problem and started to learn about data base technology. The company gradually overcame these problems, but at a non-trivial cost in time and money.

Sibley reported one manager as making a particularly pertinent observation. "Don't start the wrong way," said this manager, "or it could take you ten years or an organizational crisis to change and do it right."

'Doing it right' requires the following, according to Sibley. First, a reasonably high level of top management acceptance and support is needed; a DBMS will cause important shifts in responsibilities. Second, use competent managers, to provide good management control of the DBMS project. Third, assign competent technicians, including a data administration staff, with proper tools. And finally, get competent auditors to monitor content, security, and privacy of the entire data base system. In addition, he recommended that a data dictionary be installed before the first data base application is fully operational.

How are DBMS being used?

We are using the term 'data base management system' in this discussion in the context of a data base where the data is shared among multiple applications. Probably the first true DBMS was the General Electric MIACS system, developed for internal production control purposes by Charles Bachman in 1961. In 1963, this grew into the G.E. Integrated Data Store; (IDS is now offered by Honeywell Information Systems, Inc.). We are not including file management systems, which are mainly being used to handle tape or disk application files.

As indicated earlier in this report, probably less than 20% of today's data processing involves true data base technology. Most computer users still have their data in application files stored on magnetic tape or on magnetic disk. When stored on disk, access to records is either done sequentially (as on tape) or by a conventional access method such as index sequential.

A small percentage of computer using organizations have attempted to install a DBMS, have failed in the attempt, and have pulled back. Perhaps the wrong people were assigned to do the job, or the 'politics' was wrong, or the people tried to put in too sophisticated a system. Whatever the reason, these organizations are not anxious to try a DBMS again.

Some other organizations have a percentage of their applications running under one or more DBMS, but with little or no sharing of data among the applications. 'Data bases' in these organizations are much the same as application files, and the DBMS is used as just a fancy new access method. If the data definitions for the tape and disk data files are not under control at these organizations, then chances are that the data definitions for 'data base' data also are not under control. The result is that there are duplications of data, inconsistencies, and so on, just as in the case of the older data files.

A very small fraction of computer users have integrated data bases, serving multiple applications from the same data base. These are the companies who are most likely to have the data definitions cleaned up. They also are the ones most likely to have installed the data administrator function, for controlling data definitions and file design.

But even if a company is using a DBMS, whether on a shared data basis or not, one point is very generally true: only a small fraction of that company's mechanized data is in the data base. Such a company might well have thousands of computer programs with their asssociated application files on tape or disk. The conversion of those programs and files to data base technology would involve a relatively large cost. Frequently there is no real incentive for making the conversion effort and expense. And in some cases, the applications are inherently stand-alone, with no possibility for sharing data with other applications.

Most of the above discussion of data base technology has singled out the DBMS as *the* main feature of the technology. This is not actually the case. Let us now briefly consider the total data base environment.

The total environment

Enough use has been made by now of data base technology to get a good idea of just what is involved when an organization 'installs a DBMS' or 'converts from one DBMS to another.' Here are the elements that make up the total data base environment and that must either be installed or possibly changed, as pointed out at DBD II.

Data definitions. In addition to the definitions of data fields and data records, data bases generally involve explicitly defined relationships between records to a greater extent than has been true in conventional files. So all data items and data relationships will have to be defined to the extent required by the DBMS. Further, if changing from one DBMS to another, these definitions may well have to be changed.

Data files. The data will have to be converted from the previous files (either mechanized or manual) to the data base, or from the old data base to the new one. Not only must the current data be considered but also the backup and archival data.

DBMS. This is the component that normally receives most of the attention. The new DBMS must be installed and whatever was used previously must be phased out.

Computer programs. Data independence is really not yet within the state of the art. When the method of storing and retrieving data is changed, then the associated application programs usually must be changed. In fact, this might well be the biggest job in the conversion project.

Data dictionary. Many users of data base technology have come to realize that they should have installed a mechanized data dictionary, to get their data definitions under control, before they installed a DBMS. Computer users who have not yet converted to data base technology should keep this point in mind. Converting to a DBMS *ought* to involve the installation of a data dictionary. Converting from one DBMS to another might require a change in data dictionaries, particularly if one dictionary uses the DBMS it serves for managing the definition data.

