STANDARDS

The objective of the Standards working panel was to recommend standards that a manager should consider when converting data from present sources (manual, semi-automated or automated) to a DBMS. The panel considered administrative guidelines as well as technical standards, since both are essential to an effective conversion process.

A "standard" is a consensus or common practice sometimes established by authority derived from a desire to reduce arbitrary variety for economic reasons. A standard is one method of insuring compatibility by using accepted conventions.

Several types of standards exist. For instance, a de facto standard is a practice accepted through common usage, although it has not been subjected to official standardization. On the other hand, standards have also been approved by an authorized official for use within a particular organization.

For example, Federal Information Standards Processing (FIPS) are those that have been approved by the federal standards process. American National Standards are those standards that have been approved for voluntary use by the American National Standards Institute (ANSI), a body made up of participants from government, industry and academe. International standards are those that have been approved by the International Organization for Standardization (ISO) for voluntary use by member nations and international organizations.

In the case of FIPS, ANSI and ISO, a formal mechanism has been established to propose, review, approve and validate standards. While ANSI and ISO standards are voluntary for all participants, the FIPS standards are mandatory within the federal government. For mandatory standards to be effective, an enforcement and validation mechanism is necessary.

As yet, no formal DBMS standards exist. Many groups

The panel chairman was Milt Bryce of M. Bryce & Associates, Inc. Other participants were Robert Bemer, Honeywell Information Systems; Don Branch, Advisory Bureau for Computing (Canada); Jean Bryce, M. Bryce & Associates, Inc.; Elizabeth Fong, National Bureau of Standards; Al Gaboriault, Sperry Univac; Anthony Klug, University of Wisconsin; Henry C. Lefkovits, H. C. Lefkovits & Associates, Inc.; C. H. Rutledge, Marathon Oil Company; Philip Shaw, IBM Corporation; Jay P. Thomas, Allied Van Lines; Ewart Willey, Prudential Assurance Co., and Maj. Jerry Winkler, U.S. Air Force.

concerned about standardization are actively working in this area. The committee felt that a review of the various standards activities would be worthwhile for the reader.

Overall, the recommendations of the standards section of Data Base Directions I have two prominent characteristics: They are still valid and they have not been implemented. Specifically, the committee at that time recommended the undertaking of more international activity in data base standardization.

Despite considerable interest in data base, standards are only now beginning to occur at the national or international level. Nothing comparable to the International Telegraph and Telephone Consultative Committee (CCITT) has been formed. Little progress has been made, either, within the International Organization for Standardization (ISO).

This lack of progress exists in spite of the various admonitions from governmental operating bodies, such as Congressman Jack Brooks and the House Operations Committee in the United States, European Economic Community Committees and so on. In effect, no firm standards efforts have been undertaken. [Editor's note: Subsequent to the preparation of this report but prior to publication, ANSI initiated several DBMS activities.]

Voluntary standards bodies such as ECMA and American National Standards Institute/Standards Planning and Requirements Committee (ANSI/SPARC et al.) have been slow to agree upon firm recommendations for standards or interim action. The technical societies have not been able to recommend agreed upon and coherent roadmaps to the eventual realization of data base standards. The art and/or science of decision-making using data base data does not seem to have advanced.

Data collection procedures may have improved since the days of manual actions with paper, but the science of decision-making based on such data does not seem to have made significant progress. For example, Data Base Directions I recommended the more accurate usage of terms. It suggested that "CDBS" (Computer Data Base System) be used instead of "DBMS," since organizations have always used data base. This suggestion underscores the important new factors to be considered by the reader of this new term. However, the suggestion has not had acceptance. On the positive side of the data base standards issue, substantial additions have been made to the literature, and we anticipate eventual, improved understanding of these issues.

Many practical data base conversions have been undertaken successfully—both from conventional systems to a computer data base system, or from one computer data base system to another computer data base system. A body of evidence indicating the economic superiority for the data base approach is accruing. The reader will read of such evidence in other sections of this report.

