DATABASE MANAGEMENT SYSTEMS:

A REVIEW WITH REFERENCES

Ben Shneiderman

Computer Science Department

Indiana University

Bloomington, Indiana 47401

TECHNICAL REPORT NO. 51

DATABASE MANAGEMENT SYSTEMS:

A REVIEW WITH REFERENCES

BEN SHNEIDERMAN

MAY, 1976

This material forms the basis for a volume of the American Federation of Information Processing Societies (AFIPS) and National Computer Conference Board Technical Series entitled Database Management (1972-1975), Ben Shneiderman, Editor. The The sixteen papers and the introductory material will be published late in 1976 by AFIPS Press, Montvale, New Jersey.

DATABASE MANAGEMENT SYSTEMS

Table of Contents

Ben Shneiderman

Introduction

The remarkably pervasive conversion of data processing to the unified database approach during the past decade has made research and implementation in database management systems the leading growth area in computer software. Keeping up with the rapid developments provides a challenge to management, technical and research personnel in industry and in the university community. This volume is an attempt to survey progress in database management systems by selecting papers from seven important conferences: the Fall and Spring Joint Computer Conferences of 1972; the National Computer Conferences of 1973, 1974 and 1975; and the USA-Japan Computer Conferences of 1972 and 1975. The sixteen papers in this volume make up only 106 pages from the 6785 pages of technical material presented at the conferences. The difficulty of selecting the best papers was compounded by the attempt to provide a representative sample of current fields of interest and opposing opinions. Many good papers could not be included because of space limitations and areas such as hardware designs for database management, application experiences, analysis of database algorithms, physical file structure techniques and theoretical approaches were excluded so as to provide better coverage of the topics that have been selected. The editor regrets these omissions and accepts full responsibility for the choice of papers.

This volume should be useful as a survey of the field for management and technical staff members. Researchers may be familiar with topics in their area of specialization, but could benefit from a broader perspective. The references and overviews for each section are designed to organize the material and give guidance for further reading. An effort has been made to limit the references to widely available sources and offer a diversity of insights into current developments in database management systems. Occasionally, more obscure sources are used because of the importance of the material.

The following brief review provides an impression of some of the papers presented at the conference but not included or referenced in later sections. Two excellent papers which discuss hardware designed for database management are

RAP - An associative processor for data base management, E.A. Ozkarahan, S.A. Schuster, K.C. Smith, NCC, 1975.

The datacomputer - A network data utility, Thomas Marill, Dale Stern, NCC, 1975.

A second vital area, the application of computerized information systems to health care, is covered in numerous papers including

Automated information-handling in pharmacology research, W. Raub, SJCC, 1972.

A public health data system, J.C. Peck, F.M. Crowder, NCC, 1974.

Automated patient record summaries for health care auditing, R. Chalice, NCC, 1974.

An integrated health care information processing and retrieval system, K.C. O'Kane, R.J. Hildebrandt, NCC, 1974.

Development and implementation of a medical/management information system at the Harvard Community Health Plan, N. Justice, G.O. Barnett, R. Lurie, W. Cass, NCC, 1974.

Structured organization of clinical data bases, Gio Wiederhold, J.F. Fries, S. Weyl, NCC, 1975.

Information processing needs and practices of clinical investigators-Survey results, N.A. Palley, G.F. Groner, NCC, 1975.

The Candadian Medical Association information base - A beginning of operational systems in Canada, J.F. Brandejs, NCC, 1975.

A comparative evaluation of automated medical history systems, Ephraim F. McLean, S.V. Foote, NCC, 1975.

Clinical information system (CIS) for ambulatory care, C. McDonald, B. Bhargava, D. Jervis, NCC, 1975.

Physical file design techniques and algorithms are reviewed by

Compression parsing of computer file data, W.D. Frazier, USA-Japan, 1972.

Data management with variable structure and rapid access, D.K. Hsiao, F. Manola, USA-Japan, 1972.

Representation of sets on mass storage devices for information retrieval systems, S.T. Byrom, W.T. Hardgrave, NCC, 1973.

Optimal file allocation in multi-level storage systems, P.P.S. Chen, NCC, 1973.

A classification of compression methods and their usefulness in a large data processing center, D. Gottlieb, S.A. Hagerth, P.G.H. Lehot, H.S. Rabinowitz, NCC, 1975.

Weight-balanced trees, J.L. Baer, NCC, 1975.

Descriptions of specific implementations can be found in

HOMLIST - A computerized real estate information retrieval system, D.J. Simon, B.L. Bateman, SJCC, 1972.

SIMS - An integrated user-oriented information system, M.E. Ellis, W. Katke, J.R. Okson, S. Yang, FJCC, 1972.

Design and development of a generalized data base management system - from the FORIMS implementation point of view, K. Kohri, USA-Japan, 1972.

On the data base management system SELDAM, S. Mimura, H. Sakai, R. Yoshikawa, USA-Japan, 1972.

The COMRADE data management system, S. Willner, A. Bandurski, W. Gorman, M. Wallace, NCC, 1973.

COMRADE data management system storage and retrieval techniques, A. Bandurski, M. Wallace, NCC, 1973.

DUCHESS - A high level information sytem, B.J. Taylor, S.C. Lloyd, NCC, 1974.

TOOL-IR: An on-line information retrieval system at an inter-university computer center, T. Yamamoto, M. Negishi, M. Ushimaru, Y. Tozawa, K. Okabe, S. Fujiwara, USA-Japan, 1975.

A notable successful application is the Ohio College Library Center Network for cataloguing books described by Kilgour in Computerized Library Networks (USA-Japan, 1975).
A number of management related topics are covered in

Where do we stand in implementing information systems?, J.C. Emery, SJCC, 1972.

MIS technology - A view of the future, C. Kriebel, SJCC, 1972.

Selective security capabilities in ASAP - A file management system, R. Conway, SJCC, 1972.

User/system interface within the context of an integrated corporate data base, G. Altshuler, B. Plagman, NCC, 1974.

Data bases - uncontrollable or uncontrolled?, C.M. Traver, NCC, 1974.

Finally, some important papers which do not readily fit into any category

A graphics and information retrieval supervisor for simulators, J.L. Parker, SJCC, 1972.

Privacy and security in data bank systems - Measures, costs, and protector intruder interactions, R. Turn, N.Z. Shapiro, FJCC, 1972.

A model for a generalized data access method, R.L. Frank, K. Yamaguchi, NCC, 1974.

Evaluating inter-entry retrieval expressions in a relational data base management system, J.B. Rothnie, Jr., NCC, 1975.

Optimizing distributed data bases - A framework for research, K.D. Levin, H.L. Morgan, NCC, 1975.