Query and report writing facilities. One of the benefits of a true data base is the convenience of answering ad hoc queries and report requests. So query languages and report writing languages have been developed, for use with data bases, and their use is growing rapidly. Further, users often develop catalogs of pre-defined queries and reports. These facilities are important elements of a data base environment. If the DBMS is changed, both the facilities and the catalogued queries and reports may well have to be changed.

Data administrator function. Another thing that users of data base technology have learned is that they should have set up the data administrator function even before they installed a data dictionary, and have both installed prior to the installation of the DBMS. This function controls the data definitions and the structuring of the data; we discussed the function in our November 1972 report. Converting from one DBMS to another may require a change in this function, perhaps because the new technology may be beyond the interests or capabilities of existing staff members.

Security and integrity functions. Most computer users have recognized the need for data to be accurate, relevant, timely, and protected from destruction and unauthorized disclosures. New government regulations, such as privacy laws, raise the level of concern. Converting to a DBMS, or from one DBMS to another, may involve many associated changes in procedures to support data security and integrity.

These, then, are the major elements that we view as making up the total data base environment. They lead us into our next subject.

How should DB technology be used?

One has only to read the trade press to become aware of the growing use of data base technology. There are more and more advertisements for DBMS, for data dictionaries, for query and report writing packages that work with DBMS, and for network services that use data base technology. Articles tell about user experiences. Classified ads seek people trained in these areas. And so on.

In addition, data base technology is being extended into the mini-computer and even the micro-computer areas. Distributed data bases, using minis, are almost (but not quite) here. The complexities of data base management are such that they stretch the capabilities of most of today's minis, we believe, and are probably beyond the capabilities of today's micros. Even so, we expect to see some aspects of data base management spreading among mini-computers and beginning to appear in some micros. (See Reference 9 for information on one new CODASYL-type DBMS for some minis and micros.)

As our discussion earlier in this issue pointed out, converting to data base technology is no simple matter. Mayford Roark called it 'the most difficult conversion problem of all,' mainly because of the historical accumulation of the programs and data files that would have to be converted. But even if there is no attempt to convert existing systems, there are severe problems that can be encountered, as described by Sibley.

What can computer users do to avoid the failures and the messes? What is the best way to convert to data base technology for the first time? What steps can be taken to ease the conversion from one DBMS to another?

To help answer such questions as these, the U.S. National Bureau of Standards and the Association of Computing Machinery have jointly sponsored two working conferences with the generic name *Data Base Directions*, as we mentioned at the beginning of this report. Each working conference involved about 60 invited experts, people working near the forefront of each sub-topic addressed by the conference. The goal of each conference was to produce a report of the conclusions of the working panels that addressed the sub-topics, based on two and one-half days of deliberations. The first working conference (DBD I) was held in late 1975, and DBD II was held in late 1977. References 1a and 1b are the two reports.

The working panels of DBD I addressed the subjects: user experiences with data base technology, future trends in data base technology, the need for data base standards, the expected impact of government regulations on the use of data base technology, and the impact of this technology on auditing.

The topics addressed by the panels at DBD II included: user experiences with data base conversions, management considerations when making these conversions, how standards might help in conversions, and what conversion aids are being developed by the technology.

So the two working conferences produced what we think is a good amount of practical advice for management. The remainder of this report will summarize what was covered at the conferences, in an intermingled fashion.

DATA BASE DIRECTIONS

Our discussion of the two DBD working conferences will cover the keynote speech of DBD I, plus main recommendations of the working panels: user experience, management considerations, technology developments, the need for standards, the impact of government regulations, and the impact on auditing.

A manager's viewpoint

Daniel B. Magraw, the keynote speaker at DBD I, is Assistant Commissioner, Department of Administration, State of Minnesota. For nine years, he has been responsible for all aspects of the State of Minnesota information system activities. He described the problems of introducing data base technology into an organization, from a manager's viewpoint.

The broad question of data management lies at the very center of the organization management process, said Magraw. The essence of management is to rationalize decision making based on the best available data. So in dealing with the data of the organization, one deals with the heart and soul of management.

Magraw sees eight areas of particular concern as one thinks of installing data base technology to manage the organization's data resource. These areas are skewed heavily toward user management's understanding and involvement.