The Standards Planning and Requirements Committee (SPARC) of ANSI/X3 (Computers and Information Processing) established a study group in 1972 which was chartered to investigate the potential for standardization in the area of DBMS. During early deliberations, the study group concluded that the proper subjects for DBMS standardization were interfaces within a DBMS.

It was then necessary to define a framework for DBMS. In 1975, the study group produced an interim report[3] which described such a framework. A revised report has also been published recently[4] which represents the study group's most important ideas relevant to standardization.

The framework represents three kinds of objects: processing functions, person-roles and interfaces between these. The objects were divided into three basic levels: the internal, the conceptual and the external. The internal level is oriented toward the most efficient use of the computing facility. The conceptual level contains a logical representation of the persons, places, things and their relations of interest to the organization using the DBMS. The external level contains different views of the data base which various

applications need to do their information processing.

The three levels are associated through explicit mappings. Mappings, through their ability to isolate changes to one level, facilitate data base conversion and form a basis for orderly standardization of DBMS. In addition, the ANSI/SPARC framework defines three administrator roles, one for each of the three levels in the framework. These distinct roles are also potential aids in standardization in that responsibilities are clearly delineated.

Although the study group itself did not recommend to its parent, ANSI/X3/SPARC, action for or against standardization of any DBMS component, SPARC recommended to ANSI/X3 that standardization efforts begin on a CODASYL data definition language and on CODASYL additions to FORTRAN and COBOL for data manipulation. These efforts have been initiated.

Through the work of several technical committees coordinated by an executive committee, CODASYL has continued to develop specifications for a data definition language and data manipulation languages for incorporation into the existing COBOL and FORTRAN languages.

CODASYL has addressed a number of technical problems including concurrency and device independence and has endeavored, through task groups of its COBOL and DDL commitees, to resolve inconsistencies between the language specifications produced by the two committees. CODASYL has authorized the publication of three journals of development (DDL JOD, COBOL JOD, FORTRAN JOD) containing the respective language specifications to facilitate the work of ANSI in defining standards for the three languages.

Standardization can provide many benefits to vendors and users involved in conversion tasks. In this sense, users mean the applications development people who will actually perform the conversion. The benefits can take many forms. Among the benefits the committee felt could be realized were:

- Defining in a standard way the logical relationships of data would in particular insure that diverse hardware and/or software implementations could successfully address the same logical data. This could minimize hardware dependencies and allow the users to progress to both different and improved technology with a minimum expenditure of both time and money. Portability would provide protection against loss of the usually sizeable investment required by the user to establish a data base.
- Standards could provide a basis upon which educational institutions and corporations could structure their data processing training programs to satisfy the needs of their personnel. In effect, standards could reduce the time and expense of retraining data processing staffs.
- Standard terms and definitions would enhance communication among users and/or vendors. As it now stands, these terms are being haphazardly developed and becoming widely used with little common understanding.
- Standards would provide product specifications that could be met by all vendors. End-users would have, therefore, more precise product specifications and could exercise greater impact on what products would be manufactured to satisfy their needs.
- The standardization machinery would enable the user community to be involved in the evolution of the data base technology. In this respect, it should be self-evident that the user would have a far greater interest than the vendor in insuring that the standards applicable to data base reflect the need to minimize the problem of conversion.

This section examines the software components of the data base environment which the committee felt could be standardized. The rationale for standardization is based on four different conversion scenarios and the components which, if standardized, the committee felt would facilitate conversion. In all cases, technical feasibility of the standard has not been considered.

The term "data base environment" is used to describe an environment in which the many parts making up an organization's data base would be available for shared use and coordination. This approach is less confusing than using the term "DBMS environment," since many installations are merely using the DBMS software as a substitute for traditional access methods. In this situation, the environment has not changed; it is merely a change of access method.

The term DBMS will be used to describe the software used to manipulate the computer-resident portion of the data base. The conversion scenarios considered were:

- Moving from a nondata base environment to a data base environment using a standard DBMS.
- Moving from a standard DBMS to the same standard DBMS within a different hardware environment.
- Moving from a Type A standard DBMS to a different Type B standard DBMS.
- Moving from a standard DBMS to a nonstandard DBMS environment.