The database management revolution: A historical review

The history of database systems provides a fascinating study of the personal and institutional forces which produce progress in scientific and technical development. Thomas Kuhn's classic treatise, The Structure of Scientific Revolutions (Univ. of Chicago Press, 1962), reviews the progress of "normal science" as it leads to revolutionary change. Kuhn argues that the bulk of research involves the clarification and support of contemporary theories while providing the counterexamples for a new revolutionary reorganization of research paradigms. Other insights to the progress of scientific development can be obtained from Griffith and Mullins' "Coherent Social Groups in Scientific Change" (Science, Sept. 15, 1972, Vol. 177, pp. 959-964) which reviews recent scientific research revolutions and patterns of interaction among researchers.

In database management systems, there are two major conflicting ideological outlooks which have distinct origins but a similar chronological development. The data-structure-set model (often called CODASYL-DBTG or network approach) has the soft-spoken, bow-tied Charles Bachman as its ideological leader. The engaging E.F. Codd, with his trim, RAF pilot features and mustache, is the principal advocate of the relational model.

Codd's seminal paper, "A relational model of large shared data banks", probably the single most important paper in database management, was submitted in September 1969 and appeared in the June 1970 issue of the Communications of the ACM. Although Codd references Childs (1968) and Levien and Maron (1967), his idea was a revolutionary departure from previous work. In later articles, Codd acknowledges Ash and Sibley (1968) and Feldman and Rovner (1969), but these papers do not seem to have had any direct influence on him. Codd's paper, which challenged current docrines, was a detailed presentation requiring ten pages of the large format CACM, but a half dozen other papers were still necessary to deal with ancillary issues.

Codd created a system for manipulating data which completely avoids any discussion of storage structures. A simple tabular representation was complemented by rigorous mathematical foundations which appealed to a wide variety of users. The abstraction to logical structures enabled higher level languages to succinctly describe complex data manipulation operations with precision and elegant brevity.

The enthusiastic response to Codd's ideas prompted hundreds of papers based on the relational approach. These have appeared in the journal literature, at the Workshops of the ACM Special Interest Group on File Description and Translation (SIGFIDET, renamed in 1974 to Special Interest Group on the Management of Data - SIGMOD) and other conferences. The most intense development work occurred at the IBM San Jose, California Research Labs where Codd and his associates refined their ideas, developed query languages and implemented relational systems. An important contributor to the relationally based query facilities, SQUARE and SEQUEL, was Raymond F. Boyce whose ascending career was tragically cut short by a stroke while attending a meeting at San Jose in June 1974.

The IBM Labs in California continue to be the most prolific center for research in database management systems. For his efforts, Ted Codd has been made an IBM Fellow and continues his research on a natural language front-ends for relational systems.

On the East Coast, the IBM Labs in Yorktown Heights, New York have pursued the intriguing two-dimensional query-by-example facility for manipulating and retrieving relations. The assertive Moshe Zloof who developed query-by-example is intensely involved in expanding the capabilities of his positional notation and tuning the experimental relational system that is being developed at Yorktown Heights. Michael Senko, who left the San Jose Labs, pursues the Data Independent Accessing Model (DIAM) as an alternative conceptual framework. This level structured approach addresses end-user facilities, intermediate levels and physical device issues in a unified manner.

The second major approach to database management, the data-structure-set model, can be traced from the early work of Charles W. Bachman (1965) and George G. Dodd (1966). Bachman's ideas were the basis of the Integrated Data Store produced by General Electric and later distributed by Honeywell. Dodd, who worked for General Motors Research Labs, and Bachman were early members of the voluntary, industry supported Conference on Data Systems Languages (CODASYL) Data Base Task Group (DBTG) which began meeting in 1965. Their interim report in October 1969 and the final report in April 1971 detailed a new approach to database management systems which has become the basis of eight or more commercially available systems.

Committees are rarely the source of revolutionary developments, but the DBTG Report was successful in moving the industry from the classic COBOL file approach to the unified database approach. Although very few papers were published in the technical literature, the 269 report had an effective distribution network due to the considerable influence of the 20 to 25 task group members who were employed by some of the most important manufacturers and users of large computer systems. Unfortunately, IBM, the industry leader, had many reservations about the report. Its decision not to implement the DBTG ideas was a severe blow. Although IBM had serious technical questions (Engles, 1972), they were also concerned about the growing number of Information Management System (IMS) users who would have been upset by a major change of direction. Revisions to the DBTG Report were made by the Data Description Language Committee whose 1973 report contains a complete history of the project.

The conceptual basis for the data-structure-set model is the use of arrows to show relationships between owner records and groups of member records. This pictorial approach, supplemented by data structure diagrams (1969), is easy to comprehend but is criticized as being more closely tied to physical storage structures.

Bachman's accomplishments were acknowledged through the Turing Award of 1973 which is given to major contributors in computer science. His Turing lecture (1973) entitled "The programmer as navigator" was an appealing review of the revolution in database management which he likens to the Copernican revolution. Bachman envisions a major change from a computer-centered view of database management to a data-centered view, paralleling the transformation from geo-centric to helio-centric perceptions of the universe. In the old paradigm, the computer was central and data was treated as merely passing through the machine for processing. Now, Bachman argues, we see the database as the fundamental resource and the computer programs are merely tools for searching through the vast sea of data. Within this model, the programmer is the navigator who directs the program along complex pathways to insular data nodes. The elevation of data above the programs which process data is widely accepted, but the programmer-as-navigator approach is questioned by supporters of the relational model who feel that their approach allows automatic-pilot processing.

As the prime mover of the data-structure-set concept, Charles Bachman was asked to participate in a debate against E.F. Codd, representing the relational model. The debate on the opposing data models was held at the SIGMOD Workshop in Ann Arbor, Michigan in May 1974, but the edited transcript (1975) did not appear until almost a year later because of disagreement over the contents. Bachman calmly argued the equivalence of the two models but emphasized that ten years of experience had demonstrated the value of the data-structure-set approach. He felt that although the relational model had a fine theoretical basis, it was untried in realistic environments. Codd responded by criticizing the data-structure-set approach and stressing the simplicity of relational concepts. The numerous questions and provocative discussions following the major presentation further probed the differences, but no resolution was achieved. This dramatic debate was one of the few occasions in which leaders of two distinct technical research outlooks confronted each other in a public forum.

It is unlikely that one data model will dominate the field. The variety of application projects and the differences in the intellectual backgrounds of users will prohibit the universal adoption of a single approach. Alternative data models, query languages and procedural languages are necessary to cope with the diversity of database environments.