DBMS objectives. Why are computers not being used more widely to support management's decisions? As Magraw sees it, the fault is not with the technology but with the failure of management to define the decision system so that the needs can be addressed.

It is essential that management establish DBMS objectives, he said. Otherwise, the (ill defined) decision system can be adversely affected. "One simply does not fiddle around with the most precious of all raw materials of an organization, its data. It is simply crucial that the target be clear," he said.

Realism. In order to avoid the 'Great Expectations Cemetary,' one must be realistic; be wary of over-stated claims. A data base system just will not go in overnight nor will it automatically save tremendous amounts of money.

Organization. An unequivocal statement of responsibility and a proper structure must be provided for the data management function, or an expensive, academic exercise will result. The new organization should include strong, central, total authority over data definitions and related DBMS functions.

Cost/ benefit. Management needs a clear cost/ benefit picture before deciding on installing data base technology. Consider all significant factors, said Magraw, including the non-hardware, non-software costs that are related to the organizational, structural, and disciplinary changes that are needed.

Transition. "I am simply flabbergasted at the cavalier dismissal, by friend and foe alike, of the overwhelming prospect of putting the data bases in order for a DBMS." The problems of getting good data definitions and data structures cannot be so easily dismissed, he said.

Training. Insufficient training for computer professionals seems to center on the areas of decision theory, systems, and man-machine dialog, according to Magraw. But the greatest need is for management training in information systems in general and data base management systems in particular.

Privacy. (Note: the State of Minnesota was a pioneer in the U.S. in the adoption of a privacy law, so Magraw has had first hand experience with this type of legislation.) New government regulations will affect the ability to collect data on individuals and to inter-relate such data from various sub-systems, he said. Since data base concepts tend to inter-relate data items, this will be an area of increasing concern.

Security. "Can we justifiably expect greater security capability with data base technology? Do we become more vulnerable to some catastrophic event?" These were only two of the questions raised by Magraw. In short, he says that there are still a number of unanswered questions about the security aspects of data base technology.

As Magraw sees it, these and similar concerns should properly be addressed by user management, if the pitfalls of data base concepts are to be avoided and if the benefits are to be delivered.

Management considerations

Two types of conversions were considered by the panel (in DBD II) on management objectives, led by Richard Nolan. One type was converting from non-DBMS to DBMS, and the other type was converting from one DBMS to another. While there were some similarities between the two types, there were also some significant differences.

A general conclusion of this panel was a restatement of a point made by Sibley in Reference 2. That is, while it is appreciated that technological problems exist in data base conversions, it should be recognized that a DBMS conversion is primarily a *management* problem, for both executive management and data processing management.

The panel identified four key points for management to keep in mind, regardless of the type of data base conversion involved.

When, not whether. Management should recognize that the problem of converting to data base technology is not one that will just 'go away' if it is ignored long enough. It is a question of when, not whether. The natural evolution of data processing is moving organizations into this environment.

At the same time, the panel stressed that data base technology probably will not be suitable for all application systems, at least in the foreseeable future. Some applications will be more properly served by individual application files stored on tape or disk.

When is data base technology likely to be adopted by an organization? The panel found it very helpful to use Nolan's four stages of growth (Reference 3) for answering this question. *Stage one* is the getting started stage, when the first attempts are made to use a new technology. Functional areas are selected where the risks are not too great and where positive payoffs seem likely. *Stage two* is the proliferation stage—adding more functional areas and building on to existing ones. Both maintenance needs and overall costs grow to the point where it is hard to work on new applications. *Stage three* is where management becomes concerned with the rising costs and the 'mess' that is being created. Control is imposed and clean-up is started. Finally, in *stage four*, a mature approach to the use of the new technology is reached. The organization is ready to begin using some even newer technology.

To return to our question: *when* is an organization likely to adopt data base technology? It is near the end of stage two that the need to get things under control becomes apparent. As an organization gets into stage three, data base technology is almost a necessity, said the panel.

Most computer using organizations are still in stage two. Proliferation in the use of computers has led these organizations to develop thousands of computer programs and thousands of application oriented data files. The 'mess' is now apparent. Some are beginning to move into stage three and to install data base technology. Many (or even most) new applications are being developed using data base concepts. But it will take a long time for all of the existing programs and data files to be converted to these concepts.