In moving from the nondata base environment to the standard data base environment, several points can be identified which can be facilitated by the availability of standard tools or methods.

To determine the organization's data requirements, a data flow analysis should be undertaken. To facilitate this analysis, a standard approach is needed. It should include a complete description of the data and where it is used, including all input and output documents, both manual and automated processes.

Although it may be premature to consider a standard for a dictionary/directory system, this is an important tool for definition analysis and data flow analysis and to facilitate access to this information. A more detailed discussion of the possible capabilities is presented elsewhere in the report.

The data management function, an organizational group concerned with overall responsibility for the management of data, both computerized and noncomputerized, should have a specific functional description.

Data base administrator is a position holding primary responsibility for the overall accuracy, timeliness and availability of the corporate data through direct control over the data dictionary/directory system. This position located within the data management function should have clearly understood responsibilities.

To facilitate the conversion effort, specific conventions are needed for placing data into data bases initially.

Conversion to the data base environment can be accomplished by either moving current applications to the new environment with no major changes in process or by creating new applications. Both approaches use a data manipulation language (DML) to access the data base from a user-oriented view of the data base.

In a DBMS to DBMS conversion where each DBMS is the same standard and only the hardware changes, the steps in Scenario 1 were presumably implemented already. However, additional problems occur.

Moving the data in the source computer data base to the target computer data base requires a translation of the data from physical structure of the source computer to the physical structure of the target computer. A needed standard facility would be a method of unloading the physical data to a standard interchange format, in display format representation and then loading that data into physical structure of the target computer.

Restatement of the application programs reflecting changes to the description of data are necessary as well as changes to the application programs. Standard data description and data manipulation functions are needed in addition to those stated in Scenario 1.

Conversion between standard versions of different DBMS assumes both standard DBMS incorporated the points in Scenario 1. No additional points beyond those described above are needed, but significant differences in the details of the data models may occur.

Standard DBMS to nonstandard DBMS conversion is similar to Scenario 3 because there would be no reason to make this step unless no standard DBMS satisfied information requirements.

Additional computer components considered and mentioned as desirable, but for which standards may be premature or which may not affect the conversion effort are standard end-user facilities (a means whereby a noncomputer professional can communicate with the data base), network interfaces, restart functions, security functions, communication facilities and command languages (operating systems).

Many nonsoftware components are involved in any conversion process. Those considered by the committee include:

Administrative/management Procedures. The committee identified no standards which could be applied to these procedures during conversion. However, the development of guidelines and checklists based upon experience would be very useful.

Verification of Compliance With Standards. The committee believed there is a need for a standard approach to validating compliance with DBMS standards. An important component of such a validation would be having a set of common procedures for measuring compliance.

Two areas should be considered immediately for standardization if progress toward successful conversions is to be made. These two areas are development of a standard DBMS and development of a generalized dictionary/directory system.

A DBMS standard would benefit government and industry in their conversion efforts, and the committee encourages progress toward a DBMS standard. The work of CODASYL group, with some possible modifications, appears to offer a first step toward the complete specification of a DBMS, although some committee members felt several successful commercial DBMS may become de facto standards. The committee also believed that consideration should be given to the idea that the standard should not inhibit any future technological developments in hardware and software.

The use of a dictionary/directory facility is highly desirable as a tool in any conversion process, whether from manual or conventional file to a data base environment or from one DBMS to another DBMS. The advantages for the use of a dictionary facility will be gained primarily in the following two areas:

- It will allow the proper assessment and definition of the organization's data to be undertaken or to be made prior to the start of a project.
- It will provide an essential management mechanism to control the actual development of the new system.

Since, in this sense, the dictionary facility chronologically precedes the DBMS selected for the conversion, it appears likely that the direct support of the dictionary facility itself should not depend upon the target DBMS or, for that matter, upon any particular DBMS. In any case, the dictionary facility must be able to support descriptive facilities for the data in a logical mode.

Another useful capability is that of being able to produce

processable data descriptions in the target DBMS environment. This does not, however, imply that this target DBMS must be used in the implementation of the dictionary facility, but rather that a suitable (possibly manual) interface to this DBMS be made available. It would be highly desirable, if not required, that all DBMS had the same logical interface. Then the dictionary could interface with multiple DBMS at the same time.