Other sources of development research in database management systems have included commercial software manufacturers such as Cullinane Corp., MRI Systems, Software AG, Cincom, and the major hardware manufacturers. University contributions include MIT which produced an early relational implementation. The University of Toronto and the University of California at Berkeley are the homes of two active projects to develop full-scale relational systems, languages, supporting software and theoretical issues. Other important centers include the Universities of Florida, Maryland, Michigan, Minnesota, Ohio, Pennsylvania and Utah.

For all of the progress of the last decade, much remains to be done. The enormous expansion of batch production and online usage of databases will produce an even greater demand for simplified querying and efficient processing. The concurrency, privacy, security and integrity problems will have to be dealt with rigorously. Networking, distributed databases, and special purpose processors will offer new opportunities for systems architects and software designers. Finally, the greatest challenge will be to managers and database administrators. They will have to combine sensitivity in directing the work of their staffs and an understanding of complex technical issues, while keeping in mind the personal and societal effects of the database management systems they develop.

Ash, W.L., and Sibley, E.H., "TRAMP: A Relational Memory with Deductive Capabilities", Proc. ACM 23rd Natl. Conf., Las Vegas, Nevada, August 27-29, 1968 pp. 143-156.

Bachman, C.W., Integrated Data Store, DPMA Quarterly (January 1965).

Bachman, C.W., "Data Structure Diagrams," Data Base 1, 2, Summer 1969, Quarterly Newsletter of ACM SIGBDP.

Bachman, C.W., The programmer as navigator, CACM 16, 11 (November 1973) pp. 653-658.

Childs, D.L., "Feasibility of a set-theoretical data structure - a general structure based on a reconstituted definition of relation", Proc. IFIP Congress, North-Holland, Amsterdam, pp. 162-172, 1968.

CODASYL, Data base task group report, ACM, New York, 1971.

CODASYL Data Description Language, Journal of Development, June 1973, NBS Handbook 113, U.S. Dept. of Commerce.

Codd, E.F., A relational model of data for large shared data banks, CACM 13, 6 (June 1970), 377-387.

Dodd, G.G., APL - A language for associative data handling in PL/I, Proc. Fall Joint Computer Conf., AFIPS Press, Montvale, NJ, 1966, pp. 677-684.

Engles, R.W., An analysis of the April 1971 Data Base Task Group Report, Proc. 1971 ACM-SIGFIDET Workshop on Data Description, Access and Control, pp. 93-112, 1971.

Feldman, J.A. and Rovner, P.D., "An Algol-Based Associative Language", CACM 12 (August 1969).

Levien, R.F., and Maron, M.E., "A Computer System for Inference Execution and Data Retrieval", CACM 10 (November 1967).

Rustin, R. (ed.), Data Models: Data-structure-set versus relational, Proc. ACM SIGMOD Workshop on Data Description, Access and Control, May 1-3, 1975, ACM, 1975.

Management and utilization perspectives

Management, in the context of this book, refers to supervision of the selection, design, development and administration of database management systems. Those whose needs can be satisfied by the acquisition of a generalized system produced by others, will benefit from the hundreds of person-years invested in the development and testing of the commercially available product. Internal development of database management systems is justifiable only in situations where the application is highly specialized or if machine efficiency is an overriding issue. Management must beware of the desire of the technical staff to embark on a technical adventure which may be beyond the ability of the programming staff. The uneven history of internal database management system implementations indicates the complexity of the task. For a good analysis of some failures see Morgan and Soden (1973).

If a decision has been made to acquire a commercially available generalized database management system, the confusing set of alternatives presents a real challenge even to the most sophisticated administrator. Even if benchmark comparisons can be arranged, a non-trivial project in itself, other factors besides machine efficiency are often more critical. The cost of the system will be only a fraction of the total investment, so it usually plays a minor role. The major issues should be longer term questions such as the ease of use, ease of training, availability of experienced programmers, reliability of the product, reliability of the developer in providing assistance and improvements, flexibility of the system to accommodate new features and modifications of the data model, availability of utilities, telecommunications features, query facilities and security control.

Having chosen a specific system, the administrator must carefully monitor the implementation and development of the numerous applications programs. The integrated or unified database approach which combines a large number of independent file management programs is more than a technical problem. Unifying the information processing operations of a large organization requires numerous changes which produce confrontation among personnel who recognize that the organizational power structure is being revised. Failure to deal with these issues is possibly the greatest cause of failure for database management systems. If middle level management is not involved in the design and development process they will be uncooperative or hostile. Without widespread active cooperation, the system will fail. High level management must also be involved and committed to the success of the database system. Their participation is essential since they have the ultimate decision power and can be influential in gaining the cooperation of middle level management. The successful administrator will be the one who avoids technical showmanship and puffery and honestly solicits the involvement of all personnel by keeping them advised of progress, providing training, and asking for advice and participation.

Since a database management system must ultimately provide service for an end user, careful attention should be given to their needs. The users may be external customers of the organization, operational management, or high level management or a combination of these. Interviews with the users should be done during the design stage and also when the first phases of implementation are completed to derive feedback for modification of the system. The database management system must ultimately serve the needs of the users - if the system does not, then it will fail no matter how clever and sophisticated the technical aspects.

The management function has two components: the database administrator aspects and the application development tasks. The database administrator is responsible for the data and for providing access to the data by way of the database management system. The database administrator, who may supervise several staff members, develops the schema description (a complete description of all the data items) and is responsible for the physical storage aspects. Efficient and reliable performance for the community of users is the goal, even if some applications are implemented in less than optimal fashion. Privacy, security, integrity and concurrency must be provided since the responsibility for the data resides in the office of the database administrator. For a more thorough discussion, see Meltzer (1975), Nolan (1973a, 1973 b, 1974), and Martin (1976).

The application development teams meet with the database administrator to implement subschema descriptions (or views) of the data items necessary for a particular project. The database administrator may modify the schema description to accommodate new applications, but this should be optional if there is true data independence. Schema descriptions are modified to include new data items or to improve performance.

Schema and subschema design is a challenging problem since performance and ease of use are severely affected by improper design. This topic is discussed in Date (1975), Martin (1975), Hubbard and Raver (1975), Gerritsen (1975), Smith and Smith (1976) and Curtice (1974, 1975). The evolution of stored-data descriptions, and their relationship to conceptual structures for data, is the topic of Bachman's discussion (1975). He reviews some trends which will influence developments in the next decade.

The impact of database management systems is not limited to the organizations which adopt them for their information processing needs. The users of database and anyone dealing with the organization may be confronted with new freedom and power as well as constraints and burdens. Database management specialists must be keenly aware of the power to do good and evil that is placed in their hands. Like all tools, computers and database management systems can be applied skillfully and beneficially or crudely and maliciously. With more powerful tools, the protections must be increased to prevent mishap or misuse. Ethics and social conscience should be part of every database management training program. Systems should be designed with a sympathetic attitude towards users needs (Gottlieb, 1972; Sterling, 1974; Finerman, 1975).