Similarly, during stages three and four, DBMS users will find the need to convert from one DBMS to another, for any of a variety of reasons mentioned earlier in this report. Further, much more than just the DBMS will be involved, as we have indicated. The data files, the application programs, the data dictionary, etc. are also subject to change.

So these two types of conversions—to a DBMS and from one to another—are for all practical purposes inevitable. But we should reiterate: this does not mean that *all* applications need be on the data base.

Choose the entry application carefully. If the use of data base technology gets off to a bad start, it can delay the acceptance of this technology for a number of years. So it is very important that the first step be a good one.

The first data base application preferably should be a non-trivial one, said the panel, but also a non-vital one. Recognize that this first application will be a learning experience. Once it is running, though, it should have good visibility for the user departments so that they can see the advantages of the data base. It should not be buried deep within data processing.

By all means, said the panel, avoid the common problem of trying to do too much with this first application. Do not try to demonstrate the 'full power' of the DBMS by designing extra fancy data structures.

However, the panel recognized that, in some instances, the only reason for putting in a DBMS may be to handle an application that is critical to the organization's business. In such a case, the designers must be very careful to avoid a too-complex data structure. The combination of data base complexity and criticality is *very* risky and should be avoided.

Use good project management. Treat a DBMS conversion project like any other system project, said the panel. Use normal, good practices for justifying the project. Use good project management principles that cover the whole life cycle of the project, from justification through installation and maintenance (which we discussed most recently in our December 1977 report). Make sure the user department representatives play an important role in the project management.

Plan for future conversions now. The panel saw the use of data base technology as a way of life for the computer field. The first DBMS that an organization installs will be just that—the first of many. So it is not too soon to begin planning on how to convert from one DBMS to another, even if the first one has not yet been installed.

The panel offered some management guides for this planning. Set up the data administrator function, with proper responsibility and with adequate authority over the data resources. Also, minimize the dependence of the application systems on a specific DBMS, perhaps by building some in-house software to act as a standard interface between the application programs and the DBMS functions.

Just how practical were these recommendations by the management objectives panel? To see, let us now review what the user experience panels at DBD I and II reported.

User experience

The user experience working panels at both conferences included people from private industry and government who had gone through data base conversions.

One of these panels thought that technologists view data base technology in terms of the 'great' things this technology can do—things such as superior access to data, easy multi-file processing, single source for data entry and edit, and on-line access for answering ad hoc queries. On the other hand, users tend not to view data base technology in this manner. Instead, users are puzzled by the new buzz words (such as 'schema'), they fear the total exposure and risk that a data base involves, and they are quite suspicious of the claims for benefits that are made.

Moreover, this panel warned against over-ambitious starts in using data base technology. Such starts result in greater than necessary times and costs to implement the applications. Users may be (inadvertently) encouraged to ask for more than they need because of the great promise of the technology. Then, when problems are encountered, alientation of users and management occurs.

So instead of an overly ambitious start, plan to deliver small data base oriented systems in a short time frame. Build up confidence among the users in what the technology can do for them. Pay particular attention to designing forgiving and fault tolerant procedures for terminal operators, for each new data base application. And do not attempt to give full access to all data base data from the outset.

The other panel on user experience made many similar observations. Take a phased approach to implementing a data base; such an approach is easier to control and more likely to succeed. Design the data base so as to minimize the impact on other existing systems. In general, do not attempt to write the DBMS software inhouse.

Future developments in technology are not likely to obsolete any significant portion of the effort to use data base concepts effectively. So there is really no need to wait 'until the tools are perfected.' Today's technology provides adequate tools with which to get started.

At the two working conferences, attended by people who had been working deeply in the data base area, much the same theme was repeated over and over again. This theme was: Take it easy. Take small steps. Have realistic expectations. Treat a data base project like any other system project. Start *now* to make future converions easier.

But how about the new technological developments? Surely, with all of the progress that is being made in the computer field, there must be some exciting developments on the horizon. Will not these developments make conversions to data base concepts much easier? These were the types of questions addressed by the two panels that dealt with technology.

Technology

The members of the technology panels at the two working conferences were drawn from among people who are working at the forefront of data base concepts, on developments that have not yet reached the marketplace. So these people were in an excellent position to know just what is likely to appear in the next few years.