In discussing the architectural placement of the dictionary, the committee concluded this facility should not exist as an adjunct to any DBMS, but as a separate facility that interfaces to data in conventional files, as well as to data being managed by one or more DMBS. A distinction needs to be made in the manner in which the dictionary facility is invoked.

In the case of a free-standing dictionary, this facility is invoked under user control. As such, it is a user option if, and when, the dictionary is to be invoked in the execution of any process. In the case of an integrated dictionary, which means the data dictionary is directly available to the DBMS, this facility might be invoked automatically by the system itself in the execution of any process, thus assuring synchronization of the process with the information contained in the dictionary. The feasibility of this approach depends upon the overhead burden it may generate.

Since the use of some of the currently available dictionary facilities is believed to be of considerable value in the attainment of these and other goals, it is envisioned that a more general facility, called a generalized dictionary/directory system, would be more effective. In brief, a GD/DS is an information system in and of itself whose subject matter is all the information about the enterprise on the following classes of entities: data components, processing components, people and organizational components and events.

Attributes to be included in the data dictionary are those about the entities themselves, the relationship that exists among the entities, as well as the context in which these relationships exist. For example, among the attributes would be the relationship between systems and both automated and manual procedures. The generalized data dictionary/directory should contain not only the logical attributes and relationships between the entities described, but also attributes associated with the physical location of the entities.

In summary, the committee strongly believes that many areas for standardization exist within the conversion process and the data base environment in general. The committee notes with frustration the lack of overall progress since the first Data Base Directions workshop.

The committee noted that in actual use, DBMS are often misapplied. It is possible, however, that this misuse relates to the historical problems in the environments when DBMS are introduced. For example, most organizations seem to be obtaining DBMS as their initial introduction into the data base environment and then find a need for a dictionary/directory which is then introduced as a secondary action. Only when both of these steps do not really solve the problem does the organization look at its total organization and address the fundamental work necessary. This fundamental work includes documenting data, systems and processes to insure their more effective operation in this new DBMS environment.

The committee observed that the installation process actually should be reversed, with the data dictionary introduced first. All of the data within the environment and the systems that use them should be carefully documented and defined (including all the developmental shortcuts that had been taken in the past) and then entered into a generalized dictionary/directory. Once this is accomplished, the appropriate data bases to support their various information systems in the organization can be built.

REFERENCES

The following references are included as part of the Standards panel report as a convenience to the readers. Some of the material included was not used directly by the committee, but did influence the thinking of some of the participants.

1. Sibley, Edgar H., "Standardization and Database Systems," IFSM TR No. 23, University of Maryland, College Park, Maryland 20742, July 1977
2. Berg, John L., (editor), "Data Base Directions—the Next Step," Proceedings of the Workshop of NBS and ACM held at Fort Lauderdale, Florida, October 29–31, 1975, National Bureau of Standards Special Publication 451
3. ANSI/X3/SPARC Study on Data Base Management Systems, "Interim Report, 75-02-08," published as Vol. 7, No. 2 1975 of *FDT*, the Bulletin of ACM-SIGMOD
4. Tsichritzis, D. and Klug, A. (editors), "The ANSI/X3/SPARC DBMS Framework: Report of the Study Group on Data Base Management Systems," AFIPS Press, 1977
5. CODASYL Data Description Language Committee, "Data Description Language," *Journal of Development*, January 1978
6. Leon-Hong, B. & Marron, B., "A Technical Profile of Seven Data Element Dictionary/Directory Systems," NBS Special Publication 500-3, February 1977
7. Lefkovits, Henry C., "Data Dictionary Systems," *Q.E.D. Information Sciences*, 1977
8. CODASYL Data Description Language Committee, "DDL Journal of Development," June 1973, NBS Handbook 113, January 1974, Page 155.
9. "Data Dictionary/Directory Evaluation Criteria," M. Bryce & Associates, Inc., Cincinnati, Ohio
10. "Installing a Data Dictionary," *EDP Analyzer*, Vol. 16, No. 1, January 1978.