Bachman, C.W., Trends in database management-1975, Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1975.

Curtice, R.M., Data base design using IMS/360, Proc. Fall Joint Computer Conf., AFIPS Press, Montvale, NJ, 1972.

Curtice, R.M., Data base design using a CODASYL system, Proc. ACM Natl. Conf., ACM, New York, 1972.

Date, C.J., An Introduction of Database Systems, Addison-Wesley Publ. Co., Reading, MA, 1975.

Finerman, A., Professionalism in the computing field, CACM 18,1 (January 1975).

Gerritsen, R., A preliminary system for the design of DBTG data structures, CACM 18,10 (October 1975) pp. 551-556.

Gottlieb, A., Management information systems, public policy and social change, Proc. Spring Joint Computer Conf., AFIPS Press, Montvale, NJ, 1972.

Hubbard, G. and Raver, N., Automating logical file design, Proc. Intl. Conf. on Very Large Data Bases, 1975.

Martin, J., Computer Data-base Organization, Prentice-Hall, Englewood Cliffs, NJ, 1975.

Martin, J., Principles of Data Base Management, Prentice-Hall, Englewood Cliffs, NJ, 1976.

Meltzer, H., An overview of the administration of data bases, Proc. 2nd USA-Japan Computer Conf., AFIPS Press, Montvale, NJ, 1975.

Morgan, H.L. and Soden, J.V., Understanding MIS failures, ACM Data Base, Vol. 5, Nos. 2,3, and 4 (Winter 1973) pp. 157-167.

Nolan, R.L., "Managing the Computer Resource: A Stage Hypothesis," CACM 16,7 (July 1973) pp. 399-405.

Nolan, R.L., "Computer Data Bases: The Future is Now," Harvard Business Review, Vol. 51, No. 5 (Sept.-Oct. 1973) pp. 98-114.

Nolan, R.L., Database - an emerging organizational function, Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1974.

Smith, J.M. and Smith, D.C.P., Data base abstraction, CACM (to appear).

Sterling, T.D., Guidelines for humanizing information systems, A report from Stanley House, CACM 17, 11 (November 1974).

Implementation and design of database management systems

Data management techniques have developed in three phases. The first phase was the use of sequential tape files to store information. The two utilization paradigms during this phase were the generation approach and the batched-searching approach. The generation approach required transactions to be collected and passed against an old-master file in order to do updates and produce a new-master. This process was repeated on a regular schedule, say once a week, and produced transaction listings and file status reports. Specific queries about a single record could not be processed and the listings were potentially a full week out of date. The batched-searching approach for archival files, such as journal citations, required users to collect queries until a single sequential pass of the entire file could be scheduled. Unanticipated queries and new applications required major efforts by program development teams and modifications to storage structures had a profound affect on all user programs.

The second phase of data management was brought on by the development and widespread use of disk drives beginning in the mid-1960s. As systems made increased use of the direct access capability of disk drives, the generation and batch-searching approaches yielded to the new technology which permitted on-line systems to remain up-to-date continuously and searching systems to provide immediate response to specific queries. The success of airline reservations, library, inventory, banking and personnel systems is testimony to the accomplishments of this second evolutionary phase.

Unfortunately, these systems were all designed to handle specific needs and required substantial efforts when new applications or facilities were to be added. Application programmers often had to worry about detailed file management issues such as handling overflows, insertions, deletions, reorganizations, space management and device dependent features.

To relieve the application programmer from the burden of these low level details, database management systems were developed. By making the obvious system decomposition into storage-structure dependent (handled by the database management system) and data structure dependent issues (handled by the application programmer), it became possible to simplify program development. Database management systems provided an improved measure of data independence - the separation of programs from the data they manipulate. This was accomplished largely by the development of sophisticated data description facilities which require designers to create data descriptions independently of their programs. As we move into the late stages of this third phase, we find that database management systems are increasingly flexible and allow unanticipated queries to be handled relatively easily. New applications can be implemented quickly and front-end facilities such as report generators and simple query languages eliminate the need for many programs.

The development of database management systems for this third evolutionary phase is a challenge to even the most sophisticated and experienced implementers. Techniques invented for file management, operating systems, compilers and telecommunications must be merged with the human factors requirements of users and the efficiency constraints imposed by management while coping with hardware reliability and privacy/security problems. The three papers in this section give some insight to the design of database management systems. The design of commercially available systems can be intuited from the user manuals, but little has appeared in the open literature. E.F. Codd organized a panel session at the 1975 National Computer Conference (Anaheim, CA, May 1975) entitled "Implementation of relational data base management systems" which brought together representatives of seven projects (transcript appears in Codd, 1975). Designers of relational database systems, who are prolific paper writers, have produced the following worthwhile documents: Whitney (1972), McLeod and Meldman (1975), Astrahan and Chamberlin (1975), Winslow (1975), Manacher (1975), Held et al (1975), Astrahan et al (1976) and Stonebraker et al (1976). CODASYL-DBTG developments can be traced from Bachman's work (1965, 1973), Bachman and Williams (1964), the CODASYL-DBTG report (1971), Schenk (1974) and Managaki et al (1975).

Other important papers include: Madnick (1969), Senko et al (1973), Senko (1975a, 1975b) and Stemple (1976).

A basic trend in database management systems design is the level structured approach exemplified by the 4-level Data Independent Access Model (Senko et al; 1973), which puts hardware and storage structure dependent issues at the lowest level and abstract data model issues at the highest level. This highest level, close to the real world problem environment, has been termed the infological level (Langefors and Sundgren, 1975) while storage structure issues are referred to as the datalogical level. Another abstract description has been given by the American National Standard Committee on Computers and Information Processing (ANSI/X3) Standards Planning and Requirements Committee (SPARC) (1975) which coined the terms Conceptual Schema and Internal Schema. ANSI/X3/SPARC also describes an External Schema which is the view of an information system held by an end user.

The diversity of approaches is healthy since it reflects active interest and willingness to try new directions. The experience of compiler design is encouraging: the early efforts eventually produced a synthesis of conceptual notions with agreement about the design strategy for compilers. It is now relatively easy to develop new compilers and there is a stable framework for healthy experimentation. Operating systems designers are beginning to reach a more stable concensus about the building blocks of an operating system. These experiences suggest that at least another decade of experimentation with database management systems will have to pass before the consensus emerges. This is an exciting era in the development of database management systems since the field is open and searching for new ideas.