At DBD I, the panel concluded that new types of data base architecture were to be expected in the not-distant future. One such architecture was the back-end computer that would handle data base management, in much the same way that front-end computers now handle data communications functions. The distributed data base was recognized as an as-yet unsolved technical area, but the panel expected to see commercially available distributed data base systems by 1981.

The flow of new data models is expected to continue, said this panel. These models provide the means by which data 'models' the real-world relationships. The better known models include the CODASYL approach, the hierarchical model, and the relational model. Panel members were experienced with developing most of the leading models and were aware of their strengths and weaknesses.

Then came what we considered to be one of the more significant observations of the conference: "Each of these models has proponents who point to advantages for their model and suggest that these models are decisive. However, the panel saw no 'best model'; further, it will be hard to conclude which model is best within the next five years. We recommend that the user select the model that presently best fits immediate and near future problems. In terms of expected advantages, presently proposed new models are, at best, evolutionary rather than revolutionary."

Considerable progress was expected to occur in the area of data independence, wherein application programs are independent of certain types of changes made to the data definitions. Progress was also foreseen in improved facilities for recovering data bases from failures. But no 'great breakthrough' was foreseen, that would make conversions to data base technology much easier.

(We might add that in the interim, since DBD I was held, what has actually happened in data base technology is very much in line with what the panel expected to happen.)

Conversion technology

The technology panel at DBD II concentrated on conversion tools. Panel members included people who have been working on some of the most advanced conversion tools, to aid in converting from tape or disk files to a data base, or to aid in converting from one DBMS to another.

What the panel concluded was that technology today can address only a small fraction of the overall problem. The problem is complicated by the thousands of poorly documented programs and thousands of undefined data items at the typical computer using organization.

Data and program conversion are never-ending problems, the panel observed. The need to convert is caused by new technology becoming available (or by a dissatisfaction with inadequate old technology), by new user requirements, and by new functions to be performed.

The use of data base standards can help ease these conversion problems. But standards will not eliminate the problems.

Automated tools to help perform conversions are, today, limited both in their number and in their capability. The most successful conversions have been ones that have used teams of people who are experienced in performing conversions; further, these teams generally use manual tools and techniques.

However, some automated tools are on the horizon. For instance, the data translator being developed at the University of Michigan will read a source data base, translate the logical relationships of the data according to prescribed rules, and then write the target data base. But the translator does not change the application programs to work with the new data base, and the panel saw this program conversion problem as more difficult than the data conversion. Also, the translator will not handle changes from one brand of hardware to a dissimilar brand.

What can be done to reduce conversion problems? The panel made some recommendations.

General recommendation. Make a controlled and intelligent use of today's technology. There are tools and techniques available now that can help, if they are used intelligently. But none of them are 'magic formulas' that guarantee success.

Define the data. One of the most basic things that should be done is to explicitly define and document the data definitions and data relationships. An automated data dictionary can help get the data definitions under control, particularly within one line of equipment. Do not allow user maintained data relationships to be continued, where the users 'know' that certain data items or files are related. All relationships should be defined explicitly and maintained in the dictionary, to as great an extent as practical.

Select from current techniques. A good data base design will reduce the need to restructure the data base, so search out good design techniques. Good program design methods provide stable, maintainable modules. Develop in-house program/data interfaces and use them in the application programs; if the DBMS is changed, only the interface routines (probably) will have to be rewritten and not the application programs. Bind all parameters at execution time and not at compile time.

The defining of data, the design of the data base and application programs, and the building of stable program/data interfaces are all part of today's technology. If used intelligently, said the panel, they can materially aid the conversion effort.

Conversion is a user responsibility. "Let the vendor do our conversion" is a risky approach, the panel observed. The vendor almost certainly will limit the project to a straight conversion and would do no redesign. But redesign may well be essential, because the data structures in the new data base may be so different from the structures used previously. Conversion really cannot be delegated; the user organization must assume responsibility for it.

Develop and use standards. The technology panel at DBD I expressed concern about the effect of standards on emerging technology. But the panel at DBD II advocated a continued development of standards. If technology gets too far ahead of standards, said this panel, it may become a hopeless task to try to develop the standards. The panel also pointed out the problem of getting installation standards accepted across organization boundaries.

The value of standards is often not appreciated by senior management, so there may be no great motivation on the part of management to adopt standards. It is up to the technologists to point out how conversion and operations can benefit from the use of intelligent standards.

Which raises the question: just what is the status of data base standards? A panel at each of the working conferences addressed this question.

Standards

At DBD I, the standards working panel pointed out that data base standards must embrace more than just the DBMS; the standards should cover the various aspects of data base usage. Moreover, such standards are an international concern and responsibility, and their development should be coordinated on an international basis.

The panel saw the need for four types of standards. *Terminology standards* are needed to promote understanding and learning. *Criteria standards* are needed for use in evaluating proposed data base standards because of the variety of interests involved. *Component standards* are needed, for data elements, data definition languages, data manipulation languages, and so on. Finally, *usage standards* are needed to protect the integrity of the data bases.

This standards panel also made a significant observation. As far as a DBMS standard is concerned, the CODASYL proposal was (and still is) the only candidate submitted for possible standardization. No other candidate has been formally entered. Since it takes well over five years for a proposal to go through the standards process, the only likely standard in the next few years is the CODASYL approach.

The panel at DBD II stressed that the components of a standard DBMS should be delineated, indicating what the standard DBMS should consist of. The component standards should tell *what* logical features are to be provided, not how they will be provided, in order to provide flexibility for evolving technology. Further, there well may be a need for multiple DBMS standards, not just one, for serving different needs.

While it probably will not become a 'standard,' the panel strongly urged that the good practice be followed of installing a data dictionary and getting the data definitions under control before installing a DBMS.

Recognize, too, said this panel, that there will be a spectrum of standards. The spectrum includes installation standards, federal standards, national standards, and international standards.

Government regulations

This working panel, at DBD I, included representatives familiar with state and federal information systems. They sought to predict which regulations then in existence or expected would relate to information systems, and what the likely impact of those regulations might be.

The panel identified twenty areas for which it believed government regulations will come to exist nationwide, affecting both public and private sectors. The seven existing or anticipated regulations with the greatest expected impact appear to be the following: (1) Regulations that say the system must prohibit all data accesses except those specifically authorized. (2) Regulations requiring that information systems be certified that they meet all specified regulatory requirements. (3) Regulations that define a data subject's rights of access to information about him. (4) A standard DBMS for federal use. (5) Regulations specifying that certain types of applications must use dedicated data processing systems (example: criminal records systems). (6) Regulations on the accuracy, completeness, and timeliness of personal data held in information systems. And (7) regulations specifying that a log of all changes and disclosures of sensitive data be maintained.

The panel identified and discussed thirteen other possible regulations, but the above seven seemed to us to have the greatest expected impact.

The panel felt that the use of a DBMS might well make it easier to respond to new regulations, or to changes in regulations. But a DBMS also brings with it the risk of chaos if the single source of data (the data base) is destroyed.

The responsibility imposed upon data processing management will increase as the number of new regulations increases. Managers will have to determine that the hardware, software, and procedures will indeed carry out the intent of the regulations. Further, any penalties imposed for failure to comply adequately with the regulations might fall most heavily on individuals in data processing.

How does management determine if the data is being used and handled properly? That question was addressed by the audit panel.

Auditing

Auditors have always been concerned about control, security, and integrity in (financial) information systems. How does the advent of the data base affect this concern? The panel concluded that the roles of internal and external auditors will not change in a data base environment, but there will be a shift in emphasis on what an auditor must perform. This shift *may* involve the certification that regulations are being met, as raised by the government regulations panel.

Just what differences might data bases bring about in auditing? Here are some of the observations.

Security. As a general principle, auditors recommend the separation of duties in the handling of assets and liabilities. A data base makes control more difficult because it aggregates data rather than fragmenting it by application. An individual might be able to access a wide variety of company records and manipulate them for his or her own purposes.

Integrity. Auditors are concerned with errors of input and with file integrity. In a non-data base environment, the same transaction often serves as input (in different formats) to two or more application systems. It is possible to compare records to determine sources of input errors. But with a data base, transactions are entered only once and this facility is lost. Also, erroneous transactions detected during input validation may be destroyed, rather than just tagged and saved. An auditor may feel this to be a loss of useful information.