A natural question is: If the first phase of data management facilities focused on sequential tape files, the second on direct access devices and the third on database management systems, then what will the fourth phase be? It seems clear that the next phase will be integrated database networks. The much talked about distributed database concept is just in its infancy and will mature only when there is more standardization among database systems and when the technology of data description and translation is improved. In the future we should expect to be able to interface a magazine subscription database with a census database for market research or to do correlative studies on independently maintained databases of medical records for clinical research. The idea of networking DMSs is simple, but the technology to enable inter-database searching is extremely complex.

ANSI/X3/SPARC Study Group on Data Base Management Systems Interim Report, FDT, Bulletin of ACM-SIGMOD the Special Interest Group on Management of Data, Vol. 7, No. 2, 1975.

Astrahan, M.M. and Chamberlin, D.D., "Implementation of a Structured English Query Language", presented at ACM-SIGMOD Conf., San Jose, Calif., May 14-16, 1975, to be published in CACM Vol. 18, No. 10, October 1975.

Astrahan, M.M. et al., System R: A relational approach to database management, ACM TODS 1,2 (June 1976).

Bachman, C.W. and Williams, S.B., A General Purpose Programming System for Random Access Memories, Proc. Fall Joint Computer Conf., AFIPS Press, Montvale, NJ, 1964.

Bachman, C.W., Integrated Data Store, DPMA Quarterly, January 1965.

Bachman, C.W., "Implementation Techniques for Data Structure Sets," Data Base Management Systems, D.A. Jardine (editor), North-Holland Publishing Company, 1974, pp. 147-160.

CODASYL Data Base Task Group, April 1971 Report, available from ACM, New York.

Codd, E.F. (ed.), "Implementation of Relational Data Base Management Systems", Bulletin of ACM-SIGMOD: The Special Interest Group on Management of Data, Vol. 7, No. 3-4, 1975, pp. 3-28.

Held, G.D., Stonebraker, M.R. and Wong, E., "INGRES: A Relational Data Base System", Proc. Natl. Computer Conf., 1975, AFIPS Press, Montvale, NJ.

Langefors, B. and Lundgren, B., Information System Architecture, Petrocelli/Charter Publishing Co., New York, 1975.

Madnick, S.E. and Alsop, J.W., A modular approach to file system design, SJCC, 1969, 1-13, AFIPS Press, Montvale, NJ.

Manacher, G.K., "Very Large, Minicomputer-Implementable Relational Database with Optimal Retrieval Performance", Proc. Intl. Conf. on Very Large Data Bases, Framingham, MA, September 22-24, 1975.

Managaki, M. Doine, T., Yoshimoto, H., and Katayama, H., A structural model for systems implementation and its application to CODASYL-DBTG, Proc. 2nd USA-Japan Computer Conf., AFIPS Press, Montvale, NJ, 1975.

McLeod, D.J. and Meldman, M.J., "RISS: A Generalized Minicomputer Relational Data Base Management System", Proc. Natl. Computer Conf., 1975, AFIPS Press, Montvale, NJ.

Mylopoulos, J., Schuster, S.A. and Tsichritzis, D., "A Multi-Level Relational System", Proc. Natl. Computer Conf., 1975, AFIPS Press, Montvale, NJ.

Schenk, H., Implementational aspects of the CODASYL-DBTG proposal in Klimbie, J.W. and Koffeman, K.L. (eds.) Data Base Management, North-Holland Publishing Co., Amsterdam, 1974, pp. 399-412.

Senko, M.E., Altman, E.B., Astrahan, M.M. and Fehder, P.L., Data structures and accessing in data base systems, IBM Systems J., 1973, 12, 30-93.

Senko, M.E., The DDL in the context of a multilevel structured description: DIAM II with FORAL, Data Base Description, Douque, B.C.M. and Nijssen, G.M. (eds.) pp. 239-240, North-Holland Publishing Co., Amsterdam, 1975.

Senko, M.E., Specification of stored data structures and desired output results in DIAM II with FORAL, Proc. ACM Conf. on Very Large Data Bases, 1975.

Stemple, D.W., A Database Management Facility for Automatic Generation of Database Managers, ACM Trans. on Database Systems, Vol. 1, No. 1, March 1976, pp. 79-94.

Stonebraker, M., Wong, E., Kreps, P., and Held, G., The Design and Implementation of IN RES, Electronics Research Lab, College of Engineering, Univ. of California, Berkeley, 1976.

Whitney, V.K.M., "RDMS" A Relational Data Management System", Proc. Fourth Intl. Symp. on Computer and Information Sciences, Miami Bearch, FL, December 14-16, 1972, Plenum Press.

Winslow, L.E., "An Efficient Implementation of Codd's Relational Model Data Base", Proc. COMPCON 75 (11th Annual IEEE Computer Society Conf.), Washington, D.C., September 9-11, 1975.

Query Languages

Database management systems will reach their full potential only if they can be made available to a wide range of users. While regular production reporting will remain the basic function of most database systems, unanticipated queries will be an increasing percentage of the utilization. The database system which permits easy and efficient querying will have a strong advantage over less flexible systems.

Codd (1974) describes the "casual user" as "one whose interactions with the system are irregular in time and not motivated by his job or social role. Such a user cannot be expected to be knowledgeable about computers, programming, logic or relations." Codd foresees enormous expansion of casual use of database systems and argues for natural language communication with computer systems to facilitate access to data. He anticipates improved natural language capabilities and stresses the notion of "clarification dialog" between the system and the user to resolve ambiguities in the user's query. While a number of elegant implementations (Winograd, 1971; Shapiro, 1975; Woods, 1972; Weizenbaum, 1966) and reviews (Rustin, 1971; Simmons, 1970) support the natural language interface concept, there are a number of skeptics (Montgomery, 1972; Hill, 1972). They argue that natural language facilities allow users to pose ill-formed queries and that most users would prefer a concise unambiguous query language which would help in their query formulation process by structuring the options. Of course, both sides are partially correct: irregular users would not invest in acquiring any syntactic notions and would prefer a completely unstructured query facility which used natural language. Those who become more regular users would be annoyed by the relatively lengthy language discourse and would prefer the conciseness and precision of an artificial language. There is a continuous spectrum of usage frequencies and a successful system should respond to the experience level of the user; gradually converting the user from a natural language user to a query language user. One approach would be to use the natural language system as a training aid - once the clarification dialogue is complete the system could display the query in the shortened artificial language form for study by the user. When a similar query arises again, the user might be able to provide the query language statement immediately.

Olle (1974) perceives the growth of "parameteric users" who use terminals for specialized functions such as finding the balance in an account, making a reservation or looking up a telephone number. These users would be required to enter only an account number or other key to get a response. A somewhat more flexible parametric use might be to respond to a series of "menus" presented by the terminal. The menu based approach relieves the user of creative action and requires simple responses to a pre-stored set of questions. These systems can be surprisingly powerful and acceptable to users, but they require a high speed terminal and communication facility to prevent boredome and anger as the options are presented.