In the area of file integrity, it has been postulated that someday a company that is wedded to an integrated data base will lose all copies of that data base (through some form of destruction) and will have to go out of business. In fact, there have already been a few near misses. When a case of 'fatal corporate amnesia' does occur, there will be a lot of finger pointing, and the auditors may have to share some of the blame. In fact, a fair case can be made that the auditors will share in the blame if they did not give clear warnings of the risk of corporate amnesia, even if the probability was small.

Accountability for data. With application oriented files, it is fairly easy to identify the person or the group responsible for inputting and maintaining each data field. With a shared-use data base, the responsibility is not so clear. It should be specifically assigned (and probably in writing). In a data base environment, the responsibility might be much stricter than has been true in non-data base systems.

Asset value. The data base, as the repository of much corporate data, will be a valuable asset. It should be subjected to strict security and control procedures, probably to a greater extent than has been true in non-data base environments.

Data definitions. Because of the sharing of data, all data definitions will have to be 'standardized' and coordinated. Where this in fact is done, it will improve the control over the data definitions.

Audit independence. In a data base, the logical structure of the data may be quite different from the physical structure in which it is stored. So the auditor will probably have to use access methods provided by someone else, in order to access the data. Thus, the independence of the auditor will be reduced.

These were only some of the points (but the ones that stood out most in our mind) raised at the two working conferences.

Conclusion

Back in June 1973, our report dealt with 'the cautious path to the data base.' In it, we urged users to take small, easy steps in converting to data base technology. The two working conferences on Data Base Directions, sponsored by the National Bureau of Standards and the Association for Computing Machinery, have stressed the same points, and in much more detail.

Let us conclude with a repetition of the admonitions. Get your data definitions cleaned up *before* you install a data base. Use a data dictionary to get the data definitions in shape. Take it easy for your first DBMS application. Pick a non-trivial but also a non-vital application for your first efforts. Have realistic expectations; a DBMS will not go in overnight and it will not save bundles of money. Treat the data base project like any other system project, using good project management methods. And start *now* to take the steps to make future data base conversions easier—because data and program conversion are a never ending problem.

REFERENCES

1. Proceedings of two working conferences on data base directions. Order from (1) Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. 20402, or from (2) Association for Computing Machinery, 1133 Avenue of the Americas, New York, N.Y. 10036:
   a) *Data Base Directions, The Next Steps*, Cat. No. C13.10:451; price $2.40
   b) *Data Base Directions, The Conversion Problem*, Cat. No. C13.10:500-XX; this report had not been published as we went to press, so that the final serial number (the 'XX') and the price had not been assigned. We have been told that it may be published by late summer.
2. Sibley, E. H., "The impact of database technology on business systems", paper presented at IFIP Congress 77. The proceedings, (titled *Information Processing 77*), are published by North-Holland Publishing Company (P.O. Box 211, Amsterdam, The Netherlands, or 52 Vanderbilt Avenue, New York, N.Y. 10017); 1977; U.S. price $71.50.
3. Gibson, C. F. and R. L. Nolan, "Managing the four stages of EDP growth," *Harvard Business Review* (Reprint Department, Soldiers Field, Boston, Mass. 02163), January-February 1974; p.76-88; price $3.00.

Additional reading

4. *On-line data bases*, Infotech State of the Art Report; order from Infotech International Limited (Nicholson House, Maidenhead, Berkshire SL6 1LD, U.K.); in U.S., order from Auerbach Publishers, Inc. (6560 North Park Drive, Pennsauken, N.J. 08109), price $215.00.
5. Lefkovits, Henry C., *Data Dictionary Systems*, Q.E.D. Information Sciences, Inc. (141 Linden Street, Wellesley, Mass. 02181), 1977, 480 pages, price $85.
6. Lyon, John K., *The Database Administrator*, Wiley-Interscience (605 Third Avenue, New York, N.Y. 10016), 1976, 170 pages.
7. Davis, B., "User experience with database management systems in the U. K.," *Database Journal* (322 St. John Street, London EC1V 4QH, U.K.), Vol. 7, No. 3, 1977, price £4.50 ($15.50).
8. Two articles in *Data Base*, the ACM SIGBDP Newsletter, in Vol. 9, No. 1, Summer 1977; order from ACM (address above); price $3.50:
   a) Benbasat, I. and R. C. Goldstein, 'Data base systems for small business: miracle or mirage?', p. 5-8.
   b) Zornes, J. A. Jr., 'Data base management systems on minicomputers,' p.9-13.
9. SEED is a low-cost (below $10,000), transportable, CODASYL-type data base management system implemented to run on a number of today's mini and maxi systems; for information, write International Data Base Systems, Inc., 3700 Market Street, Philadelphia, Penna. 19104. A version has been developed to run on Z-80 (S-100) micro computers, and is being marketed by Technical Design Labs of Princeton, New Jersey, for $1,250.
10. McFadden, F. R. and J. D. Suver, "Costs and benefits of a data base system," *Harvard Business Review* (address above), Jan-Feb 1978, p. 131-139.