Parametric use is popular and successful for business clerical situations in banking and airline reservations systems while the menu approach has found its greatest success in computer assisted education systems such as PLATO IV. More complex query languages involving boolean operators have been successful with trained users in library citation searching systems (Firschein and Summit, 1975) and report generating facilities such as MARK IV and the immediate access facility of System 2000.

Although these working systems testify to the realizability of simplified access facilities for non-programmers, much work remains to be done to broaden the availability of database systems. The simple and elegant structure of the relational model of data has prompted researchers to develop non-procedural language facilities which provide users the ability to formulate complex queries. Codd demonstrated in his original paper (Codd, 1970) a relational calculus and relational algebra which were "relationally complete". The unintentially deceptive term gives the reader the impression that all possible queries can be posed using either of these facilities. Relational completeness means that queries have the expressive power of the first order predicate calculus. Although many queries fit into this definition of completeness, there are many more which do not. The classic example of a query which is outside the bounds of relational completeness involves a personnel relation which has at least an employee and an employee-manager domain. The query: "Find all the managers, at all levels, of employee #4638!" is not relationally complete since the number of times that the relation has to be searched is a function of the data in the table. Providing a query facility to handle these kind of queries remains a challenge (Zloof, 1975b; Held and Stonebraker, 1976).

A second difficulty with query facilities is that the problems that individuals have, may not be easily responded to by a series of precise questions. Montgomery (1972) gives several realistic examples from information retrieval studies in nuclear physics but it is easy to create others: How do I compare the quality of the computer science education at four year and two year colleges? How can unemployment be reduced? What is the best way to get to Times Square? How is my daughter Sara doing in school? These questions may seem contrived, but they are typical questions which arise in realistic situations. In each case, the database system probably could not provide a direct answer, but there are databases which could contain relevant information for each of these situations. Advocates of artificial language query facilities argue that a precise language is helpful in forcing users to formulate reasonable queries. Further studies of user behavior are necessary to help implementors develop useful systems.

Interpretations and extensions of Codd's calculus and algebraic approaches have appeared with regularity in the past few years (Codd, 1971a, 1971b; Boyce et al, 1974; Chamberlin and Boyce, 1974; Pirotte and Wodon, 1974; Held et al, 1975 (in this volume)). The appeal of two dimensional query facilities has provoked some creative approaches (Zloof, 1975a (in this volume), 1975b; McDonald and Stonebraker, 1975; McDonald, 1975). Zloof argues that by avoiding keywords and using a positional notation at a CRT terminal, confusion and syntactic errors are minimized. Although two dimensional notation has not been particularly successful for general purpose programming languages, Zloof's approach is very appealing, easy to learn and easy to use as has been borne out by the psychological experiments of Thomas and Gould (1975) and by this author. Psychological studies of query facilities are being made so that language designers can test their products on subjects whose background is so vastly different from the typical programmer. The comparative study of Reisner et al (1975 (in this volume)); Reisner, 1976) suggested improvements which have been incorporated in new verions of SEQUEL. Other by-products of experimentation are a better understanding of human problem solving and teaching methods for query languages (Shneiderman, 1975).

The area of query languages is attractive to researchers in computer science, information retrieval and psychology because the potential number of users is vast. No one facility is likely to emerge as universal, just as no programming language has become universal. There will be room for a wide variety of specialized facilities to service the needs of a diverse population of users. However, the more easy to use and well-designed a particular facility is, the greater will be its acceptance.

Boyce, R.F., D.D. Chamberlin, W.F. King III, M.M. Hammer, "Specifying Queries as Relational Expressions: SQUARE", Proc. ACM SIGPLAN-SIGIR Interface Meeting, Gaithersburg, Maryland, November 4-6, 1973.

Chamberlin, W.F., R.F. Boyce, "SEQUEL: A Structured English Query Language", Proc. ACM-SIGMOD Workshop on Data Description, Access, and Control, Ann Arbor, Mich., May 1-3, 1974.

Codd, E.F., "A Relational Model of Data for Large Shared Data Banks", CACM 13, No. 6, June 1970 pp. 377-387.

Codd, E.F., "A Data Base Sublanguage founded on the Relational Calculus", Proc. 1971 ACM-SIGFIDET Workshop on Data Description, Access, and Control, San Diego, available from ACM New York.

Codd, E.F., "Relational Completeness of Data Base Sublanguages", Courant Computer Science Symposia 6, "Data Base Systems", New York City, May 24-25, 1971, Prentice Hall.

Codd, E.F., "Seven Steps to Rendezvous with the Casual User", Proc. IFIP TC-2 Working Conference on Data Base Management Systems, Cargese, Corsica, April 1-5, 1974, North-Holland, Amsterdam.

Firschein, O. and Summit, R.K., Providing the Public with Online Access to Large Bibliographic Data Bases, Proc. 2nd USA-Japan Computer Conf. 1975 AFIPS Press, Montvale, N.J.

Hill, I.D., Wouldn't it be nice if we could write computer programs in ordinary English -- or would it? Honeywell Computer Journal, Vol. 6, No. 2, 1972 pp. 76-83.

McDonald, N., M. Stonebraker, "CUPID: The Friendly Query Language", Proc. ACM Pacific Conf., San Francisco, April 17-18, 1975.

McDonald, N., CUPID: A graphics oriented facility for support of non-programmer interactions with a data base. Electronics Research Laboratory, College of Engineering, Univ. of Cali. Berkeley, 1975.

Montgomery, C.A. "Is Natural Language an Unnatural Query Language?", Proc. ACM National Conference, New York, ACM 1972 pp. 1075-1078.

Olle, T.W., Current and future trends in data base management systems, Proc. IFIP Congress 1974, North-Holland Publishing Co.

Pirotte, A., P. Wodon, "A Comprehensive Formal Query Language for a Relational Data Base: FQUL", Technical Report R283, M.B.L.E. Laboratoire de Recherches, Brussels, Belgium, December, 1974.

Reisner, P., Boyce, R.F., D.D. Chamberlin, "Human Factors Evaluation of Two Data Base Query Languages: SQUARE and SEQUEL", Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1975.

Reisner, P., Use of Psychological experimentation as an aid to development of a query language, IBM Report RJ 1707, 1976.

Rustin, R. (ed.), Natural Language Processing , Courant Computer Science Symposia 8, New York, December 20-21, 1971, Algorithmics Press, NY, 1973.

Shapiro, S.C. and Kwasny, S., Interactive Consulting via Natural Language, CACM 18, 8 (August 1975) 459-462.

Shneiderman, B., Experimental Testing in Programming Languages, Stylistic Considerations and Design Techniques, Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ (1975), 653-656.

Simmons, R.F., "Natural Language Question-Answering Systems: 1969", Comm. ACM 13, No. 1, January 1970 pp. 15-30.

Thomas, J.C. and J.D. Gould, A Psychological Study of Query by Example", Proc. Natl. Computer Conf., Anaheim, Calif., May 19-22, 1975.

Weizenbaum, J., "ELIZA -- A Computer Program for the Study of Natural Language Communication between Man and Machine", Comm. ACM 9, No. 1, January 1966 pp. 36-45.

Winograd, T., Understanding Natural Language, Academic Press, NY, 1972.

Woods, W.A., Kaplan, R.M., and B. Nash-Webber, "The Lunar Sciences Natural Language Information System", Bolt Beranek and Newman, Cambridge, MA, June 1972.

Zloof, M.M., "Query by Example", Proc. Natl. Computer Conf., 1975, AFIPS Press, Montvale, NJ.

Zloof, M.M., "Query by Example: The Invocation and Definition of Tables and Forms", Proc. Intl. Conf. on Very Large Data Bases, Framingham, MA, September 22-24, 1975.

Security, integrity, privacy and concurrency

These four terms collectively describe the delicate problems of ensuring that a database management system protects the valuable information in the database from intentional or unintentional malice, incorrect operations and improper disclosure. Security is generally taken to refer to physical protection of the database from external threats of destruction, alteration or copying. Integrity includes the protection from internal software or hardware errors, incorrect data and consistency. Privacy is the protection from improper disclosure of information and is closely tied to legal issues. Concurrency refers to the problems of providing multiple simultaneous access while ensuring the integrity of the database and preventing deadlock. The central problem in concurrent processing is guaranteeing efficiency and the correctness of results when multiple updates are in progress without locking large portions of the database.

Early research on these topics was done by operating systems designers, but the database management problem is more complex. The constraints of real-time response, efficiency and integrity are aggravated by the larger number of processing requests, enormous volume of data items and complexity of operations. The penalty for failure is high and the legal constraints of privacy cannot be compromised.

The Chamberlin et al paper (1975, in this volume) discusses these problems in the context of multiple views for a relational model of data. This important paper provides a lucid introduction to the issues confronting an implementer. The Manola and Wilson paper (1975, in this volume) deals with the multiple views problem by describing the schema/subschema concept of the CODASYL-DBTG Report. Other good discussions of this area can be found in Owens, 1971; Conway et al, 1972; Martin, 1973; Stonebraker and Wong, 1974; Fernandez et al, 1975; Gray et al, 1975; and Minsky, 1974, 1976.

The final two papers in this section deal with more technical issues. Turn reviews techniques for transforming the actual data encodings so as to disguise the data items. King and Collmeyer present an algorithm for permitting concurrent processing while preventing deadlocks. Each of these papers has good references to previous work in these two critical areas.

The technical issues are challenging enough, but the legal problems of privacy, security and integrity are even more disturbing. Computer fraud and spying by modifying or gaining access to information are growing concerns. Decades of court interpretations and legislation are necessary to establish new principles for dealing with data.

The personal privacy issue is of greatest concern to civil libertarians who, understandably, fear the misuse of computerized databases. The guidelines offered in the excellent report of the Secretary's Advisory Committee on Automated Personal Data Systems: U.S. Dept. of Health, Education and Welfare (1973) are a useful philosophical basis for discussion:

- There must be no personal data record-keeping systems whose very existence is secret.
- There must be a way for an individual to find out what information about him is in a record and how it is used.

- There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent.
- There must be a way for an individual to correct or amend a record of identifiable information about him.
- Any organization creating, maintaining, using or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data.

The ten points made by the British Younger Committee (Date, 1975) on privacy are also worthy of considerations:

1. Information should be regarded as held for a specific purpose and not be used, without appropriate authorization, for other purposes.

2. Access to information should be confined to those authorized to have it for the purpose for which it was supplied.

3. The amount of information collected and held should be the minimum necessary for the achievement of the specified purpose.

4. In computerized systems handling information for statistical purposes, adequate provision should be made in their design and programs for separating identities from the rest of the data.

5. There should be arrangements whereby the subject could be told about the information held concerning him.

6. The level of security to be achieved by a system should be specified in advance by the user and should include precautions against the deliberate abuse or misuse of information.

7. A monitoring system should be provided to facilitate the detection of any violation of the security system.

8. In the design of information systems, periods should be specified beyond which the information should not be retained.

9. Data held should be accurate. There should be machinery for the correction of inaccuracy and the updating of information.

10. Care should be taken in coding value judgments.

Management and technical personnel may be irritated by the burden imposed in adhering to these recommendations, but it cannot be stressed strongly enough, that successful adoption of large scale database management systems will occur only if the privacy demands have been met. If any attempt is made to weaken these guidelines, administrators, politicians and the public will rightfully condemn and prohibit such information sytems. Only by satisfying the demands of critics will it be possible to derive the benefits of large scale information systems. The technocrat's acceptance that in a small percentage of cases it is possible that privacy will be compromised will not be sufficient. The successful manager will be the one who most intensely defends the privacy of information stored in database management systems.

Critics of computerized information systems should be reminded that privacy may be more easy to protect in computerized systems since it will be possible to monitor access of the data in a more reliable way than with paper record files. Copies of paper files can be made without any record while computerized systems could audit every usage of a data item.

The complex issue of privacy will become an increasing concern as more files are placed in machine readable form. Computer professionals should confront this issue directly and honestly and should adhere to the highest ethical standards. Failure to do so will lead to a generally antagonistic response from other professionals and the public at large.

Chamberlin, D.D., Gray, J.N., and Traiger, I.L., "Views Authorization, and Locking in a Relational Data Base System", Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1975.

Conway, R., Maxwell, W. and Morgan, H., On the implementation of security measures in information systems. CACM 15, 4 (April 1972), 211-220.

Date, C.J., An Introduction to Database Systems, Addison-Wesley, Reading, MA 1975, pp. 298-299 reporting from Younger Committee (the Committee on Privacy) Report (July 1972), Cmmd. 5012. Her Majesty's Stationery Office, London.

Fernandez, E.B., Summers, R.C. and Coleman, C.D., "An Authorization Model for a Shared Data Base", Proc. ACM-SIGMOD Conf., San Jose, Calif., May 14-16, 1975.

Gray, J.N., Lorie, R.A., and Putzolu, G.R., "Granularity of Locks in a Large Shared Data Base", Proc. Intl. Conf. on Very Large Data Bases, Framingham, MA, September 22-24, 1975.

King, P.F. and Collmeyer, A.J., Database sharing - an efficient mechanism for supporting concurrent processing, Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1973.