You probably have been reading about the rapidly growing market for 'hobbyist' micro computers. Rather impressive computers are now being sold for under $1,000; a system costing about $5,000 can begin to handle the data processing workload of a small business, or a small unit of a big business. The result of these developments will be, we think, a rapid growth in the use of 'personal' computers in business. There are some problems, of course, with system software and application programs being at the top of the list. Next month, we will discuss the present status and the near future likelihoods for 'personal' computers in business.

SUBJECTS COVERED BY EDP ANALYZER IN PRIOR YEARS

1975 (Volume 13)

Number

1. Progress Toward International Data Networks
2. Soon: Public Packet Switched Networks
3. The Internal Auditor and the Computer
4. Improvements in Man/Machine Interfacing
5. "Are We Doing the Right Things?"
6. "Are We Doing Things Right?"
7. "Do We Have the Right Resources?"
8. The Benefits of Standard Practices
9. Progress Toward Easier Programming
10. The New Interactive Search Systems
11. The Debate on Information Privacy: Part 1
12. The Debate on Information Privacy: Part 2

1976 (Volume 14)

Number

1. Planning for Multi-national Data Processing
2. Staff Training on the Multi-national Scene
3. Professionalism: Coming or Not?
4. Integrity and Security of Personal Data
5. APL and Decision Support Systems
6. Distributed Data Systems
7. Network Structures for Distributed Systems
8. Bringing Women into Computing Management
9. Project Management Systems
10. Distributed Systems and the End User
11. Recovery in Data Base Systems
12. Toward the Better Management of Data

1977 (Volume 15)

Number

1. The Arrival of Common Systems
2. Word Processing: Part 1
3. Word Processing: Part 2
4. Computer Message Systems
5. Computer Services for Small Sites
6. The Importance of EDP Audit and Control
7. Getting the Requirements Right
8. Managing Staff Retention and Turnover
9. Making Use of Remote Computing Services
10. The Impact of Corporate EFT
11. Using Some New Programming Techniques
12. Progress in Project Management

1978 (Volume 16)

Number

1. Installing a Data Dictionary
2. Progress in Software Engineering: Part 1
3. Progress in Software Engineering: Part 2
4. The Debate on Trans-border Data Flows
5. Planning for DBMS Conversions

(List of subjects prior to 1975 sent upon request)

PRICE SCHEDULE

The annual subscription price for EDP ANALYZER is $48. The two year price is $88 and the three year price is $120; postpaid surface delivery to the U.S., Canada, and Mexico. (Optional air mail delivery to Canada and Mexico available at extra cost.)

Subscriptions to other countries are: One year $60, two years, $112, and three years $156. These prices include AIR MAIL postage. All prices in U.S. dollars.

Attractive binders for holding 12 issues of EDP ANALYZER are available at $6.25. Californians please add 38¢ sales tax.

Because of the continuing demand for back issues, all previous reports are available. Price: $6 each (for U.S., Canada, and Mexico), and $7 elsewhere; includes air mail postage.

Reduced rates are in effect for multiple subscriptions and for multiple copies of back issues. Please write for rates.

Subscription agency orders limited to single copy, one-, two-, and three-year subscriptions only.

Send your order and check to:
EDP ANALYZER
Subscription Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-3233

Send editorial correspondence to:
EDP ANALYZER
Editorial Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-5900

Name____________________

Company ____________________

Address ____________________

City, State, ZIP Code____________________