Manola, F.A. and Wilson, S.H., Data Security implications of an extended subschema concept, Proc. 2nd USA-Japan Computer Conf., AFIPS Press, Montvale, NJ, 1975.

Martin, J., Security, Accuracy and Privacy in Computer Systems, Prentice-Hall, Englewood Cliffs, NJ, 1973.

Minsky, N., On interaction with data bases, Proc. ACM SIGFIDET Workshop on Data Description, Access and Control, 1974, pp. 51-62.

Minsky, N., Intentional Resolution of Privacy Protection in Database Systems, CACM 19, 3 (March 1976).

Owens, R.C., "Evaluation of Access Authorization Characteristics of Derived Data Sets", Proc. ACM-SIGFIDET Workshop on Data Description, Access, and Control, San Diego, Calif., November 11-12, 1971.

Stonebraker, M., and Wong, E., "Access Control in a Relational Data Base Management System by Query Modification", Electronics Research Lab., Univ. of California, Berkeley, ERL-M438, May 14, 1974.

Turn, R., Privacy transformations for data banks, Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1973.

U.S. Dept. of Health, Education and Welfare, Records, Computers and the Rights of Citizens, Report of the Secretary's Advisory Committee on Automated Personal Data Systems U.S. Government Printing Office, Washington, D.C., 1973.

Specification, simulation and translation of database systems

A number of high level problems associated with database management systems are covered in this section. Capraro and Berra (1974) review the difficulties in specifying the characteristics of a database system and of describing the problem environment which the system is to deal with. Teichroew (1972) presents a thorough survey of languages developed for problem specification and Teichroew and Sayani (1971) discuss the general problem of automating the construction of computer based systems. Another good source of material on this topic is the collection of papers edited by Couger and Knapp (1974). Nunamaker et al (1973) describe eight information processing systems and provide a framework for evaluation using the term "generalized database planning system". While these efforts are appealing, much work remains to be done since comparisons of database management systems are difficult to make and the diversity of needs is so great that automation seems inappropriate. Computer-based systems do poorly in complex highly varying ill-defined environments.

A promising area of research with a potentially high pay-off is the development of simulation modelling for database management systems. Nakamura et al (1975) describe a simulation system and present results for insurance company application. Simulation of database systems can be helpful in predicting performance and in selecting from the wide variety of implementation possibilities. This area is still in its infancy and must develop new strategies since the queuing model approach to simulation is not adequate. Other useful papers on simulation modeling of database structures include Senko et al (1970), Lum et al (1970), Cardenas (1973), Severance (1975), Yao and Merten (1975) and Siler (1976).

The conversion of data from one system to another, the physical reorganization of data to improve performance and the logical restructuring of data to accommodate new requirements are all considered within the topic of data translation. This area has received much attention recently because of the increasing need from database administrators. The paradigm has been to provide data description facility for describing the source and target data at logical and physical levels. These descriptions are input to a data translating system which processes the source data and produces the target data. An extension of the data translation idea to include translation of the programs is being considered by some researchers. While resolution of the data and program translation problem is not possible in general, successful translators have been created for restricted cases. Gerritsen (1975, in this volume) discusses an intriguing facility for the translation between the relational and CODASYL-DBTG data-structure-set models of data. This appealing idea sheds new light on the distinctions between the two approaches. Other references to the data translation idea can be found in Smith (1971, 1972), Taylor (1971), Fry et al (1972), Merten (1974), Shu et al (1975).

Capraro, G.T. and Berra, P.B., A database management problem specification model, Proc. Natl, Computer Conf., AFIPS Press, Montvale, NJ, 1974.

Cardenas, A.F., "Evaluation and selection of file organization - a model and system", CACM 16, 9 (Sept. 1973), 540-548.

Couger, J.D. Jr., and Knapp, R.W., Systems Analysis Techniques, John Wiley and Sons, New York, 1974.

Fry, J.P., Smith, D.P. and Taylor, R.W., An approach to stored data definition and translation, Proc. ACM SIGFIDET Workship on Data Description and Access, 1972, pp. 13-55.

Fry, J.F., Frank, R.L. and Hershey, E.A. III, A developmental mode for data translation, Proc. ACM SIGFIDET Workshop on Data Description and Access, 1972, pp. 77-105.

Gerritsen, R., The relational and network models of data bases-bridging the gap, Proc. 2nd USA-Japan Computer Conf., AFIPS Press, Montvale, NJ, 1975.

Lum, V.Y., Ling, H., and Senko, M.W., "Analysis of a Complex Data Management Access Method by Simulation Modeling," Proc. of the Fall Joint Computer Conf., 1970, AFIPS Press, Montvale, NJ.

Merten, A.G., "A data description language approach to file translation," Proc. ACM SIGFIDET Workshop of Data Description, Access and Control, 1974, pp. 191-205.

Nakamura, F., Yoshida, I, and Kondo, H., A simulation model for database system performance evaluation, Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1975.

Nunamaker, J.F. Jr., Swenson, D.E. and Whinston, A.B., Specifications for the development of a generalized data base planning system, Proc. Natl. Computer Conf., AFIPS Press, Montvale, NJ, 1974.

Senko, M.E., Lum, V.Y. and Owens, P.J., "A File Organization Evaluation Model (FOREM)", Proc. IFIP 1968, 1968, C19-C23, North-Holland Publishing, Amsterdam.

Severance, D.G., "A parametric model of alternative file structures", Information Systems 1, 2 (1975), 51-55.

Shu, N.C., Housel, B.C. and Lum, V.Y., CONVERT: A high level definition language for data conversion, CACM 18, 10 (October 1975) pp. 557-567.

Sibley, E.H. and Taylor, R.W., A data definition and Mapping Language, CACM 16, 12 (December 1973) pp. 750-759.

Siler, K.F., "A Stochastic Evaluation Model for Database Organizations in Data Retrieval Systems" CACM 19, 2 (Feb. 1976), 84-95.

Smith, D.P., An Approach to Data Description and Conversion, Ph.D. Dissertation, University of Pennsylvania, 1971.

Smith, D.P., "A method for data translation using the Stored-Data Definition and Translation Task Group Languages," Proc. of the ACM SIGFIDET Workshop on Data Description and Access, 1972, pp. 107-124.

Teichroew, D., "A Survey of Languages for Stating Requirements for Computer-Based Information Systems," FJCC, 1972, pp. 1203-1224, AFIPS Press, Montvale, NJ.

Teichroew, D. and Sayani, H., Automation of System Building, Datamation 15 (August 1971) pp. 25-30.

Yao, S.B. and Merten, A., "Selection of file organizations through analytic modeling", Proc. Very Large Data Bases (Sept. 1975).